Large-Scale Galaxy Bias

Vincent Desjacques[a,b], Donghui Jeong[c], Fabian Schmidt[d]

[a]*Physics Department, Technion, 3200003 Haifa, Israel*
[b]*Département de Physique Théorique and Center for Astroparticle Physics, Université de Genève, 24 quai Ernest Ansermet, CH-1221 Genève 4, Switzerland*
[c]*Department of Astronomy and Astrophysics, and Institute for Gravitation and the Cosmos, The Pennsylvania State University, University Park, PA 16802, USA*
[d]*Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Straße 1, 85748 Garching, Germany*

Abstract

This review presents a comprehensive overview of galaxy bias, that is, the statistical relation between the distribution of galaxies and matter. We focus on large scales where cosmic density fields are quasi-linear. On these scales, the clustering of galaxies can be described by a perturbative bias expansion, and the complicated physics of galaxy formation is absorbed by a finite set of coefficients of the expansion, called *bias parameters*. The review begins with a detailed derivation of this very important result, which forms the basis of the rigorous perturbative description of galaxy clustering, under the assumptions of General Relativity and Gaussian, adiabatic initial conditions. Key components of the bias expansion are all leading local gravitational observables, which include the matter density but also tidal fields and their time derivatives. We hence expand the definition of *local bias* to encompass all these contributions. This derivation is followed by a presentation of the peak-background split in its general form, which elucidates the physical meaning of the bias parameters, and a detailed description of the connection between bias parameters and galaxy statistics. We then review the excursion-set formalism and peak theory which provide predictions for the values of the bias parameters. In the remainder of the review, we consider the generalizations of galaxy bias required in the presence of various types of cosmological physics that go beyond pressureless matter with adiabatic, Gaussian initial conditions: primordial non-Gaussianity, massive neutrinos, baryon-CDM isocurvature perturbations, dark energy, and modified gravity. Finally, we discuss how the description of galaxy bias in the galaxies' rest frame is related to clustering statistics measured from the observed angular positions and redshifts in actual galaxy catalogs.

Keywords: cosmology, dark matter, galaxy bias, galaxy clustering, large-scale structure, primordial non-Gaussianity

arXiv:1611.09787v5 [astro-ph.CO] 14 Jan 2019

Contents

---

* Sections that are more technical in nature and not essential to the understanding of later sections are marked with an asterisk.

1 Introduction

The observed distribution of galaxies, quasars, and clusters of galaxies—the large-scale structure of the Universe, Fig. 1—is one of the foundations of our knowledge about the history of the Universe. These tracers can be observed out to cosmological distances, and can thus be used to survey significant fractions of the observable Universe. If we understand how the distribution of tracers is related to the underlying distribution of matter, we can access a wealth of information on the composition of the Universe, properties of dark matter, dark energy and gravity, as well as the nature of the process that produced the initial seeds of structure. The relation between luminous tracers and matter, which is known as *bias*, thus forms a key ingredient in the interpretation of the observed large-scale structure.

1.1 *Historical review*

Perhaps the first example of cosmological conclusions drawn from sky surveys is [1], who showed that the observed flux distribution of radio sources is inconsistent with a static homogeneous Universe. Beyond this qualitative conclusion, it is difficult to extract information from the 1-point function of galaxies. Thus, most cosmological inferences have been based on the next-order statistic of the galaxy density field, the two-point correlation function and its Fourier transform, the power spectrum [2] (a basic introduction to the description of statistical fields is provided in Appendix A). Two-point statistics were first measured with significant signal-to-noise ratio in early surveys of galaxies and clusters of galaxies [3, 4, 5, 6, 7, 8, 9, 10, 11]. Already in these first measurements it became clear that the correlation function of galaxies and clusters is not the same, which implies that they cannot both be unbiased tracers of the matter density fluctuations. Consider a simple ansatz which locally relates the density contrast of galaxies or clusters of galaxies to that of matter at a fixed time:

$$\delta_g(\boldsymbol{x}) \equiv \frac{n_g(\boldsymbol{x})}{\overline{n}_g} - 1 = b_1\,\delta(\boldsymbol{x}) = b_1 \left(\frac{\rho_m(\boldsymbol{x})}{\overline{\rho}_m} - 1 \right) , \tag{1.1}$$

where all quantities are evaluated at the same fixed time, which we leave implicit here, $\overline{n}_g$ is the mean *comoving* (see Sec. 1.3) number density of galaxies, while $\overline{\rho}_m$ is the comoving background matter density, and b_1 is a parameter that we call *bias*. Then, the two-point function of galaxies (or clusters of galaxies) is enhanced by a factor of $(b_1)^2$ over the matter two-point function. If we allow for clusters to have a larger bias parameter than galaxies, their different observed correlation functions can be explained.

A relation of the form Eq. (1.1) with $b_1 \neq 1$ implies that the galaxy density $n_g(\boldsymbol{x})$ is not linearly proportional to $\rho_m(\boldsymbol{x})$, as otherwise their fractional perturbations would be equal. Instead, the galaxy density has to be a nonlinear function of the matter density. In his seminal paper [12], Kaiser laid out a physical picture for such a nonlinear function, by positing that clusters form at the locations of rare, high-density excursions of the matter density field (we will describe this ansatz in detail in Sec. 2.1). This argument was subsequently refined by Bardeen, Bond, Kaiser, and Szalay (BBKS [13]), who derived the statistics of peaks in Gaussian random fields (Sec. 6). Further, Refs. [12, 13] and, subsequently, Refs. [14, 15, 16] formulated the "peak-background split," which establishes a connection between the bias parameters and the mean abundance of tracers (Sec. 3). Importantly, in the meantime Ref. [17] showed that bias can be made much more general than the specific examples considered in [12, 13].

The advent of larger galaxy surveys, in particular the CfA [18] and APM [19] surveys, allowed for the first robust cosmological inferences from galaxy clustering [20, 11, 21, 22], for which bias is a crucial ingredient. In particular, Ref. [23] showed that the two-point correlation function of galaxies measured in the CfA and APM
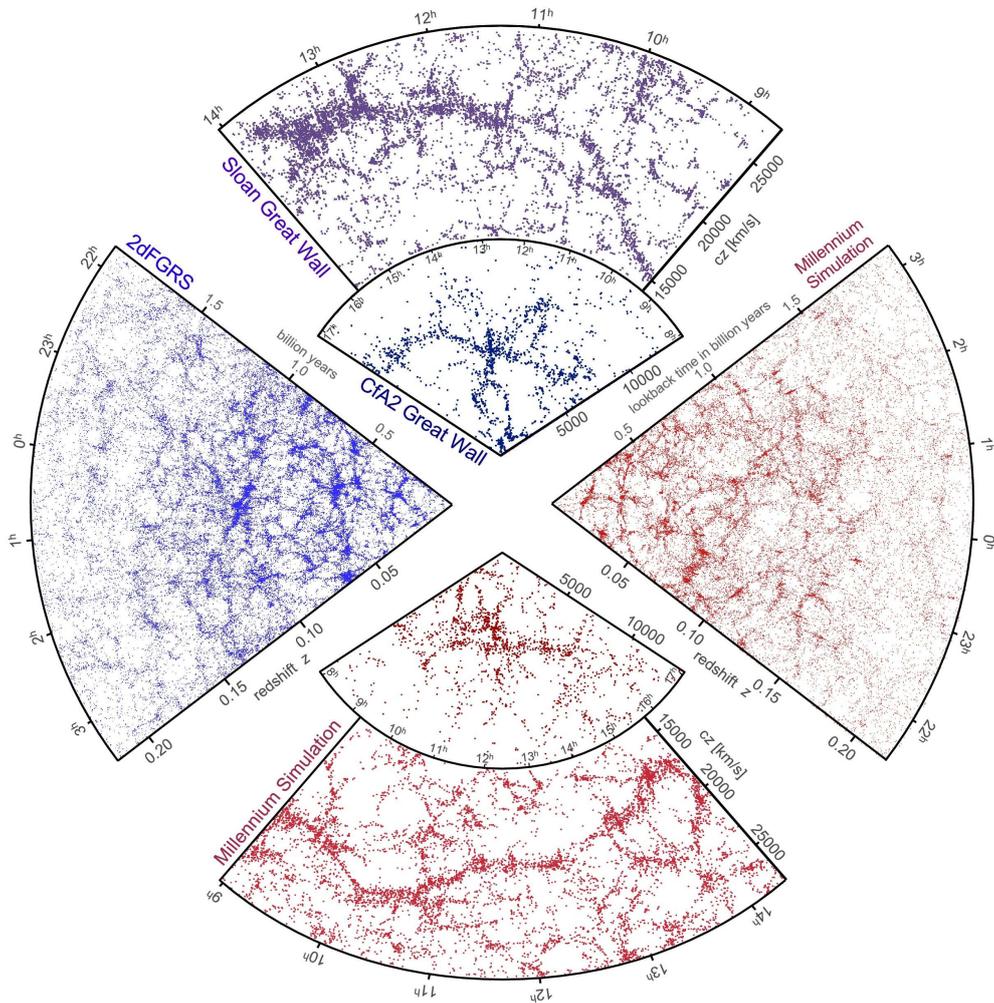
Figure 1: Two-dimensional slice projections (pie diagram) of the measured locations of galaxies in the CfA2, 2dF, and SDSS galaxy redshift surveys (top left half). The lower right half shows the location of galaxies which were assigned to dark matter halos in the *Millennium* gravity-only N-body simulation using a semi-analytical prescription. It is apparent that the simulation, which assumes a flat ΛCDM cosmology, qualitatively reproduces the observed large-scale structure of the Universe very well. From [32].

surveys was impossible to reconcile with the predictions of a cold dark matter (CDM) dominated cosmology (with matter density parameter $\Omega_m \leq 1$) unless a bias parameter b_1 is introduced following [12]. Moreover, the results of these surveys played an important role in establishing the now familiar standard model of cosmology, the spatially flat ΛCDM model. To this day, this model consistently describes both the large-scale galaxy power spectrum and the cosmic microwave background (CMB) [24, 25, 26, 27, 28, 29, 30, 31], and any significant departures from this model are by now tightly constrained. We briefly recap this history here.

In the early 1990s, the flat, matter-dominated Einstein-de Sitter (EdS) Universe [33], was favored on theoretical grounds. Ref. [34] pointed out that this scenario made inconsistent predictions for the relative amplitude between large-scale and smaller scale clustering of the APM galaxies, given the constraints on the primordial amplitude of perturbations from CMB temperature fluctuations from COBE [35]. Essentially, since the shape of the two-point function of tracers, even those that are biased according to Eq. (1.1), is the same as that of matter on large scales, the former could be used to rule out the shape of the matter

correlation function predicted by the EdS scenario. Moreover, this cosmological model did not correctly describe galaxy velocity statistics [36]. As noted in [37], the introduction of a cosmological constant, with a magnitude that corresponds to roughly 80 percent of the present-day total mean energy density, could resolve the discrepancy (see also the earlier discussion in [38]). This eventually led to the establishment of the standard flat ΛCDM cosmology, whose confrontation with observations has been enormously successful [24, 25, 26, 27, 28, 29, 30, 31, 39]. We refer the reader to [40] for a review of recent cosmological constraints from galaxy clustering.

Another milestone was reached in the detection of the baryon acoustic oscillation (BAO) feature [41] in the two-point correlation function of the SDSS [42] and the power spectrum of the 2dF [25] data sets (see Fig. 1 for representations of these surveys). This feature can be used as a robust standard ruler to measure the expansion history of the Universe [43, 44]. In addition, redshift-space distortions [45] allow for a measurement of the growth rate of structure [46, 47]. These probes are by now part of the core science of a number of ongoing[1] and next-generation experiments.[2]

Despite these successes, the exploitation of the cosmological information in large-scale galaxy surveys is still in an early stage: (i) Beyond the two-point function, many studies suggest that higher n-point functions contain significant amounts of information as well [59, 60, 61, 62] (Sec. 4). (ii) The potential of galaxy clustering to test General Relativity on scales of 30–150 Mpc has only recently been realized [63, 64, 65, 66] (Sec. 8.3). (iii) Departures from perfect Gaussian initial conditions can leave distinct imprints in the large-scale clustering of biased tracers [67, 68]. This leads to the fascinating (and surprising) prospect that galaxy clustering can provide insights on early-Universe physics that are complementary to those from the CMB (Sec. 7).

A common theme of all these, and many other possible applications of galaxy statistics is that they critically rely on a robust physical description of the relation between galaxies and matter: **galaxy bias**.

We emphasize that galaxies are not the only tracers of large-scale structure. Important other examples include groups and clusters of galaxies, voids, quasars, the Lyman-α forest, line intensity mapping of Hα, CO, the 21cm hydrogen hyperfine structure transition, and others, as well as diffuse backgrounds such as the cosmic infrared background. Keeping this in mind, in this review we will use the term "galaxies" as a convenient stand-in for general tracers of the large-scale structure; aspects that are specific to different kinds of tracers are discussed in Sec. 9.

1.2 From initial conditions to observed galaxies: the role of bias

In order to be able to extract cosmological information from the observed clustering of galaxies, we need a reliable model, or, better yet, theory for the statistics of galaxies, given the properties of the very small ($\delta \sim 10^{-4}$) initial perturbations in the early Universe, and the background cosmology. The different ingredients necessary in this endeavor are summarized in the flowchart in Fig. 2.

First, we require a theoretical description of the distribution of matter itself, which is governed by the growth under gravitational collapse starting from the initial perturbations (see [69, 70] for reviews). Unlike the CMB, which is accurately described by linear (first-order) perturbation theory, the perturbations in the matter density of the Universe at redshifts $z \lesssim 10$ are nonlinear. Crucially, the degree of nonlinearity depends on the scales considered. One can broadly divide large-scale structure into two regimes: large, *quasi-linear scales*, where perturbation theory (see below) converges to the correct result if carried out to sufficiently high order; and the smaller *nonlinear scales*, which cannot be described by perturbation theory. The description of LSS on nonlinear scales has to rely on numerical simulations and simplified heuristic models. In this review, we focus on quasi-linear scales. While this neglects the cosmological information that can be extracted from LSS on small scales, the key advantage of restricting to large scales, as we will see, is that we are able to obtain a rigorous theory of galaxy clustering. In order to describe the distribution of matter on quasi-linear scales, nonlinear cosmological perturbation theory (PT) approaches have been developed ([71]; see [72] for a review). PT provides a robust theoretical foundation for the description of the

---

[1]Dark Energy Survey (DES) [48], eBOSS [49], HETDEX [50], HSC [51], KiDS [52]
[2]DESI [53], Euclid [54], 4MOST [55], LSST [56], PFS [57], WFIRST [58]

quasi-linear matter density and tidal field, and has made significant progress in the past decade with the development of several new technical approaches.[3] In this review, we will only rely on fairly basic results of perturbation theory. These are summarized in Appendix B.

Now, galaxy surveys of course do not measure the matter density field itself, but rather the distribution of galaxies or other tracers, that is, of highly nonlinear objects which are the result of a complex formation process. *Bias* describes, in a statistical sense, the relation of the distribution of these objects to that of matter. Clearly, this is a very complicated relation in general: galaxy formation takes place over long periods of time and in interaction with the formation of structure in the matter distribution, and is currently far from being understood in detail (see [92] for an overview). Given a perturbation-theory based description of the large-scale matter density and tidal fields on quasi-linear scales, the goal of the theory of galaxy bias is to write the local number density of galaxies n_g as the most general function of the properties of the large-scale environment that is allowed by general covariance under coordinate transformations. Key theoretical advances have been made in the understanding of bias over the past decade, paralleling those for the matter density field mentioned above.

Remarkably, on quasi-linear scales, all the complexities of galaxy formation can be absorbed into a finite number of parameters (at each order in perturbations and at fixed time), the *bias parameters*. This is a nontrivial result, and relies on the fact that on large scales, structure formation is completely determined by the action of gravity. In fact, one can show that, at linear order in perturbation theory (and assuming Gaussian, adiabatic initial conditions), the ansatz Eq. (1.1) is correct and complete, up to an additional additive noise term. More generally, the statistics of galaxies at a given order in perturbation theory are determined by a well-defined set of bias parameters which can be constrained using these statistics. In this way, we effectively marginalize over the unknown details of the galaxy formation process, while robustly extracting cosmological information from galaxy surveys. This review provides a comprehensive overview of this approach, and connects it to the other aspects of the theory of galaxy clustering on quasi-linear scales, as summarized in Fig. 2.

To summarize, the *perturbative theory of galaxy clustering*, valid on quasi-linear scales, is based on two key ingredients: (i) a perturbation theory prediction for the matter density and tidal field; (ii) a complete parametrization of galaxy bias at each order in perturbation theory.

Despite the complexities of galaxy formation mentioned above, there is a robust, well-established fact that we can build on: galaxies reside in massive, gravitationally bound structures called *halos*. As dark matter makes up approximately 80% percent of all matter, the potential well of gravitationally bound structures is dominated by dark matter. Consequently, the halos that host galaxies are dark-matter dominated. The connection between galaxies and halos is well established numerically and observationally, for example through stacked weak gravitational lensing [93, 94]. The lower half of Fig. 1 shows an example of how halos identified in a gravity-only simulation can be populated with galaxies to realistically reproduce their observed distribution. Thus, the understanding of the large-scale clustering of dark matter halos is a physically well-motivated intermediate step toward the same for the clustering of galaxies themselves. Moreover, the formation, structure, and clustering of halos can be studied reliably and in detail via gravity-only N-body simulations. Hence, several sections of this review will deal with numerical results on, and physical models of, halo clustering (in particular, Sec. 4–6).

1.3 *Notation and terminology*

Throughout the bulk of this review, we work in conformal-Newtonian gauge and comoving coordinates, and restrict to scalar perturbations. Moreover, we assume a spatially flat background, as supported by observations, although this assumption is of no relevance to the topic of bias. Then, the perturbed Friedmann-

---

[3]Renormalized perturbation theory [73, 74, 75], renormalization group approach [76], closure theory [77], Lagrangian perturbation theory [78, 79, 80], TimeRG theory [81, 82], effective field theory [83, 84, 85, 86, 87, 88, 89, 90], time-sliced perturbation theory [91], among others.

Baryon density parameter	$\Omega_{b0}h^2$	0.022307
CDM density parameter	$\Omega_{c0}h^2$	0.11865
Neutrino density parameter	$\Omega_{\nu0}h^2$	0.000638
Cosmological constant parameter	$\Omega_{\Lambda0}$	0.69179
Hubble constant today	$H_0 = h\,100\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$	$67.78\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$
Scalar spectral index	n_s	0.9672
Scalar power spectrum amplitude	$\mathcal{A}_s$	2.147×10^{-9}
Matter power spectrum normalization at t_0	$\sigma_8 = \sigma(R = 8\,h^{-1}\,\mathrm{Mpc}, z = 0)$	0.8166

Table 1: Parameters of the flat ΛCDM reference cosmology used for numerical results (maximum likelihood values for "base_plikHM_TTTEEE_lowTEB_lensing_post_BAO_H080p6_JLA" from Planck 2015 [29, 97]). Here, the density parameters are defined as $\Omega_{X0} = \overline{\varrho}_X(t_0)/\varrho_{\mathrm{crit}}(t_0)$, where ρ_{crit} is the critical density, and t_0 denotes today's epoch. σ_8 is a derived parameter.

Robertson-Walker (FRW) metric can be written as

$$ds^2 = a^2(\tau)\left[-(1 + 2\Phi)d\tau^2 + (1 - 2\Psi)\delta_{ij}dx^i dx^j\right],\tag{1.2}$$

where in General Relativity $\Phi = \Psi$ in the absence of anisotropic stress. The matter density contrast δ is, in general, not a local observable. Thus, it should not really appear by itself in a bias expansion [such as in Eq. (1.1)]. On scales much smaller than the comoving horizon, $\mathcal{H} \equiv aH$, however, this issue is irrelevant for all popular gauge choices. Strictly speaking, the matter density perturbation δ should be understood as being defined in synchronous-comoving gauge throughout this review [95, 96]. We will discuss these issues in Sec. 2.9.

We will use the reference cosmology defined in Tab. 1 for all numerical results, unless otherwise indicated. Note that numerical results and figures taken from the published literature are based on different cosmological parameters. Throughout the review, the matter (ρ_m), galaxy (n_g), and halo densities (n_h) are defined as *comoving densities*, i.e. as physical densities multiplied by a^3; for example, $\rho_m = a^3 \varrho_m$, where ϱ_m is the physical matter density. Further, we define the time-dependent density parameter for any component X as $\Omega_X(\tau) \equiv \overline{\varrho}_X(\tau)/\varrho_{\mathrm{crit}}(\tau)$, where $\varrho_{\mathrm{crit}}(\tau) = 3H^2(\tau)/8\pi G$. Their values today are denoted as Ω_{X0}.

Let us now generalize the simple linear bias in Eq. (1.1), by writing

$$\delta_g(\boldsymbol{x}, \tau) = \sum_O b_O(\tau)O(\boldsymbol{x}, \tau).\tag{1.3}$$

Here, O are operators, or statistical fields, which describe properties of the galaxies' environment on which their density can depend. Each operator is multiplied by a corresponding bias parameter b_O, which, at fixed time, is merely a number. Eq. (1.1) already provides one example, with $O = \delta$ and $b_O = b_1$. This is easily generalized to a whole set of operators $O = \delta^N$, $N \geq 1$. The corresponding bias parameters have historically been known as "local bias parameters" b_N (see Sec. 2.1), defined through

$$b_N \equiv N!\,b_{\delta^N}.\tag{1.4}$$

However, as we describe below, we will broaden the definition of *local bias* to a more general, physically motivated class. Whenever we need to single out the restricted set of bias parameters in Eq. (1.4), we will refer to the expansion in a power series in δ solely as *local-in-matter-density (LIMD)* bias.

Note that we will not distinguish in notation between bias parameters of halos and those of galaxies. This is because almost all of the bias parameters we encounter in the review are equally relevant for galaxies and halos. Where results only apply to halos, we will clearly indicate this.

The subsequent sections will derive exactly which operators appear in the proper, complete bias expansion. In addition to the term *LIMD* defined above, we assign certain names to different categories of operators that will appear in the following, which we state here mainly for readers with experience in the

9

Probability density function (PDF)	$p(x)$
Spatial derivatives, Laplace operator	$\partial_i \equiv \partial/\partial x^i$, $\nabla^2 \equiv \delta^{ij}\partial_i\partial_j$,
	$\quad \mathcal{D}_{ij} \equiv \frac{\partial_i\partial_j}{\nabla^2} - \frac{1}{3}\delta_{ij}$
Fourier transform	$f(\boldsymbol{k}) \equiv \int d^3\boldsymbol{x}\, f(\boldsymbol{x})e^{-i\boldsymbol{k}\cdot\boldsymbol{x}}$
Momentum integral	$\int_{\boldsymbol{k}} \equiv \int \frac{d^3\boldsymbol{k}}{(2\pi)^3}$
Sum notation	$\boldsymbol{k}_{1\cdots n} \equiv \boldsymbol{k}_1 + \cdots + \boldsymbol{k}_n$
Connected n-point function[1]	$\langle O(\boldsymbol{k}_1)\cdots O(\boldsymbol{k}_n)\rangle_c$
Dirac delta distribution[2]	$\delta_D(\boldsymbol{x}) = \int_{\boldsymbol{p}} e^{i\boldsymbol{p}\cdot\boldsymbol{x}}$
n-point correlator without	$\langle O(\boldsymbol{k}_1)\cdots O(\boldsymbol{k}_n)\rangle'$, where
momentum conservation	$\langle O(\boldsymbol{k}_1)\cdots O(\boldsymbol{k}_n)\rangle = \langle O(\boldsymbol{k}_1)\cdots O(\boldsymbol{k}_n)\rangle'\,(2\pi)^3\delta_D(\boldsymbol{k}_{1\cdots n})$
Kronecker symbol	δ_{ij}, δ_{NM}
Heaviside step function	$\Theta_H(x) = 1$ for $x > 0$ and 0 otherwise
Complementary error function	$\mathrm{erfc}(x) = 1 - \mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_x^\infty du\, e^{-u^2}$
Laguerre polynomials	$L_n(x)$
Legendre polynomials	$\mathcal{L}_l(x)$
Probabilists' Hermite polynomials	$H_N(x)$

[1] See Appendix A.
[2] Note that this implies $(2\pi)^3 \int_{\boldsymbol{k}} \delta_D(\boldsymbol{k} - \boldsymbol{k}_0)f(\boldsymbol{k}) = f(\boldsymbol{k}_0)$.

Table 2: List of mathematical symbols and notations.

subject. The definition of these categories can be skipped on a first reading, as it jumps somewhat ahead of the proceedings. Specifically, we distinguish three categories of bias terms:

- *Local bias:* this includes all operators O that involve exactly two spatial derivatives[4] for each instance of the gravitational potential Φ. This includes the LIMD terms, i.e. powers of the density δ^N, as δ is related to $\nabla^2\Phi$ through the Poisson equation (Tab. 4). It also includes powers of the tidal field, and time derivatives of the density and tidal field. The physical reasoning behind this category is that these operators constitute the leading *local gravitational observables* of long-wavelength spacetime perturbations. Note that Φ itself, or first derivatives $\partial_i\Phi$, are not locally observable and hence not included in this category, as required by the equivalence principle.

- *Higher-derivative bias:* this includes operators that involve more than two derivatives acting on a single instance of Φ; for example, $\nabla^2\delta$ or $(\partial_i\partial_k\partial_l\Phi)^2$. These are clearly also local gravitational observables. However, they are sub-leading in the limit of very long-wavelength density perturbations. Moreover, each additional spatial derivative has to be multiplied by a spatial scale R_* in order to render the contribution to δ_g dimensionless [see Eq. (1.3)]. R_* corresponds to the characteristic spatial scale of the formation of the galaxies considered.

- *Nonlocal bias:* this class, finally, includes operators with fewer than two, or a fractional number of derivatives of the potential Φ. These terms cannot be induced by gravitational evolution or local physical processes, as they are forbidden by the equivalence principle, and so they must be imprinted in the initial conditions (with a single minor exception described in Sec. 8.2). The most important case is that of primordial non-Gaussianity, considered in Sec. 7.

The term "scale-dependent bias" has been used frequently in the literature to denote a nontrivial function

---

[4]Note that the operator $\partial_i\partial_j/\nabla^2$ counts as zero net spatial derivatives.

of wavenumber k in the Fourier-space relation between $\delta_g(\boldsymbol{k})$ and $\delta(\boldsymbol{k})$, or other operators. Unfortunately, this term can apply equally to the classes of higher-derivative bias and nonlocal bias, which, as we argued above, have distinct physical origin. We will only use the term "scale-dependent bias" in the context of primordial non-Gaussianity (Sec. 7), specifically to denote the leading nonlocal term appearing in that case.

For reference, mathematical symbols and conventions used throughout this review are summarized in Tab. 2, abbreviations are listed in Tab. 3, and a reference list of physical variables is given in Tab. 4.

1.4 List of new results

Beyond summarizing the current state of the field, this review also includes numerous new results, mainly related to connections between different aspects of bias that had previously been overlooked. We briefly list these new results here, for the benefit of expert readers.

- A complete enumeration of higher-derivative bias up to second order (Sec. 2.6), and of stochastic contributions to the galaxy density and velocity (Sec. 2.8).

- A derivation of the connection between peculiar acceleration of galaxies and the leading velocity bias and higher-derivative bias (Sec. 2.7), elucidating the reason for the different time dependences of velocity bias obtained in the literature.

- A proof that the Eulerian peak-background split bias parameters are exactly the large-scale renormalized bias parameters (Sec. 3.4).

- A previously-overlooked stochastic contribution to the halo-matter cross-power spectrum at next-to-leading (1-loop) order.

- A rigorous derivation of the connection between bias parameters inferred from moments and "scatter-plot" methods with those measured through n-point functions (Sec. 4.2–4.3). The former measure different bias parameters (which we call *moments biases*), a fact which has not been recognized previously.

- Excursion-set bias parameters obtained using a numerical solution of the exact Langevin equation for a tophat filter (Sec. 5.8).

- Forecasts for constraints on primordial non-Gaussianity from the power spectrum and bispectrum of galaxies for current and upcoming surveys (Sec. 7.6.1), including, for the first time, all relevant bias parameters and stochastic terms.

BAO	Baryon acoustic oscillation
CMB	Cosmic microwave background
EdS	Einstein-de Sitter (flat, matter-dominated Universe)
EFT	Effective field theory
ESP	Excursion-set peaks
FoF	Friends-of-friends algorithm (Appendix D)
ΛCDM	Λ cold dark matter
LIMD	Local in matter density (previously commonly known as "local bias")
LPT	Lagrangian perturbation theory
LO	Leading order (tree level)
NLO	Next-to-leading order (1-loop)
PBS	Peak-background split
PNG	Primordial non-Gaussianity
PT	Perturbation theory
RSD	Redshift-space distortions
SO	Spherical overdensity (Appendix D)
SPT	Eulerian standard perturbation theory

Table 3: List of abbreviations used in the text.

Quantity	Symbol	Defining relation
Conformal time	τ	$d\tau \equiv a^{-1}dt$
Eulerian comoving coordinate	$\boldsymbol{x}$	Eq. (1.2)
Time derivative	$\dot{f}$	$\dot{f} \equiv df/dt$
Hubble rate	H	$H \equiv \dot{a}/a$
Conformal Hubble rate	$\mathcal{H}$	$\mathcal{H} \equiv a^{-1}da/d\tau = aH$
Mean comoving (physical) matter density	$\overline{\rho}_m\ (\overline{\varrho}_m)$	$\overline{\rho}_m(\tau) \equiv a^3(\tau)\overline{\varrho}_m(\tau)$
Mean comoving halo number density	$\overline{n}_h$	$\overline{n}_h(M,z) \equiv \partial^4\overline{N}_h/(\partial^3\boldsymbol{x}\,\partial\ln M)$
Linear matter growth factor	$D(\tau)$	Eq. (B.9)
Logarithmic growth rate	$f(\tau)$	$f \equiv d\ln D/d\ln a$
Gravitational potential	Φ	Eq. (1.2)
Primordial Bardeen potential	ϕ	$\Phi(k)\vert_{\text{mat. dom.}} = T(k)\phi(k)$
Primordial curvature pert. in comoving gauge	$\mathcal{R}$	$\mathcal{R} = (5/3)\phi$ in matter domination
Lagrangian comoving coordinate	$\boldsymbol{q}$	$\boldsymbol{q} = \lim_{\tau\to 0}\boldsymbol{x}_{\text{fl}}(\boldsymbol{q},\tau)$
Comoving coordinate of fluid trajectory	$\boldsymbol{x}_{\text{fl}}(\boldsymbol{q},\tau)$	$\boldsymbol{x}_{\text{fl}}(\boldsymbol{q},\tau) \equiv \boldsymbol{q} + \boldsymbol{s}(\boldsymbol{q},\tau)$
Lagrangian displacement	$\boldsymbol{s}(\boldsymbol{q},\tau)$	Eq. (2.25)
Peculiar fluid velocity	$\boldsymbol{v}$	$\boldsymbol{v} \equiv a\,\dot{\boldsymbol{x}} = d\boldsymbol{x}/d\tau$
Matter density contrast[1]	δ	Eq. (1.1); $\delta(\boldsymbol{x},\tau) = 2/(3\Omega_m\mathcal{H}^2)\nabla^2\Phi(\boldsymbol{x},\tau)$
Density contrast of galaxies (general tracer)	δ_g	Eq. (1.1)
Halo density contrast	δ_h	$\delta_h(\boldsymbol{x},\tau) \equiv n_h(\boldsymbol{x},\tau)/\overline{n}_h(\tau) - 1$
Lagrangian halo density contrast[2]	$\delta_h^L(\boldsymbol{q},\tau_0)$	Tab. 5 on p. 24
Tidal field	K_{ij}	$K_{ij} \equiv (\partial_i\partial_j/\nabla^2 - \delta_{ij}/3)\delta$
Linearly extrapolated initial density field	$\delta^{(1)}$	$\delta^{(1)}(\boldsymbol{k},\tau) \equiv \mathcal{M}(k,\tau)\phi(\boldsymbol{k})$ [Eq. (7.1)]
Operator constructed out of density field	O	e.g., $O(\boldsymbol{x},\tau) = [\delta(\boldsymbol{x},\tau)]^2$
Smoothed field	O_R	$O_R(\boldsymbol{x},\tau) \equiv \int d^3\boldsymbol{y}\,O(\boldsymbol{x}+\boldsymbol{y})W_R(\boldsymbol{y})$
Operator at n-th order in perturbation theory	$O^{(n)}$	e.g., $O^{(2)}(\boldsymbol{x},\tau) = [\delta^{(1)}(\boldsymbol{x},\tau)]^2$
Linear matter power spectrum	$P_{\text{L}}(k,\tau)$	$P_{\text{L}}(k,\tau) \equiv \langle\delta^{(1)}(\boldsymbol{k},\tau)\delta^{(1)}(\boldsymbol{k}',\tau)\rangle'$
Variance of linear density field on scale R	$\sigma^2(R)$	$\sigma^2(R) \equiv \int_{\boldsymbol{k}} P_{\text{L}}(k)W_R^2(k)$
Generalized spectral moment[3]	$\sigma_n^2(R)$	$\sigma_n^2(R) \equiv \int_{\boldsymbol{k}} k^{2n}P_{\text{L}}(k)W_R^2(k)$ [Eq. (6.1)]
Critical density (collapse threshold)	$\delta_{\text{cr}} \simeq 1.686$	Eq. (5.9)
Peak significance	ν_c	$\nu_c \equiv \delta_{\text{cr}}/\sigma(R)$
Multiplicity function	$\nu_c f(\nu_c)$	Eq. (5.19)
Bias parameter[4] with respect to operator O	b_O	$\delta_h(\boldsymbol{x},\tau) = \sum_O b_O(\tau)[O](\boldsymbol{x},\tau)$
N-th order LIMD bias parameter	b_N	$b_N \equiv N!\,b_{\delta^N}$ [Eq. (1.4)]
Lagrangian bias parameter	b_O^L	$\delta_h^L(\boldsymbol{q},\tau_0) = \sum_O b_O^L(\tau_0)[O^L](\boldsymbol{q},\tau_0)$
Filter function[5] on scale R	$W_R(x),\ W_R(k)$	See Appendix A.2.
Lagrangian radius of halos	$R(M)$	$R(M) \equiv (3M/4\pi\overline{\rho}_m)^{1/3}$ [Eq. (5.11)]
Large smoothing scale	R_ℓ	$R_\ell \gg R(M)$
Operator smoothed on large scale	O_ℓ	$O_\ell(\boldsymbol{x}) \equiv \int d^3\boldsymbol{y}\,O(\boldsymbol{x}+\boldsymbol{y})W_{R_\ell}(\boldsymbol{y})$

[1] In synchronous-comoving gauge, see Sec. 1.3.
[2] For halos identified at time τ_0; δ_h satisfies the continuity equation by definition (see Sec. 2.3).
[3] Here we allow for $n \in \mathbb{R}$.
[4] This is the physical, renormalized bias, see Sec. 2.10.
[5] Filter functions are normalized such that $\int d^3\boldsymbol{x}\,W_R(x) = 1$ and $\lim_{k\to 0} W_R(k) = 1$.

Table 4: List of symbols and notation used throughout the review.
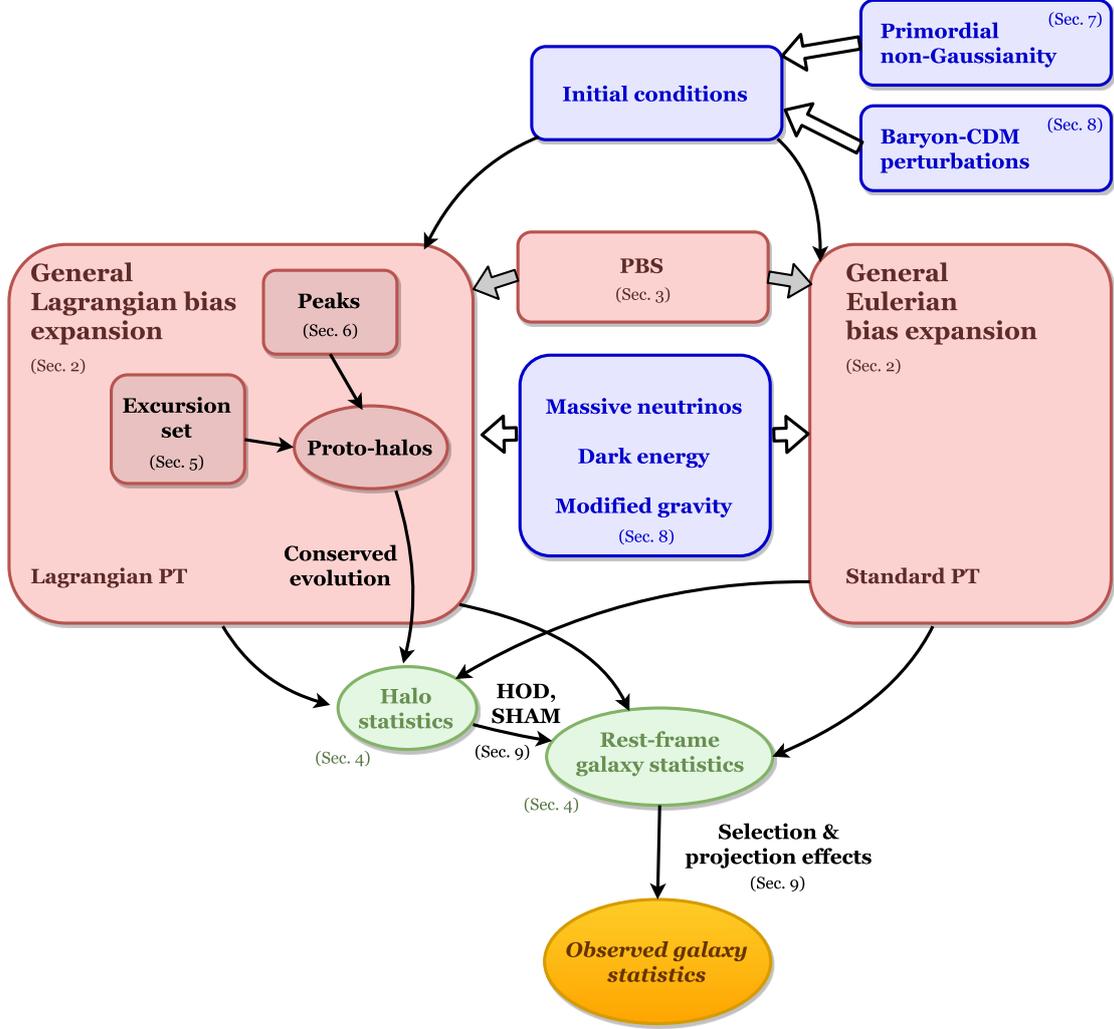
Figure 2: Schematic outline of the theoretical prediction of observed galaxy statistics. Given the statistics of the initial conditions, perturbative bias expansions predict the rest-frame galaxy density as well as that of dark matter halos. This expansion can either be done using Lagrangian (left) or Eulerian frames (right). Crucially, the general bias expansion in either frame is *mathematically equivalent*. The bias expansion is closely connected to the perturbation theory of the matter density field (Lagrangian [LPT] and standard Eulerian [SPT] perturbation theory, respectively). The peak-background split (PBS) informs the bias expansion by relating the bias parameters to responses of the mean tracer abundance. The peak and excursion-set approaches are a special case of the Lagrangian bias expansion, and predict the proto-halo density, which is connected to the statistics of halos at low redshift through conserved evolution. The statistics of halos in turn can be related to those of galaxies through halo occupation distribution (HOD) and subhalo abundance matching (SHAM) approaches. Finally, the connection between rest-frame and observed galaxy statistics involves selection and projection effects (such as redshift-space distortions). Cosmological physics enters in the initial conditions through primordial non-Gaussianity and isocurvature perturbations between baryons and CDM. It also affects the evolution of structure, and consequently the bias expansion, through the effects of massive neutrinos, dark energy, and modified gravity.

1.5   Guide for the reader

In the following, we describe the outline of the review. We recommend that readers begin with Sec. 2, and continue to Sec. 3 and 4. Sec. 5 and following are, for the most part, independent of each other and can be read in arbitrary order. The connection between the different topics is illustrated in the flowchart, Fig. 2 on the previous page. Below is a brief outline of the contents of each section:

Section 2: This section provides a pedagogical introduction to the general perturbative bias expansion. We begin with a simple example, bias that is local in the matter density (LIMD) in Lagrangian space, and then turn to a more realistic study of the gravitational evolution of proto-halos—i.e. the Lagrangian patches which collapse to form virialized halos. These examples set the stage for our derivation of the general perturbative bias expansion.

Section 3: This section presents physical arguments, which have become known under the term *peak-background split*, that can be used to derive the actual values of the bias parameters. We begin with general physical arguments valid for any tracer, and specialize to dark matter halos of a given mass afterwards.

Section 4: We present various methods of measuring bias parameters, for example auto-correlations and cross-correlations with matter, and "scatter plot" methods. In each case, we show rigorously which bias parameters are measured and derive the leading theoretical uncertainties. While we do not make any assumptions about the tracers considered, we do not include redshift-space distortions and other observational complications here (see Section 9), and consequently focus on the application to simulations. We also review measurements of the bias parameters of halos and galaxies, including assembly bias.

Section 5: This section presents the excursion-set approach to calculating the abundance and clustering of dark matter halos. We discuss in detail the so-called cloud-in-cloud problem, choice of filter, and barrier shape, as well as assembly bias.

Section 6: The second physical model of halos, peaks in the Lagrangian density field, is described in this section. We also discuss the recent framework for merging excursion sets and peaks (ESP).

Section 7: While Sec. 2–6 assume Gaussian initial conditions, which is an excellent first-order approximation, we discuss in detail the impact of non-Gaussian initial conditions on galaxy bias here. We derive both the general bias expansion (extension of Sec. 2) and the prediction in the excursion-set and peak approaches (extensions of Sec. 5–6). We also review measurements on simulations with non-Gaussian initial conditions, and present idealized forecasted constraints on primordial non-Gaussianity from galaxy surveys.

Section 8: In this section, we relax the assumption of adiabatic perturbations in a single pressureless (CDM+baryon) fluid made in the previous sections. That is, we consider the impact of massive neutrinos and relative (isocurvature) perturbations between baryons and CDM induced by pre-recombination plasma oscillations, as well as Compton drag after reionization. Finally, we consider the impact on bias of dark energy perturbations and modifications to General Relativity.

Section 9: Strictly speaking, galaxy bias provides a relation between the galaxy (or halo) density and local gravitational observables in the galaxy rest frame. This section describes the selection and projection effects that enter when relating the rest-frame galaxy statistics to observations; this includes redshift-space distortions as well as so-called relativistic and light-cone effects. We also briefly describe empirical models connecting halos and galaxies.

We conclude with an outlook in Sec. 10. The appendices contain:

Appendix A: an overview of statistical field theory.

Appendix B: an introduction to perturbation theory in large-scale structure.

Appendix C: relations between different conventions for the bias parameters.

Appendix D: a brief overview of halo finding algorithms.

2 From local-in-matter-density bias to the general perturbative bias expansion

This section provides a detailed introduction into the general perturbative description of galaxy bias. The final result of this section is fully general, and applies to any large-scale structure tracer. In particular, although bias has been studied extensively for dark matter halos identified in N-body simulations, the general bias expansion is *not restricted to halos*. Bias is a complex problem, which goes significantly beyond the simple well-known LIMD relation $\delta_g = b_1\delta + (b_2/2)\delta^2 + \cdots$. For this reason, this section provides a detailed, step-by-step treatment of the problem. Readers mostly interested in a summary of the relevant equations will find precisely that in Sec. 2.11.

The ultimate goal of bias is to describe the observed statistics of galaxies, such as the galaxy two-point correlation function $\xi_g(r)$, over a certain range of scales, in terms of a finite number of terms (various correlation functions of matter and space-time perturbations) and associated bias parameters. These bias parameters can be understood as coefficients of operators $O(\boldsymbol{x}, \tau)$ in an expansion of the galaxy number density perturbation of the general form

$$\delta_g(\boldsymbol{x}, \tau) = \sum_O b_O(\tau) O(\boldsymbol{x}, \tau) \,. \tag{2.1}$$

Once certain physical assumptions about the background cosmology and the nature of the initial conditions are made, galaxy statistics then contain sufficient information to constrain parameters of the cosmological model even after the free bias parameters have been marginalized over. The goal of the general perturbative bias expansion is to determine which operators have to be included in the sum of Eq. (2.1) in order to describe galaxy clustering down to a certain minimum scale. The relative importance of operators can be ranked by their order in cosmological perturbation theory (see Appendix B for a brief overview, and [72] for a comprehensive review). On scales where perturbation theory is valid, higher-order terms are successively suppressed so that the expansion converges (see Fig. 4 for an illustration in the context of a simple toy model); as stated in Sec. 1, in this review we restrict ourselves to these scales throughout. Thus, *the general perturbative bias expansion consists of an enumeration of all operators that are relevant at a given order in perturbation theory* (and given order in spatial derivatives, as we will see).

The derivation of this general *local* bias expansion, which contains the operators that are relevant in the large-scale limit, is the topic of Sec. 2.1–2.5. We proceed in a pedagogical fashion, beginning with

the simplest example, namely a special case of local-in-matter density (LIMD) bias in Lagrangian space, motivated by the spherical collapse approximation to halo formation (Sec. 2.1). We then turn to a more realistic study of the gravitational evolution of conserved tracers, using two complementary approaches, in Sec. 2.2–2.4. These yield the general local bias expansion up to cubic order, and set the stage for our derivation of the general bias expansion at all orders, in Sec. 2.5. Readers interested only in the full bias expansion can jump to Sec. 2.5 directly.

At this point, we stress again that we define a *local operator* to mean any leading local gravitational observable, which includes the matter density and tidal field, as well as further operators we encounter for the first time in Sec. 2.4. On the other hand, we refer to the restricted expansion which only contains powers of the matter density perturbation δ as *local-in-matter density, or LIMD* bias, a case frequently known in the previous literature under the name "local bias." Note that the LIMD expansions in Eulerian and Lagrangian frames are not equivalent (Sec. 2.3).

In addition to the perturbative order, any physical biased tracer introduces a spatial scale which controls the importance of so-called *higher-derivative operators* in the expansion Eq. (2.1). This spatial scale, which we will denote as R_* in the following, quantifies the *size of the spatial region involved in the process of galaxy formation.* That is, the abundance of galaxies depends on the detailed matter distribution (as well as the other local gravitational observables such as tidal fields) within a region of size R_*. This leads to an expansion in spatial derivatives, i.e. powers of $R_*\partial/\partial x^i$. Such terms are thus known as higher-derivative operators, and are the topic of Sec. 2.6. These higher-derivative operators also take into account all non-gravitational physics influencing the galaxy formation process, such as gas heating and cooling as well as radiative and kinetic feedback processes. In this context, we also discuss *bias of galaxy velocities* (Sec. 2.7); the latter affect the observed galaxy density field through redshift-space distortions (Sec. 9.3.2) and are thus a key ingredient in the modeling of the observed galaxy statistics.

There is one more ingredient we need to consider. The relation between the galaxy density field and the operators O written in Eq. (2.1) is deterministic. In reality, whether a galaxy forms at a given location depends on the initial conditions on very small scales, whose random phases are not determined by the large-scale perturbations included in the bias expansion. This randomness, or *stochasticity*, has to be included separately in the bias expansion, leading to

$$\delta_g(\boldsymbol{x}, \tau) = [\text{Eq. (2.1)}] + \varepsilon(\boldsymbol{x}, \tau) + \sum_O \varepsilon_O(\boldsymbol{x}, \tau) O(\boldsymbol{x}, \tau), \tag{2.2}$$

where the fields ε, ε_O are uncorrelated with the large-scale perturbations described by the operators O and uncorrelated among themselves on large scales. Their contribution to galaxy statistics on large scales can then again be described by a finite set of parameters. These terms are discussed in Sec. 2.8.

Finally, the last two sections of this section deal with the embedding of the general bias expansion within relativistic perturbation theory (Sec. 2.9) and renormalization (Sec. 2.10). Specifically, Sec. 2.10 discusses the rigorous machinery beneath perturbative galaxy bias, and provides the mathematical proof of the completeness of the bias expansion argued on physical grounds in Sec. 2.5. While essential from a theoretical perspective, it can be skipped by readers mostly interested in the observational and measurement aspects of bias. The section concludes with the above mentioned summary in Sec. 2.11.

2.1 A toy model: LIMD in Lagrangian space

Let us begin our discussion of galaxy bias with a simplified example. We assume that dark matter halos, within which observed galaxies reside, simply correspond to overdense regions (above a threshold) in Lagrangian space, that is, in the initial matter density field extrapolated to a desired reference time using the linear growth. We denote this field, often referred to as linear density field, by $\delta^{(1)}$, and assume it to be Gaussian (see Sec. 7 for the non-Gaussian case). This toy model was first studied quantitatively in [12]. In order to trace halos identified at low redshift back to the initial conditions, we imagine following the trajectory of the constituent particles of a given halo back to the initial time $\tau = 0$. The region occupied by these particles is referred to as the *proto-halo*. Since the initial density field is arbitrarily close to uniform, the proto-halo volume of a halo of mass M is $M/\overline{\rho}_m$, from which we can define the *Lagrangian radius* of

the halo via $R(M) = (3M/4\pi\bar{\rho}_m)^{1/3}$. Hence, in order to define "regions above threshold" which eventually collapse to form halos, we filter the initial (linear) density field on the scale $R(M)$, denoting this as $\delta_R^{(1)}$ (Fig. 3). The shape of the filter is not relevant for this discussion; we list popular filters in Appendix A.2. The comoving Lagrangian number density of proto-halos is then defined as

$$n_{\mathrm{thr}}^L(\boldsymbol{q}) \equiv \Theta_H(\delta_R^{(1)}(\boldsymbol{q}) - \delta_{\mathrm{cr}}) \,, \tag{2.3}$$

where δ_{cr} is a fixed density threshold and Θ_H denotes the Heaviside step function. Note that the "density" n_{thr}^L defined here corresponds, up to normalizing factors which we neglect here, to the mass-weighted cumulative number density of halos above mass M (see for example [98] and Sec. 5). In the sketch Fig. 3, the proto-halo number density is unity whenever the blue solid line crosses the threshold indicated by the horizontal line. As this toy model describes "thresholded regions," it is often referred to as "thresholding."

The statistics of the Gaussian field $\delta_R^{(1)}$ are completely described by the two-point correlation function (Appendix A)

$$\xi_{\mathrm{L},R}(r) = \langle \delta_R^{(1)}(\boldsymbol{q}) \delta_R^{(1)}(\boldsymbol{q} + \boldsymbol{r}) \rangle \,, \tag{2.4}$$

where $\xi_{\mathrm{L},R}(r)$ is the filtered version of the linear matter correlation function $\xi_{\mathrm{L}}(r) = \lim_{R\to 0} \xi_{\mathrm{L},R}(r)$ [the Fourier transform of the linear power spectrum $P_{\mathrm{L}}(k)$, Eq. (4.8)], and $\xi_{\mathrm{L},R}(0) = \sigma^2(R)$ is the variance of the filtered density field. The mean "number density" of proto-halos is obtained by taking the expectation value of Eq. (2.3),

$$\langle n_{\mathrm{thr}}^L(\boldsymbol{q}) \rangle = \frac{1}{\sqrt{2\pi}} \int_{\nu_c}^{\infty} e^{-\nu^2/2} d\nu = \frac{1}{2} \operatorname{erfc}[\nu_c/\sqrt{2}] \equiv p_1 \,, \quad \nu_c \equiv \frac{\delta_{\mathrm{cr}}}{\sigma(R)} \,. \tag{2.5}$$

This shows that, for a high threshold $\nu_c > 1$, proto-halos become exponentially rare. Note that any normalizing factors that we neglect here do not affect the calculation of the bias, which we turn to next.

The Lagrangian two-point function of equal-mass proto-halos at separation r is given by the ratio of the pair probability $p_2(\boldsymbol{q}, \boldsymbol{q} + \boldsymbol{r})$ of finding two proto-halos at positions $\boldsymbol{q}$ and $\boldsymbol{q} + \boldsymbol{r}$, divided by the 1-point probability squared $(p_1)^2$ [71, 12]:

$$1 + \xi_h^L(r) = \frac{p_2(\boldsymbol{q}, \boldsymbol{q} + \boldsymbol{r})}{p_1^2} = \sqrt{\frac{2}{\pi}} \left[\operatorname{erfc}(\nu_c/\sqrt{2})\right]^{-2} \int_{\nu_c}^{\infty} e^{-\nu^2/2} \operatorname{erfc}\left[\frac{\nu_c - \nu\hat{\xi}(r)}{\sqrt{2\{1 - \hat{\xi}^2(r)\}}}\right] d\nu \,. \tag{2.6}$$

where $\hat{\xi}(r) \equiv \xi_{\mathrm{L},R}(r)/\sigma^2(R)$. This relation follows directly from integrating over the Gaussian likelihood of the density field $\delta_R^{(1)}$. Again, $\xi_h^L(r)$ is the two-point function of proto-halos in the initial conditions, as emphasized by the superscript L, extrapolated to $z = 0$ using linear theory. If $\xi_{\mathrm{L},R}(r)$ is sufficiently small, we can expand Eq. (2.6) in a series,

$$\xi_h^L(r) = \sum_{N=1}^{\infty} \frac{1}{N!} (b_N^L)^2 \left[\xi_{\mathrm{L},R}(r)\right]^N \,, \tag{2.7}$$

where the r-independent coefficients b_N^L, the *bias parameters*, are given by [12, 99, 100]

$$b_N^L = \sqrt{\frac{2}{\pi}} \left[\operatorname{erfc}\left(\frac{\nu_c}{\sqrt{2}}\right)\right]^{-1} \frac{e^{-\nu_c^2/2}}{\sigma^N(R)} H_{N-1}(\nu_c) \stackrel{\nu_c \gg 1}{=} \frac{\nu_c^N}{\sigma^N(R)} + \mathcal{O}\left(\nu_c^{N-1}\right) \,. \tag{2.8}$$

The superscript L on the bias parameters indicates that they refer to the Lagrangian density field. Note that b_1^L is positive definite, which is not true for general tracers. Moreover, in the limit of low mass $R \to 0$, $\sigma(R) \to \infty$ and b_1^L approaches zero; this is because the simple thresholding procedure Eq. (2.3) does not correctly describe low-mass objects due to the "cloud-in-cloud" problem. We will return to this in Sec. 5.

Since in the real Universe, matter density fluctuations are small on large scales, we have $\xi_{\mathrm{L}}(r) \to 0$ as
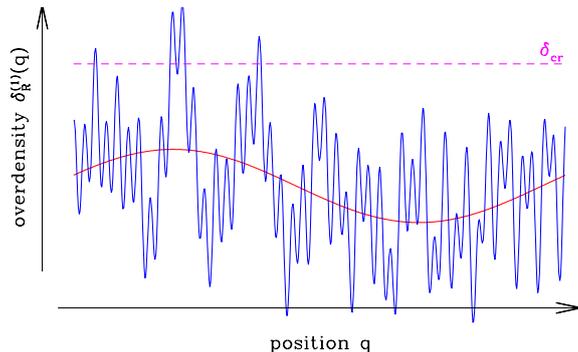
Figure 3: Illustration of the toy model of Sec. 2.1. The solid blue line shows the smoothed density field $\delta_R^{(1)}(\boldsymbol{q})$, while the red line indicates a long-wavelength perturbation. The dashed horizontal line marks the threshold overdensity $\delta_{\rm cr}$.

$r \to \infty$. This means that, as long as we are interested in the proto-halo correlation function on large scales, it is sufficient to keep only the first few terms in the expansion Eq. (2.7). In particular, the leading term is

$$\xi_h^L(r) = (b_1^L)^2 \xi_{{\rm L},R}(r) \,, \tag{2.9}$$

which is what one obtains from the simple bias relation Eq. (1.1). Thus, the proto-halo correlation function is directly proportional to the matter correlation function on large scales, i.e. both have the same r-dependence. However, ξ_h^L is enhanced by the factor $(b_1^L)^2$, which becomes large for rare, high-mass proto-halos [$\nu_c \gg 1$; Eq. (2.8)]. The picture in Fig. 3 delivers an intuitive explanation for why this happens. When adding a large-scale density perturbation (red solid line) to the matter density field, the abundance of rare regions above threshold responds much more sensitively than the matter density itself, which, as we will see in Sec. 3, is an alternative, exactly equivalent definition of bias. This larger response in turn leads to enhanced clustering on large scales [12, 13]. The latter effect, shown in Fig. 4, was the original motivation of [12], who showed that the correlation function of rare objects such as massive galaxy clusters is enhanced relative to that of the underlying matter on *all* scales.

Fig. 4 shows that the series expansion Eq. (2.7) converges rapidly to the exact result on scales $\gtrsim 10\,h^{-1}\,{\rm Mpc}$. This approximately corresponds to the scale where the matter correlation function becomes of order 1. Thus, separations $r \gtrsim 10\,h^{-1}\,{\rm Mpc}$ are amenable to a perturbative description in the context of this toy model. On the other hand, on smaller scales, higher-order terms are no longer smaller than lower-order terms. This means that a general perturbative expansion is not guaranteed to converge to the correct result. While for the toy model considered here we have an exact result, this is not the case for real galaxies, so that the restriction to perturbative scales is the only way to guarantee a theoretical error that is under rigorous control. Note that the scale where the expansion Eq. (2.7) breaks down is related to the *nonlinear scale* $R_{\rm NL}$ (see Appendix B; $R_{\rm NL} \sim 10-20\,h^{-1}\,{\rm Mpc}$ at redshift zero), at which the perturbative description of the nonlinear matter density field itself breaks down. This scale becomes smaller at higher redshifts; hence, the range of scales accessible to perturbation theory is larger at high redshifts [101].

The expression for the proto-halo two-point correlation function, Eq. (2.7), is not specific to the ansatz Eq. (2.3) we started from. Indeed, if we write $\delta_h(\boldsymbol{q}) = n_{\rm thr}^L(\boldsymbol{q})/\langle n_{\rm thr}^L \rangle - 1$ as a formal series expansion in $\delta_R^{(1)}$,

$$\delta_h^L(\boldsymbol{q}) = b_1^L \delta_R^{(1)}(\boldsymbol{q}) + \frac{1}{2} b_2^L \left([\delta_R^{(1)}(\boldsymbol{q})]^2 - \sigma^2(R) \right) + \cdots \,, \tag{2.10}$$

then Eq. (2.7) is obtained directly when discarding all terms that involve zero-lag correlators (terms containing factors of $\langle [\delta_R^{(1)}(\boldsymbol{q})]^n \rangle$; see Appendix A). One can similarly derive all higher N-point functions and cumulants of $\delta_h(\boldsymbol{q})$ [102, 100]. Historically, Eq. (2.10) is known as the *local bias expansion* [17], since $\delta_h(\boldsymbol{q})$ is a local function of the filtered matter density field. As we have discussed in the beginning of Sec. 2, Eq. (2.10) only contains a subset of the local bias terms according to the definition used in this review. To

19
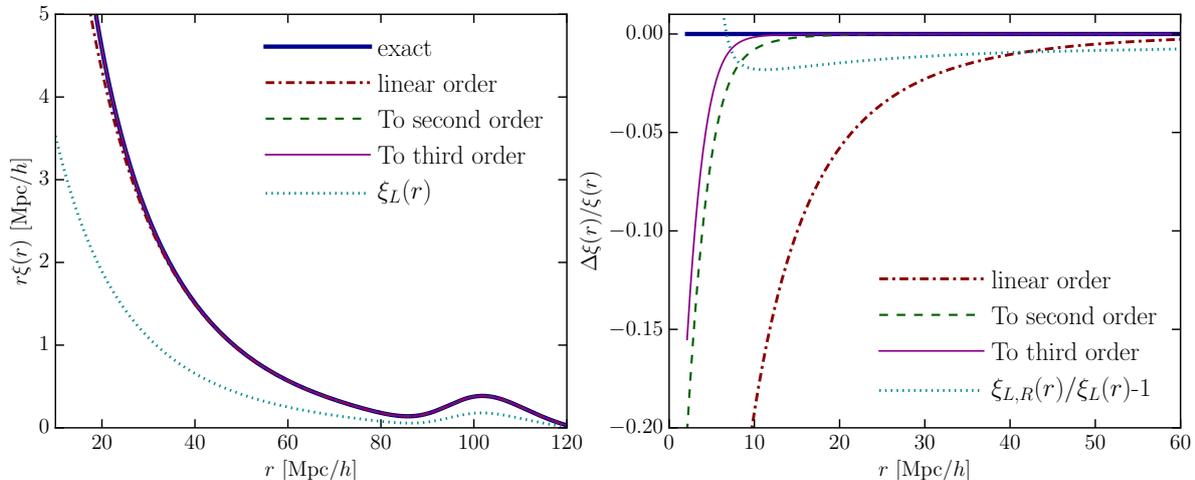
Figure 4: Correlation function in Lagrangian space of thresholded regions in the initial density field extrapolated to $z = 0$. The smoothing scale $R = 4.21\,h^{-1}\,\mathrm{Mpc}$ (mass scale $2.5 \cdot 10^{13}\,h^{-1}\,M_\odot$) is chosen to correspond to $b_1^L = 1.5$ via Eq. (2.8). *Left panel:* We show the exact result Eq. (2.6), multiplied by r to better show the large-scale behavior, as well as the series expansion Eq. (2.7) truncated at different orders. For comparison, the cyan dotted line shows the linear, unfiltered matter correlation function. The bump at $r \approx 100\,h^{-1}\,\mathrm{Mpc}$ is the BAO feature. *Right panel:* Relative deviation of the truncated series expansion from the exact result. The dotted line shows the relative deviation of the unfiltered linear contribution $(b_1^L)^2\xi_{\mathrm{L}}$ from the filtered contribution $(b_1^L)^2\xi_{\mathrm{L},R}$.

be specific, we refer to Eq. (2.10) as *local-in-matter-density (LIMD) bias*. Even though not complete, as we will see, the LIMD expansion is sufficiently general to describe an ansatz of the type Eq. (2.3) for any function $n_{\mathrm{thr}}^L(\boldsymbol{q}) = F\left(\delta_R^{(1)}[\boldsymbol{q}]\right)$ [103, 104]. Furthermore, while Eq. (2.7) is a prediction specific to Gaussian density fields, the LIMD ansatz allows statements about the statistics of n_{thr}^L to be made even for general density statistics [104], and for density fields whose N-point functions obey hierarchical scaling laws [105, 106].

This fact, combined with the fact that the series expansion Eq. (2.7) converges to the true result on large scales (Fig. 4), shows the power of the perturbative bias expansion: it can capture a very general class of tracers (in the case of Eq. (2.10), a tracer whose abundance is described by an arbitrary local function of the initial matter density field) via a small set of free parameters, while guaranteed to converge to the correct result on sufficiently large scales.

Moreover, the proto-halo number density does not have to be an *exactly* local function of $\delta_R^{(1)}$ in order for the LIMD expansion Eq. (2.10) to be effective, as already noticed by [17]: we only need to require that the nonlocality is restricted to scales much smaller than the scale r at which the correlation function is calculated. This is illustrated in the right panel of Fig. 4, which shows the ratio of the correlation function of the filtered matter density field, $\xi_{\mathrm{L},R}(r)$, to that of the unfiltered field $\xi_{\mathrm{L}}(r)$. Clearly, the smoothed density field is nonlocally related to the un-smoothed density field, however this manifests itself in the correlation function only on scales smaller than a few times the smoothing radius. This exemplifies that, as shown in [104, 106, 107], in order to change the shape of the galaxy correlation function $\xi_h^L(r)$ relative to the matter correlation function $\xi_{\mathrm{L}}(r)$ on *large* scales, say on hundreds of Mpc, galaxy formation has to be highly nonlocal, i.e. nonlocal on a scale of order a hundred Mpc. The only exception to this rule, which however does not invalidate the point, is that there can be an effect on large scales if there are sharp features in the matter correlation function, for example the BAO feature at $r \approx 100\,h^{-1}$ Mpc. As we will see in Sec. 2.6, this effect can be taken into account to high accuracy with only one additional bias parameter.

Notice that we did not include a zeroth order bias b_0^L, i.e. a constant offset, in the bias relation Eq. (2.10) because of the requirement $\langle \delta_h \rangle = 0$. This is only true for a *deterministic bias relation* as assumed in

20

Eq. (2.10). In reality, both the presence of random small-scale fluctuations (which affect the abundance of halos and galaxies, but are not correlated with the long-wavelength fluctuations) and the fact that halos are a discrete sample of the underlying density field demand that we allow for *stochastic fields* in the bias expansion. At lowest order, this adds a stochastic variable $\varepsilon(\boldsymbol{q})$ with vanishing mean to Eq. (2.10). We begin including stochastic contributions from the next section.

So far, we have provided a description of proto-halo statistics in Lagrangian space. Naturally, we need to ask how to translate these results to the observationally relevant statistics of the evolved halo field at lower redshifts (Eulerian space). Let us again consider a simplified setup, namely a large spherical region of radius R_ℓ; specifically, following [15], we assume that R_ℓ is much larger than the typical separation between the halos considered. We allow this region to have an over- or underdensity $\delta_\ell(\tau)$, but assume the Universe to be unperturbed otherwise on large scales. Then, we expand the fractional overdensity $\delta_{h,\ell}$ of halos within this region with respect to the global mean at a given redshift as

$$\delta_{h,\ell}(\tau) = \frac{\bar{N}_\ell}{\overline{n}_h V_\ell(\tau)} - 1 = b_1^E(\tau)\delta_\ell(\tau) + \frac{1}{2}b_2^E(\tau)\delta_\ell^2 + \cdots , \qquad (2.11)$$

where $\bar{N}_\ell$ is the number of virialized halos in the region, and $\overline{n}_h$ is the global mean abundance (recall that all densities are comoving and thus unaffected by the dilution due to the expansion of the Universe), while $V_\ell(\tau)$ is the comoving volume of the spherical region. Analogously to Eq. (2.10), and following [15], Eq. (2.11) defines the Eulerian bias parameters b_N^E on the right-hand side. Note that these bias parameters formally become independent of the size of the region only in the limit $R_\ell \to \infty$. For simplicity however, we shall omit this dependence here. This subtlety will be addressed in Sec. 4.2.

We now want to relate the Eulerian bias parameters to the Lagrangian bias parameters b_N^L, by using the number conservation of halos. Since by definition we refer to proto-halos in the initial conditions which are the exact progenitors of halos identified at time τ, their number is conserved. The evolution of the matter overdensity δ_ℓ or equivalently $V_\ell(\tau)$ on the other hand follows spherical collapse [108] (see Sec. 5.2.1), which is uniquely determined from the initial overdensity $\delta_\ell^{(1)}$ of the region. The spherical collapse evolution is valid up to shell crossing, i.e. as long as $R_\ell > 0$; up to that point, the mass in each shell is conserved. This implies that the Lagrangian density perturbation of halos is given by $1 + \delta_{h,\ell}^L = \bar{N}_\ell/(\overline{n}_h V_\ell^L)$ where V_ℓ^L is the Lagrangian volume of the region, which is related to the Eulerian volume V_ℓ by

$$V_\ell^L = (1 + \delta_\ell)V_\ell . \qquad (2.12)$$

Therefore, we obtain the following relation between $\delta_{h,\ell}(\tau)$ and $\delta_{h,\ell}^L$:

$$1 + \delta_{h,\ell}(\tau) = \left[1 + \delta_{h,\ell}^L\right]\left[1 + \delta_\ell(\tau)\right] = 1 + \left(1 + b_1^L\right)\delta_\ell + \mathcal{O}(\delta_\ell^2) , \qquad (2.13)$$

where on the right-hand side we have expanded the result in powers of δ_ℓ using Eq. (2.10), filtered on the scale R_ℓ. Comparison with Eq. (2.11) then shows that $b_1^E = 1 + b_1^L$. Halos which have vanishing bias ($b_1^L = 0$) in the initial conditions, so that $\delta_h^L = 0$ and they are uniformly distributed at the initial epoch, will hence remain unbiased relative to the matter at any later time τ.[5]

The calculation of the Eulerian bias parameters can be extended to all orders within this picture, since the mapping between $\delta_\ell(\tau)$ and $\delta_\ell^{(1)}$ can be calculated for any given cosmology by integrating the spherical collapse equations. Specifically, one obtains the series expansion [109, 110, 111]

$$\delta_\ell(\tau) = \sum_{k=1}^{\infty} a_k \left[\delta_\ell^{(1)}(\tau)\right]^k ; \qquad a_1 = 1 , \quad a_2 = \frac{17}{21} , \quad a_3 = \frac{341}{567} , \quad a_4 = \frac{55805}{130977} , \quad \cdots$$

$$\delta_\ell^{(1)}(\tau) = \sum_{k=1}^{\infty} a_k^{\mathrm{I}} \left[\delta_\ell(\tau)\right]^k ; \qquad a_1^{\mathrm{I}} = 1 , \quad a_2^{\mathrm{I}} = -\frac{17}{21} , \quad a_3^{\mathrm{I}} = \frac{2815}{3969} , \quad a_4^{\mathrm{I}} = -\frac{590725}{916839} , \quad \cdots . \qquad (2.14)$$

---

[5]Here, we have implicitly assumed that proto-halos comove with the matter fluid. We will discuss this in more detail in Sec. 2.2 and Sec. 2.7.
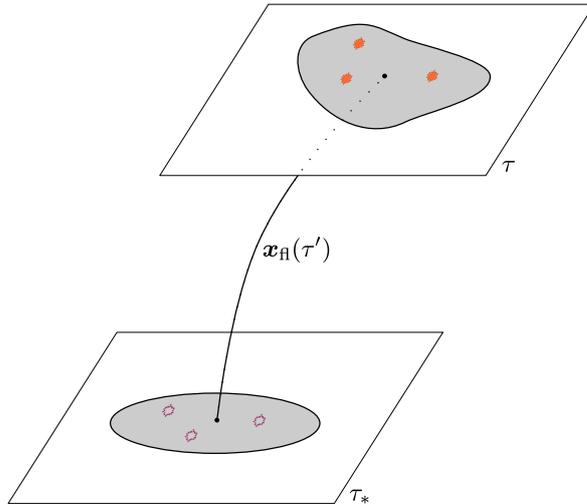
Figure 5: Sketch of the setup considered in Sec. 2.2–2.4. Galaxies form instantaneously at $\tau = \tau_*$, where they are described by an initial bias relation (lower slice), and are comoving with the matter. After this, the evolution is governed by number conservation and the comoving assumption up until time τ (upper slice), where they are assumed to be observed. The grey region denotes a Lagrangian volume encompassing three galaxies which gets deformed by nonlinear gravitational evolution. Since galaxies comove with matter, their density is similarly affected.

Here, the coefficients are only strictly valid for Einstein-de Sitter (EdS), where $\delta_\ell^{(1)}(\tau) \propto a = 1/(1+z)$. However, they are also highly accurate for other cosmologies [112], as long as we remain sufficiently far from the collapse of the spherical region of size R_ℓ. As pointed out in [113, 112], the coefficients $a_3^{\mathrm{I}}, a_4^{\mathrm{I}}$ in the inverse relation were given incorrectly in [110]. Substitution of Eq. (2.14) into Eq. (2.13) then yields a unique prediction for the *evolved* halo correlation function,

$$\xi_h(r, \tau) = \sum_{N=1}^{\infty} \frac{1}{N!} [b_N^E(\tau)]^2 \left[\xi_R(r, \tau)\right]^N , \tag{2.15}$$

with Eulerian bias parameters b_N^E given by

$$b_1^E(\tau) = 1 + b_1^L(\tau) \tag{2.16}$$
$$b_2^E(\tau) = 2\left(1 + a_2^{\mathrm{I}}\right) b_1^L(\tau) + b_2^L(\tau)$$
$$b_3^E(\tau) = 6\left(a_2^{\mathrm{I}} + a_3^{\mathrm{I}}\right) b_1^L(\tau) + 3\left(1 + 2a_2^{\mathrm{I}}\right) b_2^L(\tau) + b_3^L(\tau) .$$

These important relations were first derived by [15, 110].

We have thus achieved our goal of deriving the bias parameters describing statistics of halos in the evolved density field given a Lagrangian LIMD bias expansion Eq. (2.10). Clearly, in the context of a spherically symmetric density perturbation, a deterministic LIMD bias relation in the initial conditions is transformed into a similar relation in the evolved distribution (Eulerian LIMD). We will see in the next section that this is a consequence of the spherical symmetry assumed in this derivation: in general, gravitational evolution starting from a LIMD bias expansion generates additional terms in the Eulerian bias relation which are beyond LIMD (yet still local). In essence, we have neglected tidal fields which play a similarly important role as density perturbations. The next subsections will look more closely at the gravitational evolution of proto-halos or -galaxies.

2.2   Gravitational evolution: general considerations

We now investigate in more detail the interplay of gravitational evolution and bias. Specifically, we consider the toy example of a sample of galaxies that formed instantaneously at a fixed time τ_*, that are

comoving with matter, and whose number is conserved afterwards (Fig. 5). The discussion here is not specific to dark matter halos, and is applicable to the statistics of any conserved tracer of the large-scale structure. Hence, we will use the term "galaxies" instead of "halos" in the following. Given a bias relation for δ_g^* at time τ_*, involving the matter density and other locally observable quantities, the goal of this section is to derive how this bias relation evolves under gravity while conserving the number of tracers to some later time $\tau > \tau_*$. We will then be able to express the bias parameters b_n at time τ in terms of the bias parameters b_n^* at the "formation time" τ_*, generalizing Eq. (2.16) which is specific to a spherical region and LIMD bias in Lagrangian space.

With this ansatz, we essentially describe a galaxy sample undergoing "passive evolution." However, beyond being a mere toy model, this ansatz can be considered a *Green's function* for galaxies in the sense that a realistic galaxy sample (consisting of a broad range of formation times) can be described as a superposition of many samples with instantaneous formation at various epochs τ_*. Thus, if our bias expansion correctly describes a sample formed at an arbitrary time τ_* and evolved to time τ, then this bias expansion is sufficiently general to describe any galaxy sample provided all bias parameters are allowed to be free.

Similarly to the matter density field itself, there is no closed solution for the nonlinear evolution of a conserved galaxy density field. Instead, we adopt a perturbation theory approach which successfully describes the evolution on sufficiently large scales. Unlike the discussion in the previous section, we do not assume spherical symmetry; indeed, the lack of this symmetry is the reason why we have to add additional local terms to our bias expansion beyond the LIMD case, Eq. (2.11). Moreover, we will focus on scales much larger than the Lagrangian radius of halos throughout, and hence drop the subscript R denoting the filtering in the following. In Sec. 2.1, we found this to be accurate empirically, however this will be justified rigorously in Sec. 2.6.

The equation of motion governing the evolution of a conserved tracer is the continuity equation [114, 115, 116, 117]:

$$\frac{\mathrm{D}}{\mathrm{D}\tau}\delta_g = -\theta(1 + \delta_g)\,, \tag{2.17}$$

where

$$\frac{\mathrm{D}}{\mathrm{D}\tau} \equiv \frac{\partial}{\partial\tau} + v^i\frac{\partial}{\partial x^i} \tag{2.18}$$

is the convective (or Lagrangian) time derivative, v^i is the peculiar velocity of the cosmic matter fluid, and $\theta = \partial_{x,i}v^i$ is the velocity divergence.

In Eq. (2.17), we have implicitly assumed that there is *no velocity bias* of the galaxies with respect to matter, that is, galaxies comove with the matter fluid. A detailed discussion of velocity bias, and the proof that it is absent on large scales, is the topic of Sec. 2.7, but we elaborate briefly here. Consider a sufficiently large patch surrounding a given galaxy, say several times the Lagrangian radius of the galaxy's host halo (e.g. the grey region in Fig. 5). We then determine the center-of-mass velocity of this patch. Any peculiar velocity of the galaxy with respect to the center-of-mass has to be due either to the detailed matter distribution within the patch, or non-gravitational forces such as momentum transfer due to baryonic feedback processes or radiation pressure. That is, for a sufficiently large patch, the galaxy's relative velocity with respect to the matter is only determined by physics within the patch, which is equivalent to stating that on large scales galaxies comove with matter. In the absence of non-gravitational forces, this is already required by the equivalence principle: test bodies, regardless of their nature—be they a galaxy, black hole, or dark matter particle—fall at the same rate in an external (large-scale) gravitational field.

In order to evolve δ_g, we also need the perturbative solution of δ and v^i. In the context of standard perturbation theory (SPT; see Appendix B for a brief overview), where dark matter is treated as a pressureless ideal fluid, the fluid velocity $\boldsymbol{v}$ is curl-free and can thus be written as $v^i = (\partial^i/\nabla^2)\theta$. Then, the density perturbation δ and velocity divergence θ obey the continuity and Euler equations,

$$\frac{\mathrm{D}}{\mathrm{D}\tau}\delta = -\theta(1 + \delta), \tag{2.19}$$

$$\frac{\mathrm{D}}{\mathrm{D}\tau}\theta = -\mathcal{H}\theta - (\partial_x^i v^j)^2 - \frac{3}{2}\Omega_m\mathcal{H}^2\delta\,. \tag{2.20}$$

Quantity	Eulerian expression	Lagrangian expression
Lagrangian displacement	$s(\boldsymbol{q},\tau)$	
Deformation tensor	$M_i{}^j(\boldsymbol{q},\tau) \equiv \partial_{q,i} s^j(\boldsymbol{q},\tau)$	
Spatial coordinate	$\boldsymbol{x}$	$\boldsymbol{q}$
Spatial derivative	$\partial_{x,i}$	$\partial_{q,i} = (\delta_i{}^j + M_i{}^j)\partial_{x,j}$
Fluid trajectory	$\boldsymbol{x}_{\mathrm{fl}}(\boldsymbol{q},\tau) \equiv \boldsymbol{q} + \boldsymbol{s}(\boldsymbol{q},\tau)$	$\boldsymbol{q}$
Fluid velocity	$\boldsymbol{v} = d\boldsymbol{x}_{\mathrm{fl}}/d\tau = \partial_\tau \boldsymbol{s}(\boldsymbol{q},\tau)$	0
Convective derivative	$\mathrm{D}/\mathrm{D}\tau = \partial_\tau + v^i \partial_{x,i}$	$\mathrm{D}/\mathrm{D}\tau = \partial_\tau$
Matter overdensity	$\delta(\boldsymbol{x},\tau) = \|\boldsymbol{1} + \boldsymbol{M}(\boldsymbol{q},\tau)\|^{-1} - 1$	0
Conserved tracer overdensity[1]	$\delta_g(\boldsymbol{x}_{\mathrm{fl}}[\boldsymbol{q},\tau],\tau) = (1 + \delta(\boldsymbol{x}_{\mathrm{fl}}[\boldsymbol{q},\tau],\tau))$ $\times \left(1 + \delta_g^L(\boldsymbol{q})\right) - 1$	$\delta_g^L(\boldsymbol{q}) = \dfrac{1 + \delta_g[\boldsymbol{x}_{\mathrm{fl}}(\boldsymbol{q},\tau),\tau]}{1 + \delta[\boldsymbol{x}_{\mathrm{fl}}(\boldsymbol{q},\tau),\tau]} - 1$

[1] This only holds for a tracer whose number is conserved from time $\tau = 0$ to time τ (e.g., proto-halos of halos identified at time τ).

Table 5: Summary of relations between Eulerian and Lagrangian space.

There are essentially two equivalent approaches to deriving the mapping from $\delta_g(\tau_*)$ to $\delta_g(\tau)$. First, one can directly integrate Eq. (2.17) along the fluid trajectory, making use of the known perturbative solutions of the SPT equations. Second, one can iteratively solve the set of equations Eqs. (2.17)–(2.20) to obtain a joint perturbative solution for δ_g and δ. We will outline each approach in turn in the following sections, as they illustrate different aspects of gravitational evolution and bias.

Apart from the matter density field δ, in the following we will also encounter the tidal field, which we define through the scaled dimensionless quantity K_{ij},

$$K_{ij} \equiv \mathcal{D}_{ij}\delta = \frac{2}{3\Omega_m \mathcal{H}^2}\partial_i\partial_j\Phi - \frac{1}{3}\delta_{ij}\delta\,; \quad \mathcal{D}_{ij} \equiv \left(\frac{\partial_i\partial_j}{\nabla^2} - \frac{1}{3}\delta_{ij}\right), \tag{2.21}$$

where the second equality follows from the Poisson equation [Eq. (B.3)]. Equivalently, in Fourier space

$$K_{ij}(\boldsymbol{k}) = \left[\frac{k_i k_j}{k^2} - \frac{1}{3}\delta_{ij}\right]\delta(\boldsymbol{k})\,. \tag{2.22}$$

Furthermore, we will denote quantities at n-th order in perturbation theory with a superscript (n). For example, the second-order density field is denoted as $\delta^{(2)}$.

2.3 Evolution from the continuity equation

We now proceed to solve the continuity equation Eq. (2.17) for our conserved galaxies as follows [114, 115, 117, 118]. We first divide Eq. (2.17) and Eq. (2.19) by $1 + \delta_g$ and $1 + \delta$ respectively. This yields

$$\frac{1}{1+\delta_g}\frac{\mathrm{D}}{\mathrm{D}\tau}\delta_g = -\theta = \frac{1}{1+\delta}\frac{\mathrm{D}}{\mathrm{D}\tau}\delta\,. \tag{2.23}$$

For convenience, we now write δ_g and δ in terms of Lagrangian coordinates $\boldsymbol{x} \to \boldsymbol{q}$, in which case the convective time derivatives become partial derivatives with respect to τ; see Tab. 5, where relations between various quantities in the Eulerian and Lagrangian frames are summarized. Then, the integral along the fluid trajectory becomes trivial, and we obtain

$$\ln[1 + \delta_g(\boldsymbol{x}_{\mathrm{fl}}(\tau),\tau)] = \ln[1 + \delta(\boldsymbol{x}_{\mathrm{fl}}(\tau),\tau)] + \ln\left[\frac{1 + \delta_g(\boldsymbol{x}_*,\tau_*)}{1 + \delta(\boldsymbol{x}_*,\tau_*)}\right] \tag{2.24}$$

for $\tau \geq \tau_*$. We have fixed the integration constant by introducing the galaxy overdensity $\delta_g(\boldsymbol{x}_*, \tau_*)$ on the formation time slice. Here, $\boldsymbol{x}_{\mathrm{fl}}(\tau)$ denotes the Eulerian coordinate of the fluid trajectory corresponding to a fixed Lagrangian position $\boldsymbol{q} = \boldsymbol{x}_{\mathrm{fl}}(\tau = 0)$, and $\boldsymbol{x}_* \equiv \boldsymbol{x}_{\mathrm{fl}}(\tau_*)$ denotes the position on the formation time slice. By introducing the displacement $\boldsymbol{s}$ through $\boldsymbol{x}_{\mathrm{fl}}(\tau) = \boldsymbol{q} + \boldsymbol{s}(\boldsymbol{q}, \tau)$, we can write the equation of motion governing the fluid trajectory as

$$\left(\frac{\partial^2}{\partial \tau^2} + \mathcal{H}\frac{\partial}{\partial \tau}\right) \boldsymbol{s}(\boldsymbol{q}, \tau) = -\boldsymbol{\nabla}\Phi\Big(\boldsymbol{q} + \boldsymbol{s}(\boldsymbol{q}, \tau), \tau\Big), \tag{2.25}$$

with initial condition $\boldsymbol{s}(\boldsymbol{q}, \tau = 0) = 0$. This equation was first derived by [119]. At linear order, we can neglect the appearance of $\boldsymbol{s}$ in the argument of $\boldsymbol{\nabla}\Phi(\boldsymbol{x}_{\mathrm{fl}})$, and this yields Eq. (B.16) in Appendix B.

Thus, if we know the bias relation at τ_*, Eq. (2.24) supplies us with the bias relation at all later times. This has been derived in [79, 118] (see also [120]). It is important to note that conserved evolution relates $\delta_g(\boldsymbol{x}, \tau)$ and $\delta_g(\boldsymbol{x}_*, \tau_*)$ at two different times along *the same fluid trajectory* (solid line in Fig. 5). Mathematically, this is due to the convective derivatives in Eq. (2.23); physically, it states that galaxies and matter fall at the same rate in the large-scale gravitational field, and hence co-evolve along the same trajectories in the absence of an initial velocity bias.

We can make Eq. (2.24) even more clear by rewriting it as

$$1 + \delta_g\Big|_\tau = \frac{1 + \delta|_\tau}{1 + \delta|_{\tau_*}}(1 + \delta_g|_{\tau_*}), \tag{2.26}$$

where a vertical bar $|_\tau$ denotes a quantity evaluated at τ on a fixed fluid trajectory. This simply states that the density ratio of two conserved, initially comoving fluids remains constant under gravitational evolution, as required by the equivalence principle; we will encounter this again in Sec. 8.2 when considering the different initial conditions for baryons and CDM set in the early Universe. Moreover, letting $\tau_* \to 0$, so that $\delta|_{\tau_*} \to 0$, we recover the well-known relation between Eulerian and Lagrangian densities of a conserved tracer (see Tab. 5), derived for the special case of a spherical perturbation in Eq. (2.13).

In the following, we will solve Eq. (2.26) to second order in perturbations, that is, up to quadratic terms in $\delta^{(1)}$ and $K_{ij}^{(1)}$. Denoting second-order terms with a superscript (2), we easily obtain

$$1 + \delta_g^{(1)}(\boldsymbol{x}, \tau) + \delta_g^{(2)}(\boldsymbol{x}, \tau) = 1 + \delta^{(1)} - \delta_*^{(1)} + \delta_{g*}^{(1)}$$
$$+ \delta^{(2)} - \delta_*^{(2)} + \delta_{g*}^{(2)} + [\delta_*^{(1)}]^2 - \delta^{(1)}\delta_*^{(1)} + \left[\delta^{(1)} - \delta_*^{(1)}\right]\delta_{g*}^{(1)}, \tag{2.27}$$

where $f_* \equiv f(\boldsymbol{x}_*, \tau_*)$, while quantities without a subscript $*$ are evaluated at $(\boldsymbol{x}, \tau)$. Here, we have separated linear and second order terms into the first and second line. Note however that the distinction between $\boldsymbol{x}_* = \boldsymbol{x}_{\mathrm{fl}}(\tau_*)$ and $\boldsymbol{x}$ is itself first order in perturbations. This is simply because the fluid trajectory in an unperturbed FRW spacetime is $\boldsymbol{x}_{\mathrm{fl}} = \mathrm{const}$. Specifically, using the definition of the Lagrangian displacement [Tab. 5, see Eq. (B.16) for the linear-order expression], we have at linear order

$$\boldsymbol{x}_*(\boldsymbol{x}, \tau) = \boldsymbol{x} + \boldsymbol{s}(\boldsymbol{q}, \tau_*) - \boldsymbol{s}(\boldsymbol{q}, \tau) = \boldsymbol{x} + \left(\frac{D_*}{D} - 1\right)\boldsymbol{s}_{(1)}(\boldsymbol{x}, \tau) + \cdots, \tag{2.28}$$

where $D_*/D \equiv D(\tau_*)/D(\tau)$, and $D(\tau)$ is the linear growth factor defined in Eq. (B.9). Thus, we can neglect the distinction between $\boldsymbol{x}$ and $\boldsymbol{x}_*$ in the argument of the terms in the second line of Eq. (2.27), while those in the first line need to be expanded. For example, we obtain

$$\delta^{(1)}(\boldsymbol{x}, \tau) - \delta^{(1)}(\boldsymbol{x}_*, \tau_*) = \left(1 - \frac{D_*}{D}\right)\delta^{(1)} - \left(\frac{D_*}{D} - 1\right)\frac{D_*}{D}s_{(1)}^i\partial_i\delta^{(1)}, \tag{2.29}$$

where on the right-hand side all quantities are evaluated at $(\boldsymbol{x}, \tau)$.

Finally, we need a relation for the galaxy density $\delta_{g*} = \delta_g(\boldsymbol{x}_*, \tau_*)$. We write

$$\delta_g^{(1+2)}(\boldsymbol{x}_*, \tau_*) = b_1^*[\delta^{(1)} + \delta^{(2)}](\boldsymbol{x}_*, \tau_*) + \varepsilon^*(\boldsymbol{x}_*, \tau_*) + \frac{1}{2}b_2^*\left[\left(\delta^{(1)}\right)_*\right]^2 + b_{K^2}^*\left[\left(K_{ij}^{(1)}\right)_*\right]^2, \tag{2.30}$$

where all quantities are evaluated at $\boldsymbol{x}_*, \tau_*$. We now have allowed for a dependence of δ_g on the tidal field squared $(K_{ij})^2$, since this is a local observable and of the same order in perturbations as δ^2, and thus is expected to be of similar relevance as the term $\propto b_2^*$. Note that because the tidal field is traceless ($\mathrm{tr}[K_{ij}] = 0$), the tidal field cannot enter the bias expansion at linear order. We have also included the leading correction to a deterministic bias relation, namely a stochastic contribution ε^* to the galaxy density field which we consider to be first order. By definition, ε^* is assumed to be uncorrelated with the matter variables. We did not include any further stochastic contribution at second order, a point we will return to below. Note that b_1^* multiplies both $\delta^{(1)}$ and $\delta^{(2)}$, since b_1^* is a physical bias parameter that has to be independent of the perturbative order we are working in (it corresponds to the response of the mean density of galaxies to a change in the background matter density, as we will discuss in Sec. 3).

Now we can simply collect the linear and second-order contributions to the galaxy overdensity at $(\boldsymbol{x}, \tau)$ from Eq. (2.27), to obtain

$$\delta_g^{(1)}(\boldsymbol{x}, \tau) = \left(1 + \frac{D_*}{D}[b_1^* - 1]\right) \delta^{(1)}(\boldsymbol{x}, \tau) + \varepsilon^*$$

$$\delta_g^{(2)}(\boldsymbol{x}, \tau) = \left\{1 + [b_1^* - 1]\left(\frac{D_*}{D}\right)^2\right\} \delta^{(2)} + \left\{\frac{D_*}{D}[b_1^* - 1] - \left(\frac{D_*}{D}\right)^2 [b_1^* - 1] + \frac{1}{2}b_2^*\left(\frac{D_*}{D}\right)^2\right\} [\delta^{(1)}]^2$$

$$+ b_{K^2}^* \left(\frac{D_*}{D}\right)^2 [K_{ij}^{(1)}]^2 + \left(\frac{D_*}{D} - 1\right)\frac{D_*}{D}[b_1^* - 1]s_{(1)}^i \partial_i \delta^{(1)} - \left(\frac{D_*}{D} - 1\right)\varepsilon^* \delta^{(1)}$$

$$+ \left(\frac{D_*}{D} - 1\right)s_{(1)}^i \partial_i \varepsilon^*, \tag{2.31}$$

where on the right-hand side all quantities are evaluated at $(\boldsymbol{x}, \tau)$, except for ε^* which is evaluated at $(\boldsymbol{x}, \tau_*)$. This corresponds to Eq. (53) of [118] (who do not include stochasticity however).[6] We define

$$b_1^E = 1 + \frac{D_*}{D}[b_1^* - 1] \tag{2.32}$$

as the *linear Eulerian bias* (at time $\tau > \tau_*$). This relation first appears in [114], who pointed out that the bias of galaxies must approach unity as time goes by. We will come back to this shortly. Note that if we let $D_*/D \to 0$ while keeping $b_1^L \equiv (D_*/D)b_1^*$ finite, corresponding to a formation at $\tau = 0$ with subsequent conserved evolution, we obtain the relation $b_1^E = 1 + b_1^L$ as in the spherical collapse evolution (Sec. 2.1). We will return to this case below.

Following our discussion above, the coefficient of the second-order density $\delta^{(2)}$ in δ_g has to be b_1^E as well. Thus, we separate out $b_1^E \delta^{(2)}$ in $\delta_g^{(2)}$, and insert the expression for $\delta^{(2)}$ in terms of the linear-order $\delta^{(1)}$, $K_{ij}^{(1)}$, and displacement term [Eq. (B.14)], into the remainder. Reordering terms, we finally obtain the expression for the galaxy density contrast to second order:

$$\delta_g^{(1+2)} = b_1^E \left[\delta^{(1)} + \delta^{(2)}\right] + \varepsilon^* + \frac{1}{2}b_2^E[\delta^{(1)}]^2 + b_{K^2}^E[K_{ij}^{(1)}]^2$$

$$- \left(\frac{D_*}{D} - 1\right)\varepsilon^* \delta^{(1)} + \left(\frac{D_*}{D} - 1\right)s_{(1)}^i \partial_i \varepsilon^*, \tag{2.33}$$

where all quantities (again, except for ε^*) are evaluated at $(\boldsymbol{x}, \tau)$ and the *second-order Eulerian bias param-*

_____

[6]Note that $\boldsymbol{\Psi}$ in Ref. [118] should stand for the displacement from the formation time to the present, which we write as $(D_*/D - 1)\boldsymbol{s}_{(1)}(\boldsymbol{x}, \tau)$.

eters are given by [121]

$$b_2^E(\tau) = b_2^* \left(\frac{D_*}{D}\right)^2 + \frac{8}{21}\left(1 - \frac{D_*}{D}\right)[b_1^E - 1]$$

$$b_{K^2}^E(\tau) = b_{K^2}^* \left(\frac{D_*}{D}\right)^2 - \frac{2}{7}\left(1 - \frac{D_*}{D}\right)[b_1^E - 1]. \qquad (2.34)$$

A relation of the type Eq. (2.33) has been derived in [115], who, crucially, did not separate out the contribution $b_1^E \delta^{(2)}$ at second order however. Let us discuss this interesting result.

- The displacement term $s^i \partial_i \delta$ has canceled out of the deterministic bias relation. This is in fact expected, since the appearance of a nonzero displacement term would mean the galaxy has moved away from the fluid trajectory. Since we have only considered gravity here and neglected all non-gravitational momentum transfer effects, such a displacement cannot happen by way of the Equivalence Principle. The displacement $\boldsymbol{s}$ is still explicit in the stochastic term at second order, since the stochastic field is defined on the formation time slice. However, since ε^* is completely described by a spatially independent 1-point probability distribution function (Sec. 2.8), the displacement term does not contribute to the galaxy clustering statistics. This does change however in the case of non-Gaussianity in the initial conditions, as we will see in Sec. 7.1.2.

- For fixed formation time τ_*, $D_*/D = D(\tau_*)/D(\tau)$ monotonically decreases towards later times. Thus, the galaxy density field becomes progressively less biased with respect to matter as $\tau \to \infty$.

- At second order, evolution induces a term $\propto \varepsilon^* \delta^{(1)}$. If we imagine a realistic galaxy sample that includes galaxies with various formation times, following the "Green's function" approach mentioned above, $\delta_g^{(2)}$ contains a superposition of individual contributions $\varepsilon^* \delta^{(1)}$ from various formation times τ_*. Instead of attempting to model the distribution of formation times of all galaxies in the sample, these contributions can be absorbed by introducing a second stochastic field ε_δ in addition to ε, which enters the bias expansion as $\varepsilon_\delta \delta$. In general, there will be some nonzero (but not perfect) cross-correlation between the fields ε and ε_δ. We return to this in Sec. 2.8.

Finally, we briefly consider the cross-correlation coefficient r between galaxies and matter. This is at linear order given by [116, 122]

$$r_{gm} \equiv \frac{\langle \delta_g \delta \rangle}{\sqrt{\langle \delta_g \delta_g \rangle \langle \delta \delta \rangle}} = \left[1 + \left(\frac{D_*}{D}\right)^2 \frac{\langle (\varepsilon^*)^2 \rangle}{(b_1^E)^2 \langle (\delta^*)^2 \rangle}\right]^{-1/2}. \qquad (2.35)$$

At this order, r_{gm} differs from unity solely due to the stochastic term ε^*. Note that if we wanted to derive the leading nonlinear correction to Eq. (2.35), we would need to include terms up to third order in perturbation theory. We defer this until Sec. 4.1.4. Eq. (2.35) is equally valid in real and in Fourier space. While in Fourier space, r_{gm} is generally less than 1 on all scales, in real space it is equal to 1 if $|\boldsymbol{x_2} - \boldsymbol{x_1}|$ is sufficiently large, since the correlation function of ε^* vanishes at large separations (see Sec. 2.8 for a more precise discussion). For fixed τ_* and ε^*, the cross-correlation coefficient $r_{gm}(\tau)$ asymptotes toward 1 as $\tau \to \infty$, similar to b_1^E. In other words, $(b_1, \langle \varepsilon^2 \rangle) = (1, 0)$ is a fixed point: if galaxies are unbiased and perfectly correlated with matter at some time τ_0, this remains true for all $\tau > \tau_0$.

Bias expansions at initial and final time: Let us now go back to the gravitational evolution of a bias relation given in the initial conditions considered in Sec. 2.1, i.e. of Lagrangian bias (see also [123]). This limit is obtained by letting $\tau_* \to 0$ while keeping $(D_*/D)b_1^* \equiv b_1^L$, $\varepsilon^* \equiv \varepsilon^L$, $(D_*/D)^2 b_2^* \equiv b_2^L$, $(D_*/D)^2 b_{K^2}^* \equiv b_{K^2}^L$, fixed. This is the approach taken with the predictions from the excursion-set (Sec. 5) and peak approaches (Sec. 6). We then obtain the following expression at time τ:

$$\delta_g^{(1+2)}(\boldsymbol{x}, \tau) = (1 + b_1^L)\left[\delta^{(1)} + \delta^{(2)}\right] + (1 + \delta^{(1)})\varepsilon^L + \frac{1}{2}\left\{\frac{8}{21}b_1^L + b_2^L\right\}[\delta^{(1)}]^2 + \left[-\frac{2}{7}b_1^L + b_{K^2}^L\right][K_{ij}^{(1)}]^2,$$

$$(2.36)$$

where we have dropped the stochastic displacement term $\mathbf{s}_{(1)} \cdot \nabla \varepsilon^L$ as it does not contribute to observables (see above). To illustrate the significance of Eq. (2.36), we can contrast it with an example with no evolution, obtained by setting $\tau_* = \tau$, which contains exactly the terms included in the initial bias relation Eq. (2.30):

$$\delta_g^{(1+2)}(\boldsymbol{x}, \tau) = b_1^E \left[\delta^{(1)} + \delta^{(2)} \right] + \varepsilon + \frac{1}{2} b_2^E [\delta^{(1)}]^2 + b_{K^2}^E [K_{ij}^{(1)}]^2 \,. \tag{2.37}$$

Now, letting $b_{K^2}^L = 0$ in Eq. (2.36), we obtain the limit of *LIMD Lagrangian bias* (cf. Sec. 2.1). At finite time, it leads to a bias with respect to the tidal field squared of $b_{K^2}^E = -2/7 b_1^L = -2/7(b_1^E - 1)$. Thus, *a LIMD expansion in the initial conditions is inconsistent with a LIMD expansion at the evolved time τ* [Eq. (2.37) with $b_{K^2}^E = 0$], *and vice versa*, unless bias is trivial, $b_1^* = 1$ and $b_n^* = 0$ for $n > 1$. This was first derived by [117, 124], who pointed out that Eulerian LIMD and Lagrangian LIMD bias lead to different three-point functions. Refs. [79, 118, 125] also discuss the relation between Eulerian LIMD and Lagrangian LIMD bias. One might therefore consider a measurement of a Eulerian tidal bias of $b_{K^2}^E = -2/7(b_1^E - 1)$ as a confirmation of Lagrangian LIMD bias. However, this interpretation assumes a formation time of $\tau_* \to 0$ with subsequent conserved evolution. In fact, the same relation would be measured for any conserved tracer with finite formation time τ_* and nonzero $b_{K^2}^*$, as long as

$$b_{K^2}^* = -\frac{2}{7} [b_1^* - 1] \,. \tag{2.38}$$

Indeed, there is no reason why the formation of halos or galaxies should in general be independent of the local tidal field [117, 126, 127, 128, 118, 129].

Finally, comparing Eq. (2.36) and Eq. (2.37), we also see that a stochastic term $\propto \varepsilon \, \delta^{(1)}$ appears in the former, but is absent in Eq. (2.37). Hence, bias expansions at initial time and evolved time are only equivalent if we also allow for a second-order stochastic term $\varepsilon_\delta \delta$ at formation time. Essentially, while ε corresponds to stochastic fluctuations in the galaxy density, ε_δ corresponds to stochastic fluctuations in the linear bias. This effect is an integral part of the general bias expansion starting at second order.

At this point, it is worth pointing out that the time evolution of the bias parameters of a realistic, observationally selected galaxy sample is difficult to predict. As mentioned at the beginning of Sec. 2.2, a realistic sample consists of galaxies that have formed at various different times. Consider such a sample, where the normalized distribution of formation times is denoted as $p(\tau_*)$. Then, the bias parameter $b_O(\tau)$ corresponding to an operator O at time of measurement τ is given by

$$b_O(\tau) = \int b_O^E(\tau | \{b_{O'}^*\}, \tau_*) p(\tau_*) d\tau_* \,, \tag{2.39}$$

where $b_O^E(\tau | \{b_{O'}^*\}, \tau_*)$ is the Eulerian bias at time τ given a set of bias parameter $\{b_{O'}^*\}$ at formation time τ_*, which at second order is given exactly by the relations in Eq. (2.32) and Eq. (2.34); we can write the stochastic fields ε, ε_O in a similar way. Thus, without detailed knowledge of the formation-time distribution, constraints on the bias parameters at one redshift will not lead to a prediction of the bias parameters at another redshift. Moreover, the distribution of formation times itself at a given point could well depend on the large-scale density and tidal field, which would also affect the time evolution of the bias parameters.

We stress, however, that as long as we keep all relevant terms at a given order, the presence of a distribution of formation times does not lead to new bias parameters. Thus, in most applications, as long as we consider the bias parameters to be effectively free functions of time, we arrive at the correct description regardless of the formation history of galaxies. Moreover, the calculation above shows that the bias parameters are expected to evolve slowly, namely on the same Hubble time scale as the growth of matter perturbations themselves: $d \ln b_O / d\tau \sim \mathcal{H}$.

2.4 Evolution from a joint perturbative solution

An alternative to the approach described in the previous section is to directly solve the full set consisting of Eq. (2.17) and Eqs. (2.19)–(2.20) for δ_g, δ, and θ [130, 118, 131] (see also [132]). At linear order in PT,

one obtains second-order ordinary differential equations (ODE) for δ and δ_g, which can be combined to yield an ODE for $b_1(\tau)$ [133]. The standard approach to solving these equations at nonlinear order is to work in Fourier space, in which case Eqs. (2.19)–(2.20) become Eqs. (B.4)–(B.5) in Appendix B. In the context of the evolution of bias, this approach has the disadvantage that it mixes local physical effects (which are important for bias) with advection terms from the fluid flow such as $s^i \partial_i \delta$, $s^i \partial_i \delta_g$ (which are identical between matter and galaxies and thus are of no relevance for bias).

One option to circumvent this issue is to use Lagrangian perturbation theory (LPT) [78, 134, 135, 79, 120]. The fundamental quantity of LPT is the deformation tensor $M_{ij} \equiv \partial_{q,i} s_j$. The evolved matter density is given by

$$1 + \delta(\boldsymbol{x}, \tau) = |\mathbf{1} + \boldsymbol{M}|_{\boldsymbol{q}, \tau}^{-1}, \tag{2.40}$$

where we have assumed the single-stream regime (in the multi-stream regime, one should sum over all solutions of $\boldsymbol{x} = \boldsymbol{q} + \boldsymbol{s}(\boldsymbol{q}, \tau)$). Now, to describe biased tracers, we have to correspondingly introduce a galaxy deformation tensor $\boldsymbol{M}_g$. Note that the displacement $\boldsymbol{s}$ itself is still the same for matter and galaxies, as long as velocity bias can be neglected (see Sec. 2.7).

One then immediately obtains the Lagrangian version of Eq. (2.24),

$$-\ln[1 + \delta_g(\boldsymbol{x}[\boldsymbol{q}, \tau], \tau)] = \ln|\mathbf{1} + \boldsymbol{M}_g|_{\boldsymbol{q}, \tau} = \ln|\mathbf{1} + \boldsymbol{M}|_{\boldsymbol{q}, \tau} + \ln\left[\frac{|\mathbf{1} + \boldsymbol{M}_g|}{|\mathbf{1} + \boldsymbol{M}|}\right]_{\boldsymbol{q}, \tau_*}. \tag{2.41}$$

So far, this is very similar to the derivation of Sec. 2.3, with the advantage that there is no need to deal with the displacement terms as all terms are explicitly evaluated at a fixed Lagrangian coordinate. On the other hand, the initial bias relation at τ_* now relates $\boldsymbol{M}_g$ to $\boldsymbol{M}$:

$$|\mathbf{1} + \boldsymbol{M}_g(\boldsymbol{q}, \tau_*)|^{-1} = b_{\mathrm{tr}\,M}^* \operatorname{tr} \boldsymbol{M} + b_{\mathrm{tr}(M^2)}^* \operatorname{tr}[\boldsymbol{M}\boldsymbol{M}] + b_{(\mathrm{tr}\,M)^2}^* (\operatorname{tr} \boldsymbol{M})^2 + \varepsilon^* + \varepsilon_{\mathrm{tr}\,M}^* \operatorname{tr} \boldsymbol{M} + \cdots, \tag{2.42}$$

where all terms on the right-hand side are evaluated at $(\boldsymbol{q}, \tau_*)$. This is equivalent to the bias expansion in Eq. (2.30), in the sense that, at second order in perturbations, we can convert each term in Eq. (2.42) into a linear combination of the terms in Eq. (2.30). However, this bias relation does not allow us to easily read off the standard bias parameters (b_1, b_2 and b_{K^2} to second order), since their contributions are spread over all terms in Eq. (2.42), and terms at all higher orders as well.

Another approach to solving Eq. (2.17) and Eqs. (2.19)–(2.20) proceeds by integrating the equations along the fluid flow. In this approach, which was introduced by [131] and dubbed "convective SPT," the advection terms never appear explicitly, and only terms relevant for bias are present. On the other hand, we still deal explicitly with the Eulerian density and tidal field, allowing us to connect to the results of the previous section as well as subsequent sections. We briefly present this approach here, with details given in Appendix B.5. We stress that regardless of the approach taken, the final result has to be the same, and we follow the convective SPT approach only because the resulting expressions offer a clear physical interpretation.

We continue to assume the absence of velocity bias; that is, galaxies are comoving with the matter. The system which we started from, Eq. (2.17) and Eqs. (2.19)–(2.20), can be written in compact form as

$$\frac{\mathrm{D}}{\mathrm{D}\tau}\boldsymbol{\Psi} = -\boldsymbol{\sigma} \cdot \boldsymbol{\Psi} + \boldsymbol{S} \tag{2.43}$$

where we have defined the vector $\boldsymbol{\Psi} = (\delta_g, \delta, \theta)$, $\boldsymbol{\sigma}$ is a matrix that solely depends on the FRW background, and $\boldsymbol{S}$ is a source term which is at least second order in perturbation theory [see Eq. (B.36)].

The system Eq. (2.43) can be solved order by order in a straightforward manner, as detailed in Appendix B.5. Care needs to be taken in deriving the source term in order to allow for an integration along the fluid trajectory. In the following, we will present the solution up to third order. Earlier results up to this order can be found in [136, 118, 137, 138]. To begin, we need to provide an expression for the galaxy density at the initial "formation" time τ_*. Assuming instantaneous formation as before, we include all terms

composed of the density and tidal field, and include all relevant stochastic terms, up to third order:

$$
\delta_g^* = \sum_{n=1}^{3} \frac{b_n^*}{n!} [\delta^*]^n + b_{K^2}^* \, \mathrm{tr} \left[(K_{ij}^*)^2 \right] + b_{K^3}^* \, \mathrm{tr} \left[(K_{ij}^*)^3 \right] + b_{\delta K^2} \delta^* \, \mathrm{tr} \left[(K_{ij}^*)^2 \right]
$$
$$
+ \varepsilon^* + \varepsilon_\delta^* \delta^* + \varepsilon_{\delta^2}^* [\delta^*]^2 + \varepsilon_{K^2}^* \, \mathrm{tr} \left[(K_{ij}^*)^2 \right] \,, \tag{2.44}
$$

where here and throughout, a superscript $*$ indicates that a quantity is evaluated at $\boldsymbol{x}_* \equiv \boldsymbol{x}_{\mathrm{fl}}(\tau_*)$ and τ_*. The stochastic fields ε^*, ε_X^* ($X = \delta$, δ^2, K^2) are assumed to be first-order random fields. Note that a term of the form $\varepsilon_\delta \delta$ has already appeared through second-order evolution in the previous section, indicating that it should be included at second order. Similar reasoning applies to the new third order stochastic terms.

Up to second order, we recover the results of Sec. 2.3. The solution is given by [Eq. (B.45)]

$$
\delta_g^{(1+2)}(\boldsymbol{x}, \tau) = b_1^E(\tau) \left[\delta^{(1)} + \delta^{(2)} \right] + \frac{1}{2} b_2^E(\tau) (\delta^{(1)})^2 + b_{K^2}^E(\tau) (K_{ij}^{(1)})^2 + \varepsilon_\delta^E(\tau) \delta^{(1)} \,, \tag{2.45}
$$

where on the right-hand side, all quantities are evaluated at the *Lagrangian* position $\boldsymbol{q}$ corresponding to $(\boldsymbol{x}, \tau)$, and the Eulerian bias parameters are

$$
b_1^E(\tau) = 1 + \frac{D_*}{D} [b_1^* - 1]
$$
$$
b_2^E(\tau) = b_2^* \left(\frac{D_*}{D} \right)^2 + \frac{8}{21} (b_1^* - 1) \frac{D_*}{D} \left(1 - \frac{D_*}{D} \right)
$$
$$
b_{K^2}^E(\tau) = b_{K^2}^* \left(\frac{D_*}{D} \right)^2 - \frac{2}{7} (b_1^* - 1) \frac{D_*}{D} \left(1 - \frac{D_*}{D} \right) \,, \tag{2.46}
$$

while the Eulerian stochasticity is

$$
\varepsilon_\delta^E(\tau) = \varepsilon_\delta^* \frac{D_*}{D} - \left(\frac{D_*}{D} - 1 \right) \varepsilon^* \,, \tag{2.47}
$$

as derived independently in Sec. 2.3. Note that ε_δ^E has one power of D_*/D less than the other second-order Eulerian quantities, since the fields ε_X^* are defined at τ_*. As mentioned above, $\delta_g(\boldsymbol{x}, \tau)$ is written here as a local function of the matter fields and stochastic variables evaluated at $\boldsymbol{q} = \boldsymbol{x}(\tau = 0)$, i.e. Eulerian quantities evaluated at a fixed Lagrangian position. In order to transform to a Eulerian position $\boldsymbol{x}$, one adds the second-order displacement term to δ_g (as well as analogously to δ and θ) through

$$
- s_{(1)}^i(\boldsymbol{x}, \tau) \partial_i \delta_g(\boldsymbol{x}, \tau) \,. \tag{2.48}
$$

Going to third order, we obtain

$$
\delta_g^{(3)}(\boldsymbol{x}, \tau) = b_1^E \delta^{(3)} + b_2^E \delta^{(1)} \delta^{(2)} + 2 b_{K^2}^E K_{ij}^{(1)} K_{ij}^{(2)} + \frac{1}{6} b_3^E (\delta^{(1)})^3 + b_{K^3}^E (K_{ij}^{(1)})^3 + b_{\delta K^2} \delta^{(1)} (K_{ij}^{(1)})^2
$$
$$
+ b_{\mathrm{td}}^E O_{\mathrm{td}}^{(3)} + \varepsilon_\delta^E \delta^{(2)} + \varepsilon_{\delta^2}^E (\delta^{(1)})^2 + \varepsilon_{K^2}^E (K_{ij}^{(1)})^2 \,, \tag{2.49}
$$

where

$$
O_{\mathrm{td}}^{(3)} \equiv \frac{8}{21} K_{ij}^{(1)} \mathcal{D}^{ij} \left[(\delta^{(1)})^2 - \frac{3}{2} (K^{(1)})^2 \right] \,. \tag{2.50}
$$

Here and throughout, we use the short-hand notation

$$
K^2 \equiv (K_{ij})^2 \equiv \mathrm{tr}[K\,K] = K_{ij} K_{ji} \quad \text{and} \quad K^3 \equiv (K_{ij})^3 \equiv \mathrm{tr}[K\,K\,K] = K_{ij} K_{jk} K_{ki} \,. \tag{2.51}
$$

Again, in order to obtain the density at a fixed order in standard Eulerian perturbation theory, we need to displace δ_g from a fixed Lagrangian position $\boldsymbol{q}$ to the Eulerian position, by expanding in the argument

$\boldsymbol{x}_{\mathrm{fl}}[\boldsymbol{q}, \tau]$. This is given in Eq. (B.48). As this coordinate shift is somewhat tangential to the topic of bias, we do not repeat the results here.

The third-order Eulerian bias parameters are given by

$$b_3^E(\tau) = b_3^* \left(\frac{D_*}{D}\right)^3 + \left[(b_1^* - 1)\frac{4}{1323}\left(199 - 35\frac{D_*}{D}\right) + \frac{13}{7}b_2^*\frac{D_*}{D}\right]\frac{D_*}{D}\left(\frac{D_*}{D} - 1\right)$$

$$b_{\delta K^2}^E(\tau) = b_{\delta K^2}^* \left(\frac{D_*}{D}\right)^3 + \left[-(b_1^* - 1)\frac{1}{147}\left(33 + 7\frac{D_*}{D}\right) + \left(\frac{2}{7}b_2^* + b_{K^2}^*\right)\frac{D_*}{D}\right]\frac{D_*}{D}\left(\frac{D_*}{D} - 1\right)$$

$$b_{K^3}^E(\tau) = b_{K^3}^* \left(\frac{D_*}{D}\right)^3 + \left[(b_1^* - 1)\frac{2}{63}\left(-11 + 7\frac{D_*}{D}\right) + 2b_{K^2}^*\frac{D_*}{D}\right]\frac{D_*}{D}\left(\frac{D_*}{D} - 1\right)$$

$$b_{\mathrm{td}}^E(\tau) = \left[(b_1^* - 1)\frac{1}{6}\left(\frac{D_*}{D} - \frac{23}{7}\right) + \frac{5}{2}b_{K^2}^*\frac{D_*}{D}\right]\frac{D_*}{D}\left(\frac{D_*}{D} - 1\right). \tag{2.52}$$

We see that, as expected, all b_O^E converge to b_O^* for $D_*/D \to 1$ (when $\tau_* \to \tau$), with the exception of $O_{\mathrm{td}}^{(3)}$ since we did not allow for it at the formation time [Eq. (2.44)]. Also, they vanish for $D_*/D \to 0$, unless b_O^* diverge in the limit $\tau_* \to 0$, which is usually assumed when writing a Lagrangian bias relation. Specifically, we can easily obtain the prediction for b_O^E assuming Lagrangian LIMD bias, by letting $\tau_* \to 0$ and all cubic $b_O^* = 0$ apart from $b_3^*(D_*/D)^3 = b_3^L$. This yields

$$b_3^E(\tau) = b_3^L - \frac{796}{1323}b_1^L - \frac{13}{7}b_2^L$$

$$b_{\delta K^2}^E(\tau) = \frac{11}{49}b_1^L - \frac{2}{7}b_2^L$$

$$b_{K^3}^E(\tau) = \frac{22}{63}b_1^L$$

$$b_{\mathrm{td}}^E(\tau) = \frac{23}{42}b_1^L. \tag{2.53}$$

The first line matches Eq. (2.16) exactly, since it is the only term that remains for a spherically symmetric perturbation, whose evolution is governed exactly by the spherical collapse solution.

Finally, the Eulerian stochastic terms at third order become

$$\varepsilon_{\delta^2}^E(\tau) = \varepsilon_{\delta^2}^* \left(\frac{D_*}{D}\right)^2 - \frac{4}{21}(\varepsilon_\delta^* - \varepsilon^*)\frac{D_*}{D}\left(\frac{D_*}{D} - 1\right)$$

$$\varepsilon_{K^2}^E(\tau) = \varepsilon_{K^2}^* \left(\frac{D_*}{D}\right)^2 + \frac{2}{7}(\varepsilon_\delta^* - \varepsilon^*)\frac{D_*}{D}\left(\frac{D_*}{D} - 1\right). \tag{2.54}$$

In analogy to the second-order case, we see that the third-order stochastic terms ε_{δ^2}, ε_{K^2} are induced by gravitational evolution, even if they are absent initially. Thus, following the discussion at the end of Sec. 2.3, we should allow for these terms in the general bias expansion.

Let us now turn to the operator $O_{\mathrm{td}}^{(3)}$ which has appeared in the third-order evolved bias relation Eq. (2.49). While the operator $O_{\mathrm{td}}^{(3)}$ is third order in perturbations, it cannot be expressed locally in terms of the linear density and tidal field, as is clear from Eq. (2.50). Therefore, including this operator goes beyond the initial bias relation Eq. (2.44): that is, starting from third order in perturbation theory, *a bias expansion involving only the local density and tidal field is not sufficient.* Interestingly, the operator $O_{\mathrm{td}}^{(3)}$ can be expressed in a variety of equivalent ways at third order, as derived in Appendix C:

- $O_{\mathrm{td}}^{(3)} = 2K^{ij}\mathcal{D}_{ij}[\delta + (f\mathcal{H})^{-1}\theta] + F(\delta, K_{lm})$, where $f \equiv d\ln D/d\ln a$ is the logarithmic linear growth rate, and $F(\delta, K_{lm})$ denotes cubic local combinations of density and tidal field as enumerated in Eq. (2.44) [see Eqs. (C.15)–(C.16) for the precise relation]. Thus, $O_{\mathrm{td}}^{(3)}$ is related to the local difference of the tidal field, $\propto \mathcal{D}_{ij}\delta$, and velocity shear, $\propto -\mathcal{D}_{ij}\theta$. This type of operator was first considered in [128].

31

- $O_{\rm td}^{(3)} = (4/5)K^{ij}\left([\mathcal{H}f]^{-1}{\rm D}/{\rm D}\tau - 1\right)K_{ij} + G(\delta, K_{lm})$ [Eq. (C.11) with Eq. (C.9)]. We see that $O_{\rm td}^{(3)}$ is proportional to the *convective time derivative* of the tidal field (at second order) along the fluid flow.

- $O_{\rm td}^{(3)} = (8/3)[M^{ij} - (1/3)\delta^{ij}\,{\rm tr}\,M]\left([\mathcal{H}f]^{-1}\partial_\tau - 1\right)M_{ij}$ [Eq. (C.14)]. This is the Lagrangian expression of $O_{\rm td}^{(3)}$, showing that this operator is related to the time derivative of the Lagrangian deformation tensor (recall that in Lagrangian coordinates, convective time derivatives reduce to partial derivatives; Tab. 5).

All these equivalent formulations show that the operator $O_{\rm td}^{(3)}$ is a local observable. An observer comoving with the galaxy could measure it, for example, by measuring the proper time derivative of the local tidal field. Indeed, we would expect any quantity that emerges from the gravitational evolution of a conserved tracer to be a local gravitational observable. Thus, given that $O_{\rm td}^{(3)}$ has exactly two spatial derivatives for each power of the potential, counting $\mathcal{D}_{ij}$ as zero net derivatives, it is justified to include it in our category of *local bias operators*.

Let us briefly pause to consider how our treatment, which has gone to third order in perturbations, might continue to higher orders. In particular, what terms beyond simple combinations of δ and K_{ij} would one expect at fourth and higher order? Our first encounter of such a term, $O_{\rm td}^{(3)}$ in Eq. (2.50), shows that at third order, only a certain combination of $\mathcal{D}^{ij}(\delta^2)$ and $\mathcal{D}^{ij}(K^2)$ appears, not each one of them individually. One might then wonder if, following [128], it is sufficient to include the velocity shear $(\partial_i\partial_j/\nabla^2)\theta$ in addition to δ, K_{ij}. As we will see in the next section, this is not the case. At fourth order, the velocity shear is no longer sufficient.

Before continuing to the general bias expansion, it is worth considering how these results change when allowing for an expansion history beyond EdS. Specifically, we continue to assume the validity of GR, but allow for a nonzero curvature, and cosmological constant Λ or dark energy component (where we neglect the effect of dark energy perturbations; we will discuss further generalizations in Sec. 8.1 and Sec. 8.3). As shown in Appendix B.6, the equations of motion maintain the same structure as in EdS. We also show there that no new type of bias operator appears up to including third order for a general expansion history, although the time evolution of bias parameters for conserved tracers [Eq. (2.52)] is modified. Further, the departure from EdS of the equations of motion is completely quantified by the quantity $\Omega_m(a)/f^2(a) - 1$. Since for ΛCDM, as well as most viable dark energy expansion histories, we have $f \approx \Omega_m^{0.55}$ [139], this quantity is approximately

$$\Omega_m(a)/f^2(a) - 1 \approx -0.1\ln\Omega_m\,. \tag{2.55}$$

For $\Omega_{m0} \geq 0.3$, this quantity remains less than 0.13 at all redshifts. This explains why the EdS approximation in perturbation theory calculations, with $a(\tau)$ replaced with $D(\tau)$ in the final result, is numerically accurate.

2.5 General perturbative bias expansion

Building upon the specific second- and third-order results we have presented above, the goal of this section is to derive a general framework for the perturbative bias expansion. That is, we aim to derive a set of bias parameters that is *sufficient to describe the statistics of any large-scale structure tracer*, within the realm of perturbation theory. In the following, we will continue to use the term "galaxy", although one should keep in mind that the results are more generally applicable to other tracers. In fact, the general bias expansion can also be applied to the local matter power spectrum [140].

In Sec. 2.3–2.4, we have seen specific bias relations involving the density and tidal field, as well as an additional operator $O_{\rm td}^{(3)}$, and their stochastic counterparts such as ε_δ, ε_{K^2}. While the previous derivations assumed a conserved tracer, this is easily generalized by employing the functions $b_O^E(\tau; \{b_{O'}^*\}, \tau_*)$ as Green's functions, as discussed at the end of Sec. 2.3. We now abandon the assumption of number conservation entirely, and allow our galaxies to form and merge arbitrarily. In general, a deterministic bias relation for an arbitrary galaxy sample can be written as

$$\delta_g(\boldsymbol{x}, \tau) = \sum_O b_O(\tau)O(\boldsymbol{x}, \tau)\,, \tag{2.56}$$

where $\delta_g = n_g/\overline{n}_g - 1$ is the density contrast of the galaxies, b_O are bias parameters, and O are operators constructed out of the matter density field, gravitational potential, and in general other perturbations. The statistics of the galaxies at a fixed time are then given by the statistics of the matter density field, potential, and so on at the same time, and a finite set of numbers b_O. Note that this relation is only useful if we have a finite set of operators and hence a limited number of bias parameters which can be marginalized over while retaining cosmological information. Refs. [113, 141] have considered generalizations of bias parameters to bias functions, $b_N \to b_N(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N)$. However, in order to retain predictive power, we need to condense these functionals into a finite set of terms. This is precisely the goal of this section.

In the context of cosmological perturbation theory, we thus need to identify which operators should be included in the bias relation up to any given order in perturbations. We will refer to such a set of operators as *basis*. Of course, such a basis is not unique, since any linearly independent combination of the operators in one basis represents another basis. One significant example is choosing operators given in the Lagrangian frame, for example in terms of the distortion tensor $\boldsymbol{M}$ (Tab. 5) (*Lagrangian bias*), or in terms of quantities at the final, observation time (*Eulerian bias*), such as Eq. (2.49). We will give explicit examples of both.

In the previous section, we have seen that, when starting from a bias expansion that includes powers of the density field and tidal field, i.e. $O \in \{\delta, \delta^2, (K_{ij})^2, \dots\}$ at one time τ_*, then conserved evolution under gravity introduces additional bias operators, in particular $O_{\mathrm{td}}^{(3)}$ in Eq. (2.49), at a later time τ. In this sense, a bias expansion restricted to powers of the density and tidal field is incomplete, since it does not allow us to describe, for example, a galaxy that has formed at some earlier time (or even in the initial conditions) and then passively evolved following number conservation; see also the introduction of Ref. [113] which nicely describes this problem.

As a starting point, we will work under the following assumptions in the remaining subsections:

- Gravitation is described by General Relativity (GR).

- The impact of massive neutrinos and dark energy perturbations can be neglected.

- The initial conditions are Gaussian and adiabatic, and any isocurvature perturbations induced by physics in the early Universe can be neglected.

We will relax the first two assumptions in Sec. 8, and the last assumption in Sec. 7 as well as Sec. 8 (for isocurvature perturbations between baryons and CDM). The last point implies that we only consider the adiabatic growing mode. In a standard ΛCDM cosmology, this is accurate at the percent-level (see Sec. 8.2).

2.5.1 Spacetime picture of bias and evolution

The key physical feature of the formation of tracers such as halos and galaxies is that it happens over *long time scales* [92], while the formation process is limited to relatively *small spatial scales* [142]. With long time scales here we mean that the formation takes place over an appreciable fraction of the age of the Universe. On the other hand, the matter that forms dark matter halos comes from within a region of a few Mpc in comoving size. In other words, the spacetime region that encompasses the formation and evolution of galaxies is of the "spaghetti" shape sketched in Fig. 6. This is related to the well-known fact that the nonlinear scale, at which the fractional density perturbations become order unity, is much smaller than the Hubble horizon. This latter fact forms the basis of all perturbation-theory approaches to large-scale structure, and it is of similarly crucial importance in the derivation of bias relations.

Suppose that the abundance of galaxies at position $\boldsymbol{x}$ only depends on the distribution of matter in a finite region around $\boldsymbol{x}$, of characteristic comoving length scale R_*. We will call R_* the "nonlocality scale" which is understood to be defined on a certain time slice (for example, the final or initial times). We will further discuss its physical significance in Sec. 2.6. For dark matter halos, R_* is expected to be of order the Lagrangian radius. Now consider the case where we look at statistics of galaxies or halos on scales much larger than R_*. Then, we can approximate the bias relation as effectively *local in space*, thus reducing the bias expansion from a functional expansion to an ordinary Taylor expansion as in Eq. (2.56) (see Sec. 2.6 for the functional expansion). In effective field theory (EFT) language ([143, 83, 86, 88]; see [144] for a review), the spatially local approximation provides the low-energy effective description of the full,
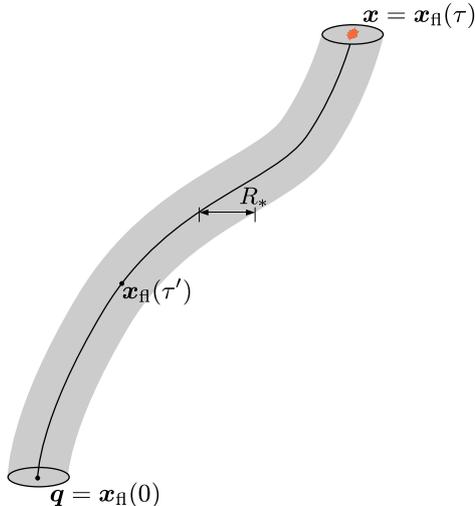
Figure 6: Sketch of the spacetime region involved in the formation of tracers such as halos or galaxies. Time is running vertically. The solid line denotes the fluid trajectory $\boldsymbol{x}_{\mathrm{fl}}(\tau')$ from a Lagrangian position $\boldsymbol{q} = \boldsymbol{x}_{\mathrm{fl}}(\tau = 0)$ to a Eulerian position $\boldsymbol{x} = \boldsymbol{x}_{\mathrm{fl}}(\tau)$ at time τ. The shaded region with a comoving spatial extent of order R_* denotes the region from which the matter within the galaxy and its host halo originates, or the region of influence feedback processes—whichever is larger.

complicated dynamics of the formation of galaxies (Sec. 2.10).

When considering galaxy formation as effectively local, the only quantities that are relevant for the formation of galaxies are then the *density and the tidal field* $\partial_i \partial_j \Phi(\boldsymbol{x}_{\mathrm{fl}}(\tau'), \tau')$ *along the trajectory of a Lagrangian patch enclosing the galaxy* ([88, 131]; Ref. [125] only considered the matter density along the fluid trajectory). One way to prove this statement is to invoke the equivalence principle, which states that in a free-falling frame, such as that comoving with the trajectory $\boldsymbol{x}_{\mathrm{fl}}(\tau)$, the leading locally observable gravitational effect is given by second derivatives of the metric tensor. Moreover, essentially all tracers of the LSS are non-relativistic. Then, the only relevant component of the metric tensor is the time-time-component. On sub-horizon scales, this is in turn equivalent to the tensor $\partial_i \partial_j \Phi$, where Φ is the gravitational potential defined in Eq. (1.2). This tensor can further be decomposed into the trace $\nabla^2 \Phi$ which is directly related to the density perturbation δ through the Poisson equation; and the trace-free part K_{ij} [Eq. (2.22)], which quantifies the tidal field proper. An alternative, more rigorous derivation of the same result is given by the Conformal Fermi Coordinate (CFC) approach [145, 146], which clarifies the meaning of the density perturbation and K_{ij} in the relativistic context. We will return to this in Sec. 2.9.

This reasoning provides the physical justification for our definition of *local bias* (Sec. 1.3) as encompassing all terms in the general bias expansion that are constructed (without any further spatial derivatives) out of the density and tidal field along the fluid trajectory: these are precisely the leading local gravitational observables for a comoving observer. In conformal-Newtonian gauge, these terms are characterized by exactly two spatial derivatives acting on each power of the potential Φ. Note that we do not need to assume a conserved, passively evolving galaxy sample here. Any gravitational interactions such as mergers [147] do not, on sufficiently large scales, depend on any property apart from the local density and tidal field. A galaxy sample that preferentially resides in halos formed from recent major mergers might have a larger nonlocality scale R_* than that of typical halos of the same mass. Nevertheless, it will be a finite scale, and presumably still of order the Lagrangian radius of these halos as argued above.

In our reasoning we did however implicitly assume that the small-scale initial conditions, i.e. those of much smaller scale than the large-scale correlations we are interested in, are statistically uncorrelated over large scales. This is the case for Gaussian initial conditions, which we assume in this section.

Now, let us formalize our reasoning. The dependence on $\delta(\boldsymbol{x}_{\mathrm{fl}}(\tau'), \tau')$ and $K_{ij}(\boldsymbol{x}_{\mathrm{fl}}(\tau'), \tau')$ can be written as multiple time integrals over the fluid trajectory. For example, in the simplest case, for a given operator

34

O constructed out of δ and K_{ij}, we can formally expand the time integral as [88]

$$
\delta_g(\boldsymbol{x}, \tau) \supset \int^\tau d\tau'\, f_O(\tau, \tau') O(\boldsymbol{x}_{\mathrm{fl}}(\tau'), \tau')
$$
$$
= \left[\int^\tau d\tau'\, f_O(\tau, \tau') \right] O(\boldsymbol{x}, \tau) + \left[\int^\tau d\tau'\, (\tau' - \tau) f_O(\tau, \tau') \right] \frac{\mathrm{D}}{\mathrm{D}\tau} O(\boldsymbol{x}, \tau) + \cdots , \qquad (2.57)
$$

where $\mathrm{D}/\mathrm{D}\tau$ is the convective derivative along the fluid flow [Eq. (2.18)]. In order to provide a basis of operators at a fixed time, as demanded by Eq. (2.56), we thus have to include time derivatives along the fluid flow, such as $\mathrm{D}(\partial_i \partial_j \Phi)/\mathrm{D}\tau$, in the basis of operators. Including time derivatives, of arbitrary order, of powers of the density field and tidal field then provides a complete basis of operators for the local bias expansion (that is, at lowest order in *spatial* derivatives). However, this is not very satisfying: since the formation of galaxies happens on long time scales, the higher-order terms not written in Eq. (2.57) are not necessarily smaller than the ones that we include; in fact, this seems to suggest that we need infinitely many operators in our bias expansion. Fortunately however, at fixed order in perturbation theory, only a finite number of time derivatives are linearly independent, and thus the basis can be completed with a finite number of operators. The physical reason is that the time evolution of the large-scale, quasi-linear perturbations is predicted at any given order in perturbation theory. Moreover, given our assumptions, they evolve at exactly the same rate on linear scales, since the linear growth factor is scale-independent. Thus, the departure from the linear growth rate is higher order in perturbations. We will now show precisely how to use this fact to obtain a finite set of operators.

2.5.2 Lagrangian basis of operators

To begin with, let us work on the initial time slice, in Lagrangian coordinates $\boldsymbol{q} = \boldsymbol{x}_{\mathrm{fl}}(\tau = 0)$. This simplifies the treatment, since in Lagrangian coordinates convective time derivatives $\mathrm{D}/\mathrm{D}\tau$ [Eq. (2.18)] reduce to simple time derivatives $\partial/\partial\tau$.

Consider a Lagrangian operator $O_L(\boldsymbol{q}, \tau)$ constructed out of d_O factors of $\partial_{q,i} \partial_{q,j} \Phi^{(1)}(\boldsymbol{q})$ (or, equivalently, the distortion tensor M_{ij}). In perturbation theory, it can be written as

$$
O_L(\boldsymbol{q}, \tau) = D^{d_O}(\tau) O_L^{(d_O)}(\boldsymbol{q}, \tau_0) + D^{d_O+1}(\tau) O_L^{(d_O+1)}(\boldsymbol{q}, \tau_0) + \cdots , \qquad (2.58)
$$

where $D(\tau)$ is the linear growth factor, normalized to some reference time τ_0. By construction, d_O is the perturbative order of the leading contribution to O_L. The operators $O_L^{(n)}(\boldsymbol{q}, \tau_0)$, $n = d_O,\, d_O + 1, \ldots$, correspond to the contributions to O_L at n-th order in perturbation theory, evaluated at the reference time τ_0. Here, we have assumed for simplicity that the n-th order growth factor is given by the linear growth factor to the n-th power. This is only strictly valid for an EdS (flat matter-dominated) Universe where $D(\tau) = a(\tau)$, although also generally very accurate for other cosmologies such as ΛCDM. We will discuss this at the end of Sec. 2.5.3.

Given the general relation Eq. (2.58), allowing for time derivatives of O_L in the bias expansion [see Eq. (2.57)] is, at n-th order in perturbation theory, equivalent to including the contributions $O_L^{(m)}$ ($m \leq n$) at *each order* individually. This is because, at this order in perturbation theory, the time derivatives of any operator O_L are given by linear combinations of the $O_L^{(m)}$ ($m \leq n$). Consider for example a second-order Lagrangian operator, which in perturbation theory can be written as, following Eq. (2.58), $O_L(\boldsymbol{q}, \tau) = D^2(\tau) O_L^{(2)}(\boldsymbol{q}, \tau_0) + D^3(\tau) O_L^{(3)}(\boldsymbol{q}, \tau_0) + \cdots$. Then, in third-order perturbation theory, the n-th convective time derivative of O_L is given by

$$
\left(\frac{\mathrm{D}}{\mathrm{D}\tau} \right)^n O_L(\boldsymbol{q}, \tau) \bigg|^{(3)} = \frac{\partial^n}{\partial\tau^n} O_L(\boldsymbol{q}, \tau) \bigg|^{(3)} = \left(\frac{d^n}{d\tau^n} D^2(\tau) \right) O_L^{(2)}(\boldsymbol{q}, \tau_0) + \left(\frac{d^n}{d\tau^n} D^3(\tau) \right) O_L^{(3)}(\boldsymbol{q}, \tau_0). \quad (2.59)
$$

At any given fixed time, this is just a linear combination of $O_L^{(2)}(\boldsymbol{q}, \tau_0)$ and $O_L^{(3)}(\boldsymbol{q}, \tau_0)$. This continues to hold correspondingly at any fixed, higher order in perturbation theory, and becomes even more obvious

35

when transforming the time coordinate from τ to $\ln D(\tau)$: then, the right-hand side of Eq. (2.59) simply becomes $2^n D^2(\tau) O^{(2)} + 3^n D^3(\tau) O^{(3)}$.

Note that, even when starting with an operator O_L that is a local combination of $\partial_{q,i} \partial_{q,j} \Phi$, the higher-order terms $O_L^{(n)}$ generated by time evolution are in general not expressible as local combinations of $\partial_{q,i} \partial_{q,j} \Phi$. Instead, terms involving $\partial^i \partial^j / \nabla^2$ acting on powers of $\partial_l \partial_m \Phi$ appear, just as we have seen with $O_{\mathrm{td}}^{(3)}$ in Eq. (2.50). Fundamentally, this is a consequence of the fact that gravity acts over long distances, so that the gravitational evolution of the tidal field cannot be approximated as local [148, 149]. In particular, the invariant definition of the tidal field is a certain projection of the Weyl tensor [150], which corresponds to the specific part of the Riemann tensor that is not locally related to the stress-energy tensor via the Einstein equations.

Crucially however, while we assume that galaxy formation is local (we relax this assumption in Sec. 2.6), we do *not* have to assume that gravity is local. Indeed, one can straightforwardly derive the evolution of the tidal field in perturbation theory, and take that into account in the bias expansion, namely through the terms $O_L^{(m)}$ described above. One finds that the time derivatives of the tidal field only contain a small subset of all possible operators constructed out of $\partial^i \partial^j / \nabla^2$ acting on powers of $\partial_l \partial_m \Phi$. Only these specific operators should be included in the bias expansion, because only these terms correspond to local observables, essentially time derivatives of the tidal field along the fluid flow.

Let us now construct an explicit Lagrangian basis of bias operators. It is convenient to write these in terms of the Lagrangian distortion tensor introduced in Sec. 2.4,

$$M_{ij} \equiv \frac{\partial s_j}{\partial q^i}\,. \tag{2.60}$$

Note that at linear order, $M_{ij}^{(1)}$ is directly proportional to $\partial_{q,i} \partial_{q,j} \Phi^{(1)}$. Knowing that we can always recast the time derivatives as a sum of higher-order operators, we simply have to take all scalar contractions of the contributions $M_{ij}^{(n)}$ at each perturbative order; up to quadratic order, this was already written in Eq. (2.42). However, we do not need to include $\mathrm{tr}[M^{(n)}] \equiv \delta^{ij} M_{ij}^{(n)}$ with $n > 1$, as these terms can always be expressed in terms of lower-order operators through the equations of motion for M_{ij} (see [80, 151] for the explicit expression of the latter). The basis up to fourth order then is [131][7]

$$
\begin{array}{ll}
1^{\text{st}} & \mathrm{tr}[M^{(1)}] \\[4pt]
2^{\text{nd}} & \mathrm{tr}[(M^{(1)})^2],\ (\mathrm{tr}[M^{(1)}])^2 \\[4pt]
3^{\text{rd}} & \mathrm{tr}[(M^{(1)})^3],\ \mathrm{tr}[(M^{(1)})^2]\,\mathrm{tr}[M^{(1)}],\ (\mathrm{tr}[M^{(1)}])^3,\ \mathrm{tr}[M^{(1)} M^{(2)}] \\[4pt]
4^{\text{th}} & \mathrm{tr}[(M^{(1)})^4],\ \mathrm{tr}[(M^{(1)})^3]\,\mathrm{tr}[M^{(1)}],\ \left(\mathrm{tr}[(M^{(1)})^2]\right)^2,\ (\mathrm{tr}[M^{(1)}])^4, \\[4pt]
& \mathrm{tr}[M^{(1)}]\,\mathrm{tr}[M^{(1)} M^{[2]}],\ \mathrm{tr}[M^{(1)} M^{(1)} M^{(2)}],\ \mathrm{tr}[M^{(1)} M^{(3)}],\ \mathrm{tr}[M^{(2)} M^{(2)}]\,.
\end{array}
\tag{2.61}
$$

At fourth order, we have used the fact that, as a 3×3 symmetric matrix, $M_{ij}^{(1)}$ is characterized by three independent rotational invariants, allowing us to eliminate $\mathrm{tr}[(M^{(1)})^2](\mathrm{tr}[M^{(1)}])^2$. Starting at third order, M_{ij} is no longer symmetric; that is, the displacement vector also has a curl component. However, at each order in perturbations, the antisymmetric part of M_{ij} can be re-expressed in terms of combinations of lower-order contributions to the symmetric part, $M_{(ij)} \equiv (M_{ij} + M_{ji})/2$, via the equations of motion [80, 151]. Hence, it is sufficient to write the bias expansion purely in terms of the contributions to the symmetric part, $M_{(ij)}^{(n)}$.

All of the operators in Eq. (2.61) are easily evaluated in Lagrangian perturbation theory. The first instance of a convective time derivative appears at third order in the bias expansion through the operator

---

[7]We have added two operators at fourth order that were missing in [131]; specifically, the first two terms in the very last line of Eq. (2.61).

tr$[M^{(1)}M^{(2)}]$, which is precisely related to the operator $O_{\rm td}^{(3)}$ introduced in Sec. 2.4 (see p. 31). Note however that, while $M_{ij}^{(1)}$ and $M_{ij}^{(2)}$ can be rephrased in terms of second derivatives of the gravitational and velocity potentials, this is no longer true for $M_{ij}^{(3)}$ which appears at fourth order. Fortunately, with this construction it is now obvious how to extend the complete set of bias operators to higher orders. Thus, Eq. (2.61) is directly applicable to calculate statistics of biased tracers in LPT. The disadvantage, as has already been mentioned after Eq. (2.42), is that Eq. (2.61) is not very convenient to connect to well-known perturbative bias expansions. This is because the matter density contrast δ and the tensor $\boldsymbol{M}$ are nonlinearly related through $\delta(\boldsymbol{q}) = |\mathbf{1} + \boldsymbol{M}|^{-1} - 1$. Thus, the well-known linear bias term $b_1\delta(\boldsymbol{q})$ contributes to many terms in the list Eq. (2.61). In the next section, we will derive an equivalent basis that is closer to standard perturbative bias expansions.

2.5.3 Eulerian basis of operators

A Eulerian basis can similarly be constructed out of $\partial_{x,i}\partial_{x,j}\Phi(\boldsymbol{x},\tau)$ and its convective time derivatives. Here, we follow Ref. [131] who have defined[8]

$$\Pi_{ij}^{[1]}(\boldsymbol{x},\tau) = \frac{2}{3\Omega_m\mathcal{H}^2}\partial_{x,i}\partial_{x,j}\Phi(\boldsymbol{x},\tau) = K_{ij}(\boldsymbol{x},\tau) + \frac{1}{3}\delta_{ij}\delta(\boldsymbol{x},\tau)\,, \qquad (2.62)$$

which in the notation of Sec. 2.3 and Sec. 2.4 contains $\delta = \operatorname{tr}\Pi^{[1]}$ and K_{ij} as the trace-free part of $\Pi_{ij}^{[1]}$. Note that the superscript [1], to be distinguished from (1), refers to the fact that $\Pi^{[1]}$ *starts* at first order in perturbation theory, but contains higher-order terms as well. We then define the higher-order tensors $\Pi^{[n]}$ recursively by convective time derivatives:

$$\Pi_{ij}^{[n]} = \frac{1}{(n-1)!}\left[(\mathcal{H}f)^{-1}\frac{\mathrm{D}}{\mathrm{D}\tau}\Pi_{ij}^{[n-1]} - (n-1)\Pi_{ij}^{[n-1]}\right]\,. \qquad (2.63)$$

Note that $\Pi_{ij}^{[n]}$ is symmetric for any n. By construction, the lowest-order contribution to $\Pi^{[n]}$ in perturbation theory is at n-th order. This helps us keep track of all the relevant terms at any given order. However, unlike in the Lagrangian case, here the $\Pi^{[n]}$ are not simply the n-th order perturbative contributions to $\Pi^{[1]}$. This is because in the Eulerian case, convective time derivatives [Eq. (2.18)] are not equal to ordinary derivatives $\partial/\partial\tau$. The reason why a basis constructed out of $\Pi_{ij}^{[n]}$ is complete, however, is exactly analogous to the Lagrangian case discussed above: at n-th order in perturbation theory, $\Pi_{ij}^{[1]}$, which contains the matter density and tidal field up to n-th order, only has n different time dependences $D(\tau), D^2(\tau), \cdots D^n(\tau)$. Therefore, any higher convective time derivative $(\mathrm{D}/\mathrm{D}\tau)^m$ with $m > n$ can be expressed in terms of the first, second, $\cdots$, n-th time derivatives, when neglecting terms higher than n-th order in perturbation theory.

The quantity $\operatorname{tr}[\Pi^{[n]}]$ is a linear combination of convective time derivatives of the Eulerian density perturbation δ. At any given order, these can be written as combinations of lower-order operators, by way of the Eulerian fluid equations (for example, $\operatorname{tr}[\Pi^{(2)}] = (17/21)\delta^2 + (2/7)(K_{ij})^2$ in second-order PT; see Appendix C), and can thus be excluded from the basis for $n > 1$. This is in analogy with $\operatorname{tr}[M^{(n)}]$ in the Lagrangian basis. Correspondingly, the bias coefficients of these terms can be seen as integrals over the kernels $f_{\operatorname{tr}\Pi^{[1]}}(\tau,\tau')$, $f_{(\operatorname{tr}\Pi^{[1]})^2}(\tau,\tau',\tau'')$ and so on, introduced in Eq. (2.57) above, against progressively higher powers of the growth factor [88].

---

[8]Note that the prefactor was absorbed into the definition of Φ there.

Up to fourth order, we therefore have, in exact formal analogy to Eq. (2.61),

$$
\begin{aligned}
&1^{\text{st}} && \operatorname{tr}[\Pi^{[1]}] && (2.64)\\
&2^{\text{nd}} && \operatorname{tr}[(\Pi^{[1]})^2]\,,\ (\operatorname{tr}[\Pi^{[1]}])^2 \\
&3^{\text{rd}} && \operatorname{tr}[(\Pi^{[1]})^3]\,,\ \operatorname{tr}[(\Pi^{[1]})^2]\operatorname{tr}[\Pi^{[1]}]\,,\ (\operatorname{tr}[\Pi^{[1]}])^3\,,\ \operatorname{tr}[\Pi^{[1]}\Pi^{[2]}] \\
&4^{\text{th}} && \operatorname{tr}[(\Pi^{[1]})^4]\,,\ \operatorname{tr}[(\Pi^{[1]})^3]\operatorname{tr}[\Pi^{[1]}]\,,\ \left(\operatorname{tr}[(\Pi^{[1]})^2]\right)^2\,,\ (\operatorname{tr}[\Pi^{[1]}])^4\,, \\
& && \operatorname{tr}[\Pi^{[1]}]\operatorname{tr}[\Pi^{[1]}\Pi^{[2]}]\,,\ \operatorname{tr}[\Pi^{[1]}\Pi^{[1]}\Pi^{[2]}]\,,\ \operatorname{tr}[\Pi^{[1]}\Pi^{[3]}]\,,\ \operatorname{tr}[\Pi^{[2]}\Pi^{[2]}]\,.
\end{aligned}
$$

This basis offers the advantage of having a close connection to the standard Eulerian bias expansion. For example, the coefficient of the term $(\operatorname{tr}[\Pi^{[1]}])^n$ is precisely $b_{\delta^n} = b_n/n!$, since $\operatorname{tr}[\Pi^{[1]}(\boldsymbol{x},\tau)] = \delta(\boldsymbol{x},\tau)$ at all orders. The term $\operatorname{tr}[(\Pi^{[1]})^2] = (K_{ij})^2 + \delta^2/3$ on the other hand contains the tidal field squared. Of course, as in the Lagrangian case, explicit time derivatives appear in the bias expansion at third order through the operator $\operatorname{tr}[\Pi^{[1]}\Pi^{[2]}]$, which again is directly related to the operator $O_{\text{td}}^{(3)}$ (see p. 31 and Appendix C). As mentioned after Eq. (2.61), gravitational and velocity shear are no longer sufficient in the bias expansion starting at fourth order, as evidenced by the appearance of $\Pi_{ij}^{[3]}$.

We now discuss the key approximation made in constructing the convenient bases Eq. (2.61) and Eq. (2.64), namely that all operator contributions at a given perturbative order n have the same time dependence, $[D(\tau)]^n$. This is only strictly true in an EdS Universe, while in ΛCDM and quintessence cosmologies new time dependences appear at each new order. For example, second-order operators can have a time dependence given by $[D(\tau)]^2$ or by $D_2(\tau)$, where $D_2(\tau) \propto \int D^2 d\ln D$ is the second-order growth factor. This means that the operators in the bases described above are in general not sufficient anymore. However, we show in Appendix B.6 that the first instance of a new term in the bias expansion appears only at fourth order. Specifically, the $\operatorname{tr}[\Pi^{[1]}\Pi^{[3]}]$ term in Eq. (2.64) splits into two terms which, however, have to have very similar bias coefficients if the nonlinear growth factors approximately obey $D_n(\tau) \simeq [D(\tau)]^n$; the departures from this relation are at the percent level for a standard ΛCDM cosmology. This means that the additional operators added to complete the operator bases described here will be (i) fourth and higher order; (ii) suppressed by a numerical prefactor $\lesssim 0.1$ relative to the terms included in Eq. (2.64). They will thus be irrelevant in most practical applications. Note that, since we only work to third order in perturbation theory there, all results given in Sec. 4 hold in ΛCDM and quintessence cosmologies.

2.6 Higher-derivative bias

In the treatment of bias so far, we have approximated the formation of halos and galaxies as perfectly local in a spatial sense. After reordering the time derivatives along the fluid trajectory, we have written the bias expansion for $\delta_g(\boldsymbol{x},\tau)$ in terms of operators evaluated at the same location: $O(\boldsymbol{x},\tau)$ in the Eulerian basis, or $O_L(\boldsymbol{q},\tau)$ in the Lagrangian basis. However, we know that the formation of halos and galaxies involves the collapse of matter from a finite region in space, and thus, the *local bias* expansion derived above cannot be completely correct on all scales. In this section, we study the limitation of the spatially-local approximation and derive the set of additional operators to include in the expansion Eq. (2.56). We refer to these operators as *higher-derivative operators*. They naturally arise in peak theory or the excursion-set approach [13, 152, 153] (see Sec. 5 and Sec. 6 for a detailed discussion).

In order to incorporate the deviation from perfect locality of galaxy formation, we should replace the local operators $O(\boldsymbol{x},\tau)$ appearing in Eq. (2.56) with *functionals* [154, 155]. For example, the linear-order operator in the Eulerian basis, $O = \delta$, now becomes

$$
b_\delta(\tau)\delta(\boldsymbol{x},\tau) \to \int d^3\boldsymbol{y}\, F_\delta(\boldsymbol{y},\tau)\delta(\boldsymbol{x}+\boldsymbol{y},\tau)\,, \tag{2.65}
$$

where $F_\delta(\boldsymbol{y},\tau)$ is a kernel that is in general time dependent. Here, we have used the homogeneity of the Universe, or the absence of preferred locations, which dictates that F_δ is independent of $\boldsymbol{x}$. We can now

perform a formal series expansion of δ around $\boldsymbol{x}$, leading to

$$b_\delta(\tau)\delta(\boldsymbol{x},\tau) \to \left[\int d^3\boldsymbol{y}\, F_\delta(\boldsymbol{y},\tau)\right]\delta(\boldsymbol{x},\tau) + \left[\frac{1}{6}\int d^3\boldsymbol{y}\, |\boldsymbol{y}|^2\, F_\delta(\boldsymbol{y},\tau)\right]\nabla_x^2\delta(\boldsymbol{x},\tau) + \cdots$$
$$= b_\delta(\tau)\delta(\boldsymbol{x},\tau) + b_{\nabla^2\delta}(\tau)\nabla_x^2\delta(\boldsymbol{x},\tau) + \cdots, \tag{2.66}$$

where statistical isotropy demands the absence of any preferred directions with which the derivative operators could be contracted.

Hence, there is no term linear in spatial derivatives, and the leading higher-derivative term involves the Laplacian of $\delta(\boldsymbol{x},\tau)$. We then identify the standard linear bias $b_\delta = b_1$ with the "total mass" of the kernel, while the second integral, the "moment of inertia" of the kernel, defines a new higher-derivative bias parameter $b_{\nabla^2\delta}$. This bias parameter has dimension [length]^2; the characteristic scale R_* that sets the magnitude of $|b_{\nabla^2\delta}| \sim R_*^2$ is the scale of the spatial support of the kernel $F_\delta(\boldsymbol{y},\tau)$, which we identify with the nonlocality scale of the tracer R_*. Note that $\nabla^2\delta$ is a local observable for a comoving observer and, following our discussion in Sec. 2.5.1, should thus be included in the general bias expansion in any case. Such a term naturally appears for example in the peak approach (Sec. 6) without any explicit nonlocality, induced by the constraint of a negative curvature of the smoothed density field. The formal derivation from a kernel as in Eq. (2.66) provides another physical interpretation for the bias parameter $b_{\nabla^2\delta}$ and its relation to the spatial scale R_*.

Let us briefly discuss how baryonic effects are also captured by higher-derivative terms. While this statement holds for general perturbations of the stress tensor of matter, we consider the effect of gas pressure perturbations δp here for simplicity. These are sourced by density perturbations, so that at linear order

$$\delta p^{(1)} = c_s^2 \overline{\varrho}_b \delta^{(1)}, \tag{2.67}$$

where c_s is the sound speed, and $\overline{\varrho}_b$ is the mean physical baryon density. The relevant quantity for dynamics is the gradient $\boldsymbol{\nabla}\delta p$ of the pressure. Thus, if we allow for the galaxy density to depend on $\boldsymbol{\nabla}\delta p$, we have to add the leading scalar quantities that can be constructed out of $\boldsymbol{\nabla}\delta p$ to the bias expansion, leading to

$$\delta_g \supset b_{\nabla^2\delta p}\nabla^2\delta p + b_{(\boldsymbol{\nabla}\delta p)^2}(\boldsymbol{\nabla}\delta p)^2 \propto b_{\nabla^2\delta p}\nabla^2\delta + b_{(\boldsymbol{\nabla}\delta p)^2}(\boldsymbol{\nabla}\delta)^2. \tag{2.68}$$

We see that these terms are higher-derivative and, moreover, the only term linear in perturbations is exactly of the same form as in Eq. (2.66). We will discuss nonlinear higher-derivative terms, such as $(\boldsymbol{\nabla}\delta)^2$, below. On the other hand, in contrast to gas, radiation can travel large distances and hence, in principle, lead to a dependence of the local galaxy density on the matter distribution within a very large region; i.e. it can significantly increase the scale R_*. We will discuss realistic estimates of these effects below. For now, let us continue with the general discussion based on Eq. (2.66).

In Eq. (2.65), the kernel F_δ describes how the formation of galaxies depends on the precise distribution of matter in the vicinity of $\boldsymbol{x}$. Therefore, it is sensible to connect the second moment of F_δ with the nonlocality scale R_* introduced in Sec. 2.5.1. If F_δ is given by a typical filtering kernel, e.g. a Gaussian or tophat [cf. Eqs. (A.26)–(A.28)], then we expect that $b_{\nabla^2\delta} > 0$: if we are at the location of a peak in the density field, then the smoothed density will be smaller than the un-smoothed one. On the other hand, $\nabla^2\delta < 0$ at such a location, and so a positive $b_{\nabla^2\delta}$ leads to the expected behavior, i.e. that peaks are damped by smoothing. Consider for example our toy model of Sec. 2.1. At linear order, the proto-halo density perturbation can be written as $\delta_{h,L}(\boldsymbol{q}) = b_1^L\delta_R^{(1)}(\boldsymbol{q})$. A formal expansion of the smoothed density field $\delta_R^{(1)}$ around the un-smoothed linear density $\delta^{(1)}$ yields

$$\delta_{h,L}(\boldsymbol{q},\tau) = b_1^L\delta^{(1)}(\boldsymbol{q},\tau) + b_{\nabla^2\delta}^L\nabla_q^2\delta^{(1)}(\boldsymbol{q},\tau) + \cdots, \quad \text{where} \quad b_{\nabla^2\delta}^L = \frac{1}{10}b_1^L R^2. \tag{2.69}$$

Here we have assumed a tophat filtering kernel to be specific. Note that $b_{\nabla^2\delta}^L$ does not necessarily have to be positive; peaks of the density field provide an illuminating example for the case $b_{\nabla^2\delta}^L < 0$ (see Sec. 6). Continuing the derivative expansion, the next higher correction in Eq. (2.69) scales as $R^4\nabla^4\delta$.
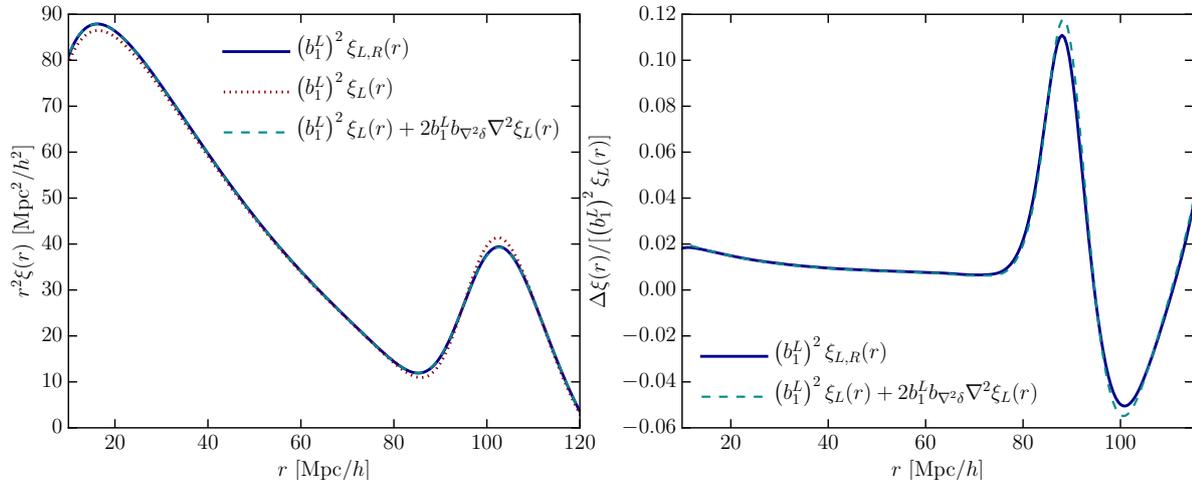
Figure 7: Illustration of the effect of higher-derivative bias in the context of the thresholding toy model of Sec. 2.1. We show two-point correlation functions multiplied by r^2 in order to better illustrate the effect on the BAO feature. The values for $R = 4.21\,h^{-1}\,\mathrm{Mpc}$ (mass scale $2.5 \cdot 10^{13}\,h^{-1}\,M_\odot$), $b_1^L = 1.5$ and $z = 0$ are the same as in Fig. 4. *Left panel:* Linear-order contribution $(b_1^L)^2\xi_{L,R}(r)$ to Eq. (2.7) (solid), and the linear LIMD bias prediction $(b_1^L)^2\xi_L(r)$ (dotted); note that the former contains the filtering kernel while the latter does not. Some differences are seen on small scales and especially around the BAO feature. When adding the leading higher-derivative bias [Eq. (2.69)] to the LIMD prediction (dashed), the perturbative bias expansion matches the smoothed two-point correlation function very well. *Right panel:* Relative deviation of the results shown in the left panel from the linear LIMD prediction.

In Fourier space, the term proportional to $b_{\nabla^2\delta}$ corresponds to a "scale-dependent bias" $-b_{\nabla^2\delta}k^2\delta \propto (R_*k)^2\delta$, with higher corrections scaling as $(R_*k)^{2n}\delta$. However, let us emphasize again that this is an expansion in powers of k^2, rather than a general function $f(k)$, which is how the term "scale-dependent bias" is sometimes interpreted. To make this distinction clear, we will use the term higher-derivative bias throughout.

At linear order, which we have assumed in Eq. (2.66) and Eq. (2.69), the distinction between higher-derivative terms in Eulerian and Lagrangian frames, i.e. $\nabla_x^2\delta$ vs. $\nabla_q^2\delta$, is irrelevant. This is no longer the case at nonlinear order in perturbations. Moreover, the relation between Lagrangian and Eulerian higher-derivative biases is complicated by *velocity bias*, unlike the case in the local bias expansion, as we will see in the next section. Crucially, the difference between derivatives with respect to $\boldsymbol{q}$ and those with respect to $\boldsymbol{x}$, which involves the distortion tensor $\boldsymbol{M}$ [Tab. 5], can be absorbed by nonlinear higher-derivative terms, which then render bias expansions in Eulerian and Lagrangian frames equivalent again. We will return to this below.

In terms of the two-point correlation function of galaxies, the leading contribution is of the form

$$\xi_{gg}(r)\Big|_{\text{higher deriv.}} = 2b_1 b_{\nabla^2\delta}\, \nabla_r^2\xi(r)\,, \tag{2.70}$$

where $\xi(r)$ is the matter two-point correlation function. Note that this term can become observationally relevant not just if the correlation scale r is of order R_*, but on the much larger scale of the BAO feature ($r \sim 100\,h^{-1}\,\mathrm{Mpc}$) because of the narrow width of this feature [156, 152]. This is illustrated, in the context of the thresholding toy model of Sec. 2.1, in Fig. 7 (left panel), which shows the effect of smoothing (on the scale $R \simeq 4\,h^{-1}\,\mathrm{Mpc}$) on the two-point correlation function. The smoothing damps the BAO feature slightly. We also show the effect of including the leading higher-derivative contribution through Eq. (2.70) and Eq. (2.69). The right panel of Fig. 7 clearly shows that the leading higher-derivative term captures the bulk of the smoothing effect. Thus, by including one additional bias parameter, we incorporate the leading effect of the finite size of the thresholded regions, and improve the precision of the bias expansion significantly. At this point, it is worth emphasizing that the higher-derivative terms primarily modify the

amplitude of the BAO feature. If the BAO scale is determined by marginalizing over the broad-band shape of the galaxy two-point correlation function, as is usually done in BAO studies, then the estimated scale is insensitive to $b_{\nabla^2\delta}$. On the other hand, if our goal is to extract the full information from the galaxy two-point correlation function, this term has to be included (Sec. 4.1).

As we have seen, higher-derivative biases introduce an additional spatial scale, R_*, into the perturbative bias expansion (unlike the *local* bias expansion at leading order in derivatives, which only involves the same Hubble time scale that governs the evolution of matter itself). The significance of this new scale is that it provides a fundamental cutoff for the perturbative approach to galaxy clustering: when the scale r on which we measure correlations approaches R_*, all higher-derivative terms become relevant, and the perturbative description loses all predictive power. The same effect happens in Fourier space for $k \sim R_*^{-1}$. Thus, even if we were able to predict the properties of the matter density field perfectly, the nonlocality of the formation of galaxies sets a fundamental limit on the scales over which we can describe the statistics of galaxies perturbatively. In practice, even for halos, it is still unclear whether higher-order local bias terms or higher-derivative operators provide the actual cutoff of the perturbative theory. For galaxies, the cutoff will most likely depend strongly on the specific galaxy sample considered. In the following, we discuss approximate estimates of and constraints on R_* for different tracers:

- ***Dark matter halos:*** Since halo formation in N-body simulations is governed exclusively by gravity, one expects R_* to be comparable to the Lagrangian radius $R(M)$ of halos. This is because the matter within a given halo originates from a region of size $R(M)$. This is also borne out by nonlinear models of halo formation such as the excursion set (Sec. 5) and peaks of the Lagrangian density field (Sec. 6), and indeed the toy model of Sec. 2.1, where $R(M)$ is the filter scale used to define the significance $\nu_c = \delta_{\mathrm{cr}}/\sigma(R)$ [just as we found in Eq. (2.69)]. As discussed above however, the precise value and indeed *sign* of $b_{\nabla^2\delta}$ depend on the details of the model considered.

- ***Galaxies:*** If the properties of galaxies in a given sample are completely determined by those of their host halos, as assumed in the halo occupation distribution and abundance matching approaches (Sec. 9.1), then the scale R_* for these galaxies is given by that of the host halos, i.e. $R(M)$. On the other hand, if the local rate of galaxy formation is significantly modified by the radiation field (e.g., the flux of ionizing UV radiation), then R_* could be as large as the absorption length of this radiation [157, 158, 159], which can be several hundred Mpc. Thermal heating of the intergalactic medium by high-energy cosmic-ray cascades is another possibility for long-range influences [160, 161], as these cosmic rays have large mean-free paths as well. These effects can in turn strongly modify the shape of the galaxy two-point function and thus affect the measured position of the BAO feature [162, 155]. Similar conclusions hold for large-scale outflows, for example, or strong jets launched by active galactic nuclei, although the scale of these phenomena is expected to be at most tens of Mpc [142] and thus significantly smaller than the mean free path of UV photons.

- ***Line emission from diffuse gas:*** This is an interesting case, which includes the Lyman-α forest as well as intensity mapping. Ignoring the effects of any large-scale fluctuations in the ambient radiation field, the nonlocality scale of the gas is of order the Jeans scale, $R_* \sim 1/k_J$, which is very small ($1/k_J \sim 0.1\,h\,\mathrm{Mpc}^{-1}$ [163, 164]) for the relatively cold gas observed using these channels. Combined with the fact that these tracers are observed at fairly high redshifts, where the nonlinear scale is small, this suggests that line emission from the intergalactic medium can be modeled accurately to very small scales. Unfortunately, the line emission depends on the ionization state of the medium, which in turn is controlled by the ambient radiation field [165, 166]. As mentioned above, the mean free path of ionizing radiation in the intergalactic medium is very large, so that in fact R_* for these tracers is *not* small. Ref. [167] provides a very clear illustration of this effect on large-scale statistics of the Lyman-α forest. The information loss by being restricted to very large scales above R_* can however be reduced by explicitly modeling this effect through measured cross-correlations with sources of ionizing radiation such as galaxies and quasars [168].

Finally, we turn to higher-derivative corrections to nonlinear operators in the bias expansion. While technically more complicated, this follows in strict analogy to the linear case leading to Eq. (2.66); the remainder of the section is not essential for the subsequent developments and can be skipped on a first reading. Let us work again in Eulerian coordinates, noting that exactly the same reasoning goes through in the Lagrangian basis in terms of derivatives with respect to $\boldsymbol{q}$. As described in Sec. 2.5.3, each operator in the local basis can be written as

$$O(\boldsymbol{x}, \tau) = \Pi^{[i_1]}(\boldsymbol{x}, \tau) \cdots \Pi^{[i_n]}(\boldsymbol{x}, \tau) \,, \tag{2.71}$$

where $\Pi_{ij}^{[n]}$ is defined in Eqs. (2.62)–(2.63), and we have suppressed tensor indices in this expression since they are irrelevant for the following arguments. Going beyond locality then means that we should introduce a kernel $F_O(\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_n; \tau)$ to replace terms in the local bias expansion so that, again suppressing tensor indices,

$$b_O(\tau) O(\boldsymbol{x}, \tau) \to \int d^3\boldsymbol{y}_1 \cdots d^3\boldsymbol{y}_n \, F_O(\boldsymbol{y}_1, \cdots, \boldsymbol{y}_n; \tau) \Pi^{[i_1]}(\boldsymbol{x} + \boldsymbol{y}_1, \tau) \cdots \Pi^{[i_n]}(\boldsymbol{x} + \boldsymbol{y}_n, \tau) \,. \tag{2.72}$$

Note that the operators $\Pi_{ij}^{[n]}$ are local observables, and the formation of halos can depend on the detailed distribution of the $\Pi_{ij}^{[n]}$ within the scale R_*. Thus, they also form the building blocks for the general higher-derivative expansion. It is then straightforward to perform the same Taylor expansion around $\boldsymbol{x}$ as in Eq. (2.66), for each factor $\Pi^{[i_j]}$, resulting at leading order in terms of the type

$$\Pi^{[i_1]}(\boldsymbol{x}, \tau) \cdots \left[\nabla_x^2 \Pi^{[i_j]}(\boldsymbol{x}, \tau) \right] \cdots \Pi^{[i_n]}(\boldsymbol{x}, \tau) \quad \text{and}$$
$$\Pi^{[i_1]}(\boldsymbol{x}, \tau) \cdots \left[\partial_{x,k} \Pi^{[i_j]}(\boldsymbol{x}, \tau) \right] \cdots \left[\partial_x^l \Pi^{[i_k]}(\boldsymbol{x}, \tau) \right] \cdots \Pi^{[i_n]}(\boldsymbol{x}, \tau) \,. \tag{2.73}$$

Note that the indices k and l in the second line can be contracted among themselves or with the tensor indices of the $\Pi_{ij}^{[n]}$. Thus, *in order to obtain the complete set of higher-derivative operators, we have to allow for all contractions of derivatives on each Π factor in the operator basis Eq. (2.64), including contractions with the tensor indices of the Π themselves.* The analogous construction works for the Lagrangian basis Eq. (2.61) in terms of ∂_q acting on $M_{ij}^{[n]}$. Including the complete set of higher-derivative terms is also necessary and sufficient to ensure that the Lagrangian and Eulerian bias expansions are equivalent at higher order in derivatives.

Explicitly, the leading higher-derivative terms $\mathcal{O}(R_*^2)$ in the Eulerian basis are, up to second order in perturbations, given by

$$\mathcal{O}(R_*^2): \ 1^{\text{st}} \qquad \nabla^2 \operatorname{tr}[\Pi] \tag{2.74}$$
$$2^{\text{nd}} \qquad \operatorname{tr}[(\nabla^2 \Pi)\Pi] \,, \ \operatorname{tr}[(\partial_i \Pi)(\partial^i \Pi)] \,, \ (\nabla^2 \operatorname{tr}[\Pi]) \operatorname{tr}[\Pi] \,, \ (\partial_i \operatorname{tr}[\Pi])(\partial^i \operatorname{tr}[\Pi]) \,,$$
$$\partial_k \Pi_{ij} \partial^i \Pi^{kj} \,, \ \Pi^{kl} \partial_k \partial_l \operatorname{tr}[\Pi] \,, \ \Pi^{kl} \partial_k \partial_i \Pi^i{}_l \,, \ \partial^i \Pi_{ij} \partial^j \operatorname{tr}[\Pi] \,, \ \partial^i \Pi_{ij} \partial_k \Pi^{jk} \,,$$

where we have denoted $\Pi \equiv \Pi^{[1]}$ for clarity. Note that the second line includes the $(\nabla \delta)^2$ term already discussed in the context of pressure perturbations, Eq. (2.68). Clearly, the number of higher-derivative operators and corresponding bias coefficients increases rapidly toward higher order in perturbation theory. However, as outlined above, all the contributions quoted here are suppressed by a factor of $(R_* k)^2$ on large scales relative to the leading terms in Eq. (2.64).

2.7 Velocity bias

Starting from Eq. (2.17), we have assumed that galaxies and matter comove along the same fluid trajectories. We now show why this is consistent within the local bias expansion. Galaxies in general experience different peculiar forces than the matter fluid. This is both due to the fact that galaxies are strongly influenced by baryonic physics, and that galaxy formation happens within a spatial region of finite size R_*

(Sec. 2.6), so that their center-of-mass acceleration is some weighted mean of the local gravitational acceleration field within this region. Let us denote the peculiar acceleration of galaxies with respect to the matter fluid as $\boldsymbol{a}_g$. The Euler equation for the galaxy velocity field $\boldsymbol{v}_g$ is then

$$\frac{\partial}{\partial \tau} \boldsymbol{v}_g + \mathcal{H} \boldsymbol{v}_g + (\boldsymbol{v}_g \cdot \nabla) \boldsymbol{v}_g = -\nabla \Phi + \boldsymbol{a}_g \,. \tag{2.75}$$

Note that, in the effective field theory of the matter fluid (EFT [83], Appendix B.3), a peculiar acceleration $\boldsymbol{a}_m$ also appears in the Euler equation for the matter fluid velocity $\boldsymbol{v}$. There, $a_m^i = -\partial_j \tau^{ij}/\rho_m$ is sourced by the effective stress tensor of matter τ^{ij} that is induced by integrating out the small-scale non-perturbative modes (see Appendix B.3). Strictly speaking, we define $\boldsymbol{a}_g$ here as the difference between the large-scale acceleration field of galaxies and $\boldsymbol{a}_m$ (denoted as $\boldsymbol{f}_g - \boldsymbol{f}$ in [131]). Subtracting the Euler equation from Eq. (2.75) and defining $\boldsymbol{v}_{\rm rel} \equiv \boldsymbol{v}_g - \boldsymbol{v}$, one can then easily show that [131]

$$\frac{\mathrm{D}}{\mathrm{D}\tau} \boldsymbol{v}_{\rm rel} + \mathcal{H} \boldsymbol{v}_{\rm rel} + (\boldsymbol{v}_{\rm rel} \cdot \nabla) \boldsymbol{v}_g = \boldsymbol{a}_g \,, \tag{2.76}$$

where $\mathrm{D}/\mathrm{D}\tau$ is throughout defined with respect to the matter velocity $\boldsymbol{v}$. We see that $\boldsymbol{a}_g$ is the source of galaxy velocity bias $\boldsymbol{v}_g - \boldsymbol{v}$. If we assume a conserved tracer in the spirit of Sec. 2.4, the continuity equation becomes

$$\frac{\mathrm{D}}{\mathrm{D}\tau} (\delta_g - \delta) = -\theta(\delta_g - \delta) - \nabla \cdot \left[(1 + \delta_g) \boldsymbol{v}_{\rm rel} \right] . \tag{2.77}$$

Our general results on velocity bias derived below are independent of the conserved-tracer assumption, however.

Since an observer in a given galaxy can in principle measure the relative acceleration between the galaxy's center of mass and the matter fluid, $\boldsymbol{a}_g$ is a locally observable quantity. As such, its effective large-scale expansion can again only involve gravitational observables, that is, $\partial_i \partial_j \Phi$ and its derivatives. Expanding $\boldsymbol{a}_g$ in powers of Φ, isotropy implies that each term involving n Φ fields must have at least $2n + 1$ derivatives. This is simple to see: any local observable composed of n powers of $(\partial_i \partial_j \Phi)$, say, has an even number of indices. In order to construct a vector a_g^k, we have to add an additional derivative ∂^k, leading to $2n + 1$ derivatives in total. The single leading term ($n = 1$) is given by $a_g^i \propto \partial^i \nabla^2 \Phi$. Physically, the peculiar acceleration can be sourced both by (statistical) differences in the local gravitational potential gradient for galaxies and matter, induced for example by smoothing or the peak constraint, and by non-gravitational effects such as the pressure forces discussed around Eqs. (2.67)–(2.68). Indeed, we have seen in the previous section that both effects lead to higher-derivative terms [Eq. (2.66) and Eq. (2.68), respectively].

We have thus proven that galaxy velocity bias is a *higher-derivative* effect, which justifies why we were able to consistently set $\boldsymbol{v}_g = \boldsymbol{v}$ in Sec. 2.2–2.4. The general velocity bias expansion then consists of all 3-vectors that can be constructed from the local gravitational observables and their spatial derivatives, following the same procedure as described in the previous section. At leading order, there is only a single contribution, which we will consider below.

Note that a velocity bias that is not suppressed by derivatives has been studied in the literature [130, 118]. However, as argued here, such a velocity bias violates the equivalence principle. On the other hand, if the galaxy density depends on multiple fluids, such as baryons and CDM, a *relative* velocity perturbation between these fluids is a local observable even without any derivatives. If present in the initial conditions on large scales, such a relative velocity between baryons and CDM also needs to be taken into account in the bias expansion, as described in Sec. 8.2.

By construction, our expansion only includes the leading terms in the large-scale limit. On small scales, a velocity bias of sub-halos within massive dark matter halos is well established in simulations [169, 170, 171]. These fully nonlinear effects are beyond the reach of the perturbative treatment described here. Further, while the galaxy velocity is unbiased at the level of the local bias expansion, the galaxy momentum density $\boldsymbol{j}_g = (1 + \delta_g) \boldsymbol{v}_g$ is not, since it is weighted by the galaxy density. All measurements of velocities of halos and galaxies in simulations and observations naturally yield the momentum density and must be carefully

re-weighted to obtain an accurate estimate of $\boldsymbol{v}_g$ [172, 173]. Ref. [173] empirically verified that halo velocity bias on scales $k \leq 0.1\,h\,\mathrm{Mpc}^{-1}$ is less than 2%.

Let us now consider the leading higher-derivative velocity bias at linear order, and its effect on the evolution of bias. In order to solve Eq. (2.76) at linear order, we need an expression for the peculiar acceleration $\boldsymbol{a}_g$ (again, strictly speaking it is the effective relative acceleration between galaxies and matter on large scales). As discussed above, $\boldsymbol{a}_g$ has at least three derivatives acting on Φ. At linear order and leading order in derivatives, $\boldsymbol{a}_g$ thus has to be proportional to $\boldsymbol{\nabla}\delta$, i.e.

$$\boldsymbol{a}_g(\boldsymbol{x}_{\mathrm{fl}}(\tau), \tau) = A_g(\tau)\boldsymbol{\nabla}\delta(\boldsymbol{x}_{\mathrm{fl}}(\tau), \tau)\,. \tag{2.78}$$

Other possible choices, such as $\nabla^2 \boldsymbol{v}$, are equivalent to $\boldsymbol{\nabla}\delta$ at linear order. The dimensionless proportionality constant $A_g(\tau)$, essentially a bias parameter quantifying the effective acceleration due to small-scale perturbations, is in general a free function of time. A_g is expected to be of the same order as the leading higher-derivative biases (Sec. 2.6), i.e. $A_g = \mathcal{O}(R_*^2 \mathcal{H}^2)$. Note that since R_* is a spatial scale and there are no preferred directions in the galaxy's rest frame, the lowest nontrivial dependence on R_* has to be of this form (cf. the low-k expansion of a generic spherically symmetric convolution kernel $W_R(k) = c_1 + c_2 k^2 R^2 + \cdots$).

We can then immediately integrate the linearized version of Eq. (2.76) to obtain

$$\boldsymbol{v}_g(\boldsymbol{x}, \tau) = \boldsymbol{v}(\boldsymbol{x}, \tau) + \frac{1}{a(\tau)} \int_0^\tau d\tau'\, a(\tau')\boldsymbol{a}_g(\boldsymbol{x}, \tau') = \boldsymbol{v}(\boldsymbol{x}, \tau) + \beta_{\boldsymbol{\nabla}\delta}(\tau)\boldsymbol{\nabla}\delta(\boldsymbol{x}, \tau)\,, \quad \text{where}$$

$$\beta_{\boldsymbol{\nabla}\delta}(\tau) \equiv \frac{1}{a(\tau)D(\tau)} \int_0^\tau d\tau'\, a(\tau')D(\tau')A_g(\tau')\,. \tag{2.79}$$

Thus, we can trade the acceleration bias $A_g(\tau)$ for a bias in the galaxy velocity $\beta_{\boldsymbol{\nabla}\delta}(\tau)$. We will reserve the notation $\beta_O(\tau)$ for velocity-bias parameters. It is instructive to consider two simple limiting cases for the time dependence of $\boldsymbol{a}_g(\tau)$. If $\boldsymbol{a}_g(\tau) = \mathrm{const}$, implying $A_g(\tau) \propto D^{-1}(\tau)$, and assuming an EdS Universe for simplicity, Eq. (2.79) immediately yields $\beta_{\boldsymbol{\nabla}\delta}(\tau)\nabla^2\delta(\boldsymbol{x}, \tau) \propto \tau$, i.e. the velocity-bias term has the same time dependence as $\boldsymbol{v}$ in the standard growing mode itself. In this case, one can write

$$\boldsymbol{v}_g(\boldsymbol{x}, \tau) \overset{\boldsymbol{a}_g = \mathrm{const}}{=} \left[1 + \beta_{\nabla^2 \boldsymbol{v}}\nabla^2\right]\boldsymbol{v}(\boldsymbol{x}, \tau)\,, \tag{2.80}$$

where $\beta_{\nabla^2 \boldsymbol{v}}$ is constant. This time evolution was first proposed in the context of the peak model [152, 174] (Sec. 6.9.1).

If on the other hand $\boldsymbol{a}_g(\tau) \propto \delta_D(\tau - \tau_*)$, corresponding to an instantaneous boost of galaxy velocities relative to matter at time τ_*, then $(\boldsymbol{v}_g - \boldsymbol{v})(\boldsymbol{x}, \tau) \propto a^{-1}(\tau)\boldsymbol{\nabla}\delta(\boldsymbol{x}_{\mathrm{fl}}(\tau_*), \tau_*) \propto a^{-1}(\tau)$ for $\tau > \tau_*$ [equivalently, $\beta_{\boldsymbol{\nabla}\delta} \propto (aD)^{-1}$, Eq. (2.79)]. This corresponds to the decaying relative-velocity mode of a system of two fluids coupled by gravity [118], and can be understood as the usual decay of peculiar velocities (when not sourced) in an expanding Universe. Thus, the different results on the evolution of large-scale velocity bias obtained in the literature are a consequence of different assumptions on the time evolution of the relative acceleration $\boldsymbol{a}_g$.

Finally, we consider the setup studied in Sec. 2.2–2.4. That is, we prescribe a bias relation at a "formation time" τ_*, and assume conserved evolution of the tracers afterwards. We can then integrate the continuity equation Eq. (2.77) to obtain the galaxy density δ_g. At linear order and including the leading higher-derivative term, we write the galaxy density at an initial time τ_* as

$$\delta_g(\boldsymbol{x}, \tau_*) = b_1^* \delta(\boldsymbol{x}, \tau_*) + b_{\nabla^2\delta}^* \nabla^2\delta(\boldsymbol{x}, \tau_*)\,. \tag{2.81}$$

At some later time τ, δ_g is then given by

$$\delta_g(\boldsymbol{x}, \tau) = b_1^E(\tau)\delta(\boldsymbol{x}, \tau) + b_{\nabla^2\delta}^E(\tau)\nabla^2\delta(\boldsymbol{x}, \tau)$$

$$\text{where} \quad b_{\nabla^2\delta}^E(\tau) = b_{\nabla^2\delta}^* \frac{D_*}{D} - \int_{\tau_*}^\tau d\tau'\, \beta_{\boldsymbol{\nabla}\delta}(\tau')\frac{D(\tau')}{D(\tau)}\,. \tag{2.82}$$

We have used $D_*/D \equiv D(\tau_*)/D(\tau)$ and $b_1^E(\tau) = 1 + (b_1^* - 1)(D_*/D)$ as defined in Sec. 2.3. This generalizes the result of [152] to a time-dependent $\beta_{\boldsymbol{\nabla}\delta}$. Following our arguments after Eq. (2.78), the two contributions to $b_{\nabla^2\delta}^E$ are expected to be of the same order $\sim R_*^2$, noting that $A_g \sim R_*^2 \mathcal{H}^2$, while $\beta_{\boldsymbol{\nabla}\delta} \sim R_*^2 \mathcal{H}$. We see that the evolution of the higher-derivative bias in the density depends on the time evolution of the velocity bias. Unlike the case of the local bias expansion, where the evolution of bias parameters for a conserved tracer at a given order is uniquely determined by the knowledge of the bias parameters *at fixed time*, at higher order in derivatives we require knowledge of the velocity bias $\beta_{\boldsymbol{\nabla}\delta}$ along the *entire trajectory*. This is because higher-derivative terms allow for non-gravitational effects to play a role, whose evolution is not necessarily tied to the gravitational evolution of matter. We stress that it is still possible to completely describe galaxy bias at a fixed time, on sufficiently large scales, in terms of a finite number of bias parameters.

Going back to the two limiting cases for A_g considered above, we find that, for $\boldsymbol{a}_g = \text{const}$, the velocity bias contribution to $b_{\nabla^2\delta}^E(\tau)$ is also constant (in EdS), so that for $\tau \gg \tau_*$ this eventually becomes the dominant contribution, as is the case in the peak model ([152]; see Sec. 6.9.1). On the other hand, for an initial boost, $A_g \propto \delta_D(\tau - \tau_*)$, this term decays as τ^{-3} in EdS (this corresponds to the adiabatic decaying mode of the density field). In this case, the initial higher-derivative bias contribution, which scales as $b_{\nabla^2\delta}^* D_*/D \propto \tau^{-2}$, dominates over the velocity bias contribution at late times. Of course, all of these results are only valid under the assumption of conserved evolution.

2.8 Stochasticity

The bias expansion described so far captures the impact of long-wavelength perturbations on the galaxy density. That is, we have ignored the influence of small-scale perturbations on the formation of galaxies, which is stochastic, as the small-scale initial conditions are not correlated over long distances (this is a consequence of the assumed Gaussianity of the initial conditions). In order to take this into account, we have to introduce stochastic fields in the bias relation, as already done in Sec. 2.3–2.4. This is related to a phenomenon known in the literature as *stochastic bias* [175, 176, 154]. Further, in Sec. 2.3–2.4 we have seen that stochasticity on one time slice couples to gravitational evolution, and thereby introduces further stochastic terms at higher order such as the term $\varepsilon_\delta \delta$ in Eq. (2.33).

The fully general set of stochastic contributions then consists of all terms of the deterministic bias expansion O, with independent stochastic parameter ε_O. Crucially, since the ε_O are uncorrelated with large-scale perturbations, they are *completely described by their joint 1-point PDF* on large scales, or equivalently their real-space moments $\langle \varepsilon_O \varepsilon_{O'} \rangle$, $\langle \varepsilon_O \varepsilon_{O'} \varepsilon_{O''} \rangle$, and so on. Consider as an example the galaxy two-point function, which we will describe in detail in Sec. 4.1. The leading stochastic contribution is

$$\langle \delta_g(\boldsymbol{x}_1)\delta_g(\boldsymbol{x}_2) \rangle \big|_{\text{leading stoch.}} = \langle \varepsilon(\boldsymbol{x}_1)\varepsilon(\boldsymbol{x}_2) \rangle = P_\varepsilon^{\{0\}} \, \delta_D(\boldsymbol{x}_1 - \boldsymbol{x}_2) \,, \tag{2.83}$$

where

$$P_\varepsilon^{\{0\}} \equiv \lim_{k \to 0} \langle \varepsilon(\boldsymbol{k})\varepsilon(\boldsymbol{k}') \rangle' \,. \tag{2.84}$$

is the power spectrum of the field ε in the large-scale limit. Here and throughout, a prime on an expectation value denotes that the momentum-conserving Dirac delta function is to be dropped. Note that in real space, the stochastic contributions are localized at zero lag (but see below).

Next, we consider the galaxy three-point function, whose leading stochastic contributions are

$$\langle \delta_g(\boldsymbol{x}_1)\delta_g(\boldsymbol{x}_2)\delta_g(\boldsymbol{x}_3) \rangle \big|_{\text{leading stoch.}} = \langle \varepsilon(\boldsymbol{x}_1)\varepsilon(\boldsymbol{x}_2)\varepsilon(\boldsymbol{x}_3) \rangle + \{\langle \varepsilon(\boldsymbol{x}_1)(\varepsilon_\delta \delta)(\boldsymbol{x}_2)\delta(\boldsymbol{x}_3) \rangle + 5 \text{ perm.}\}$$

$$= B_\varepsilon^{\{0\}} \, \delta_D(\boldsymbol{x}_1 - \boldsymbol{x}_2)\delta_D(\boldsymbol{x}_2 - \boldsymbol{x}_3) + \left\{ 2 P_{\varepsilon\varepsilon_\delta}^{\{0\}} \, \delta_D(\boldsymbol{x}_1 - \boldsymbol{x}_2)\langle \delta(\boldsymbol{x}_2)\delta(\boldsymbol{x}_3) \rangle + 3 \text{ perm.} \right\} \,, \tag{2.85}$$

where

$$B_\varepsilon^{\{0\}} \equiv \lim_{k,k' \to 0} \langle \varepsilon(\boldsymbol{k})\varepsilon(\boldsymbol{k}')\varepsilon(\boldsymbol{k}'') \rangle' \quad \text{and} \quad P_{\varepsilon\varepsilon_\delta}^{\{0\}} \equiv \lim_{k \to 0} \langle \varepsilon(\boldsymbol{k})\varepsilon_\delta(\boldsymbol{k}') \rangle' \,. \tag{2.86}$$

We see that, when expanding correlators using Wick's theorem, only correlators of the ε_O among themselves remain. This is a consequence of the fact that the ε_O are uncorrelated with the long-wavelength perturbations, such that $\langle \varepsilon_O \delta \rangle = 0$.

Since the coefficients ε_O themselves are first-order perturbations, the complete set of stochastic terms up to third order becomes, working in the Eulerian local basis and thus neglecting higher-derivative operators,

$$
\begin{aligned}
&1^{\text{st}} && \varepsilon \\
&2^{\text{nd}} && \varepsilon_\delta \, \text{tr}[\Pi^{[1]}] \\
&3^{\text{rd}} && \varepsilon_{\text{tr}[\Pi^2]} \, \text{tr}[(\Pi^{[1]})^2] \,, \ \varepsilon_{(\text{tr}\,\Pi)^2} (\text{tr}[\Pi^{[1]}])^2 \,,
\end{aligned}
\tag{2.87}
$$

with an analogous expansion in the Lagrangian case. Note that these terms have already been included in the third-order derivation of Sec. 2.4. Going back to Eq. (2.47) and Eq. (2.54), we see that gravitational evolution mixes the various stochastic terms. This shows that we also have to allow for covariance (cross-correlation) between different stochastic fields ε_O and $\varepsilon_{O'}$, as written in Eq. (2.85). The higher-derivative terms will similarly have stochastic counterparts, for example $\varepsilon_{\nabla^2 \delta} \nabla^2 \delta$.

Finally, we also need to take into account that the galaxy density contrast is not expected to depend on the small-scale perturbations in an exactly local sense, but rather depends on their spatial distribution within a finite region of order R_* (much like its dependence on the long-wavelength perturbations). This implies the necessity of including higher derivatives of the stochastic moments. In Fourier space, this translates into a series expansion in k^2, that is,

$$
\langle \varepsilon_O(\boldsymbol{k}) \varepsilon_{O'}(\boldsymbol{k}') \rangle' = P^{\{0\}}_{\varepsilon_O \varepsilon_{O'}} + P^{\{2\}}_{\varepsilon_O \varepsilon_{O'}} \, k^2 + P^{\{4\}}_{\varepsilon_O \varepsilon_{O'}} \, k^4 + \cdots \,,
\tag{2.88}
$$

where the $P^{\{n\}}_{\varepsilon_O \varepsilon_{O'}}$ are only functions of time. Moreover, following the discussion above, we expect the higher-derivative moments to scale approximately as $|P^{\{n\}}_{\varepsilon_O \varepsilon_{O'}}| \sim R_*^n P^{\{0\}}_{\varepsilon_O \varepsilon_{O'}}$. Note however, that the stochastic correlators arise through the coupling of two or more small-scale (non-perturbative) modes to a large-scale (perturbative) mode. The boundary between these two regimes is the nonlinear scale k_{NL} [Eq. (4.25)]. Thus, stochastic contributions with a different scaling which involves k_{NL} instead of R_* are also possible. We return to observational constraints on the stochastic correlators in Sec. 4.5.3.

In the context of field theory, these terms are referred to as "contact terms", as they formally correspond, in real space, to taking n powers of the Laplacian on a Dirac-delta two-point correlation function. However, care must be taken when deriving the corresponding result in real space. Eq. (2.88) cannot be simply Fourier transformed, since it is a low-k expansion and the Fourier transform relies on the contribution of all k modes. Physically, we expect the correlations of the stochastic terms to be localized to scales of order R_* or less, i.e.

$$
\langle \varepsilon_O(\boldsymbol{x}) \varepsilon_{O'}(\boldsymbol{x} + \boldsymbol{r}) \rangle \xrightarrow{r \gg R_*} 0 \,.
\tag{2.89}
$$

On scales much larger than R_*, they can then be approximated as effective delta functions, $\langle \varepsilon_O(\boldsymbol{x}) \varepsilon_{O'}(\boldsymbol{x} + \boldsymbol{r}) \rangle = P^{\{0\}}_{\varepsilon_O \varepsilon_{O'}} \delta_D(\boldsymbol{r})$, equivalent to keeping the leading term in Eq. (2.88). As discussed above, the real-space counterparts to the sub-leading terms in Eq. (2.88) are formal derivatives of $\delta_D(\boldsymbol{r})$. This reflects the fact that on scales $r \sim R_*$, any perturbative description has to break down, similar to the breakdown of the expansion Eq. (2.88) for $k \sim 1/R_*$.

In order to illustrate these considerations, let us adopt a toy model for the power spectrum of $\varepsilon(\boldsymbol{k})$ that is presumed valid on all scales:

$$
P_\varepsilon(k) \equiv \langle \varepsilon(\boldsymbol{k}) \varepsilon(\boldsymbol{k}') \rangle' = \frac{1}{\overline{n}_g} \left[1 - c \, e^{-k^2 R_*^2} \right] \,,
\tag{2.90}
$$

where $\overline{n}_g$ is the mean galaxy number density and $c \in [0, 1]$ is a constant. For $k \gg 1/R_*$, this approaches the Poisson shot noise $1/\overline{n}_g$, while in the low-k limit, this asymptotes to a smaller value $P^{\{0\}}_\varepsilon = (1 - c)/\overline{n}_g$. Qualitatively, this behavior matches the expectation from halo exclusion, where c corresponds to the

Lagrangian volume fraction occupied by the halos considered [127, 177, 178, 179]. We can Fourier transform this power spectrum to obtain the corresponding two-point correlation function, yielding

$$\xi_\varepsilon(r) = \frac{1}{\overline{n}_g}\left[\delta_D(\boldsymbol{r}) - c(8\pi^{3/2}R_*^3)e^{-r^2/4R_*^2}\right].\tag{2.91}$$

Clearly, this obeys Eq. (2.89). Note that the apparently unphysical delta function will always yield finite results in practice, since the correlation function cannot be measured at strictly zero separation, and is instead integrated over a finite volume. If the volume that is integrated over is much larger than R_*^3, then the variance of ε becomes independent of scale, just as expected for an effective white-noise distribution. Thus, while the stochastic terms contribute to Fourier-space statistics of galaxies on all scales, in real space they only contribute at small separations $r \sim R_*$, and to statistics that involve zero-lag correlators.

Finally, there are also stochastic contributions to galaxy velocities. As in Sec. 2.7, we consider the effective relative velocity between galaxies and matter. In addition to the leading-order deterministic velocity bias $\beta_{\nabla^2\boldsymbol{v}}$, we now include a stochastic contribution:

$$\left[v_g^i - v^i\right](\boldsymbol{x},\tau) = \beta_{\nabla^2\boldsymbol{v}}(\tau)\nabla^2 v^i(\boldsymbol{x},\tau) + \varepsilon_v^i(\boldsymbol{x},\tau),\tag{2.92}$$

where the vector ε_v^i is characterized by the same properties as the stochastic contributions to the galaxy density: it is uncorrelated with the large-scale operators, and completely described by its zero-lag moments. Compared to the leading stochastic field ε in the galaxy density, an additional constraint is imposed on these moments by the equivalence principle. For the same reasons as described for the deterministic velocity bias $\beta_{\nabla^2\boldsymbol{v}}$ in Sec. 2.7, the stochastic contribution to $\boldsymbol{v}_g - \boldsymbol{v}$ has to be constructed from local observables, such as the density and tidal field; this holds both for gravitational and non-gravitational sources of $\boldsymbol{v}_g - \boldsymbol{v}$. Thus, on large scales the relative velocity can always be written as a total derivative of a nonlinear function of the density and tidal field: $[v_g^i - v^i](\boldsymbol{x}) = \partial^i F(\delta(\boldsymbol{x}), K_{lm}(\boldsymbol{x}))$. This fact implies that the power spectrum of the stochastic relative velocity between galaxies and matter has the following scaling on large scales:

$$\lim_{k\to 0}\langle\varepsilon_v^i(\boldsymbol{k})\varepsilon_v^j(\boldsymbol{k}')\rangle' = k^i k^j P_{\varepsilon_v}^{\{2\}},\tag{2.93}$$

where $P_{\varepsilon_v}^{\{2\}}$ is a positive constant. That is, there is no white-noise term $P_{\varepsilon_v}^{\{0\}}$ unlike for ε. Similarly, we have, for the cross-correlation with ε, $\lim_{k\to 0}\langle\varepsilon_v^j(\boldsymbol{k})\varepsilon(\boldsymbol{k}')\rangle' = ik^j P_{\varepsilon\varepsilon_v}^{\{1\}}$. The stochastic velocity bias plays an important role for galaxy statistics in redshift space [180].

Above, we have argued that $P_\varepsilon^{\{2\}} \sim R_*^2 P_\varepsilon^{\{0\}} \sim R_*^2/\overline{n}_g$, if stochasticity in the galaxy number density is close to Poisson. We now give a rough order-of-magnitude estimate for the stochastic amplitudes of the velocity. First, for $P_{\varepsilon\varepsilon_v}^{\{1\}}$, one can follow the reasoning described after Eq. (2.82), to roughly estimate that this term should not be larger than of order $R_*^2\mathcal{H}P_{\varepsilon\varepsilon}^{\{0\}}$, and, similarly, $P_{\varepsilon_v}^{\{2\}} \lesssim R_*^4\mathcal{H}^2 P_{\varepsilon\varepsilon}^{\{0\}}$. We can provide a sharper estimate by adopting a concrete physical model. Let us assume that the stochastic velocity contribution of a given galaxy sample is due to the virial velocities within the host halos of mass M and Eulerian radius R_E of these galaxies. The variance of virial velocities is $\sigma_{v,\mathrm{vir}}^2 \approx GM/R_E$. Matching this variance to Eq. (2.93) yields

$$\sigma_{v,\mathrm{vir}}^2 = \int_{\boldsymbol{k}} P_{\varepsilon_v}(k) \sim \frac{1}{2\pi^2}\int_0^{1/R_E} k^4 dk\, P_{\varepsilon_v}^{\{2\}} \quad\Rightarrow\quad P_{\varepsilon_v}^{\{2\}} \sim \frac{GM}{R_E}R_E^5.\tag{2.94}$$

Here, we have cut off the integral at $k \sim 1/R_E$, and neglected all prefactors since this is a very rough estimate. Since $\Phi \sim GM/R_E$ is approximately mass-independent, we find that $P_{\varepsilon_v}^{\{2\}}$ scales as R_E^5, or $M^{5/3}$. This is the same scaling as dynamical mass estimates based on velocity statistics (e.g., [181]).

Including the stochastic terms listed here completes the parametrization of bias under the assumptions listed in the beginning of Sec. 2.5: General Relativity, a non-relativistic matter fluid, and adiabatic Gaussian initial conditions. The general bias expansion is summarized in Sec. 2.11.

2.9 Galaxy bias in the relativistic context

So far in this section, we have restricted our discussion of galaxy bias to sub-horizon scales $k \gg aH$, where most of the information in large-scale structure resides. This is because most of the measured modes are small-scale modes, while a survey with comoving volume of order H_0^{-3} will only measure a handful of modes with $k \sim H_0$. However, future large-scale galaxy surveys will have sufficiently large volume to probe such scales. We now consider how the bias can be extended to describe the galaxy density contrast on arbitrarily large scales.

It is important to note that the leading corrections to the sub-horizon treatment of bias, which describes the rest-frame galaxy density, scale as $(aH/k)^2$, and are thus very small unless we consider very large scales. On those large scales, linear perturbation theory is an excellent approximation. Hence, for most practical purposes, it is sufficient to consider relativistic effects at linear order. In fact, the only requirement necessary to make the bias expansion consistent in the relativistic context at linear order is that the time slicing (constant-time hypersurface) chosen to perform the bias expansion must correspond to a constant proper time of comoving observers, which is realized by working in the synchronous-comoving gauge [182, 95, 96] (Ref. [183, 184] showed how to extend this to second order in relativistic perturbations). The reason is that, apart from the local density, the galaxy density depends on the local age of the Universe, or proper time along the fluid trajectory up to the time of observation. In other words, at fixed proper time, galaxies are at a fixed evolutionary stage, in which case the bias expansion in terms of the galaxy density is sufficient at linear order.

The conformal Fermi coordinates (CFC) introduced in [145, 146] provide a useful way of defining bias operators in the relativistic context. The CFC are the natural coordinates a cosmologist living in a given galaxy would use. Briefly, the construction of CFC starts from a time-like geodesic, namely the fluid trajectory $x_{\mathrm{fl}}^\mu(t_F)$ parametrized by the proper time t_F, with tangent vector $U^\mu \equiv \partial x_{\mathrm{fl}}^\mu(t_F)/\partial t_F$. This frame can be constructed locally for any spacetime, as long as a geodesic congruence exists within a neighborhood of the geodesic considered, and is thus not restricted to the assumption of a perturbed FRW spacetime. However, in the actual Universe, which has perturbations on all scales, we perform the construction on a coarse-grained metric which only has contributions from Fourier modes below some cutoff $k < \Lambda$. Then, the velocity 4-divergence $\vartheta \equiv \nabla_\mu U^\mu$ along the fluid defines the local Hubble rate $H_F(t_F) \equiv \vartheta/3$. The Hubble rate can be integrated to yield the local scale factor $a_F(t_F)$ (up to an arbitrary overall normalization). One can then construct a coordinate system with time coordinate τ_F defined through $a_F d\tau_F = dt_F$ so that the metric becomes

$$g_{\mu\nu}^F(x_F) = a_F^2(\tau_F) \left[\eta_{\mu\nu} + A_{\mu\nu,kl}(\tau_F) x_F^k x_F^l + \mathcal{O}(\boldsymbol{x}_F^3) \right] . \tag{2.95}$$

That is, *along the entire trajectory $x_{\mathrm{fl}}^\mu(\tau)$*, all large-scale cosmological perturbations ($k < \Lambda$) are absorbed into the local scale factor a_F, and the corrections $\propto \boldsymbol{x}_F^2$, $A_{\mu\nu,kl}$. Note that $A_{00,kl}$ is trace-free with respect to kl, $A_{00,kl}\delta^{kl} = 0$.

For nonrelativistic tracers such as galaxies, the spatial components of the metric are dynamically irrelevant, and the CFC spacetime is characterized completely by the local expansion rate $H_F(\tau_F)$ and the purely tidal perturbation to the time-time component of the metric, $K_F^{ij} \equiv A_{00,ij}$. These are the same 6 degrees of freedom as found in the quasi-Newtonian description described in the previous sections. Thus, a fully relativistic basis of bias operators is given by all scalar combinations of H_F and K_F^{ij}, as well as their derivatives with respect to τ_F and x_F^i. The latter are suppressed by the spatial scale R_* as before.

To illustrate this, let us consider an adiabatic perturbation at linear order. As discussed above, this is the most relevant case. As shown in [185], the isotropic part of the perturbation is absorbed into a_F, which then obeys the Friedmann equation,

$$H_F^2 = \frac{8\pi G}{3}\varrho_F + \frac{\mathrm{K}_F}{a_F^2} , \tag{2.96}$$

in terms of the *local physical* (not comoving) CFC-frame matter density ϱ_F and curvature $\mathrm{K}_F = \mathrm{const}$. That is, the local matter density together with the initial conditions (curvature) completely specifies the isotropic part of the spacetime along the entire fluid trajectory. The CFC approach thus provides one proof of the *separate universe* conjecture [186, 187, 188, 95, 189]. Moreover, K_F is directly related to the matter density

perturbation δ_{sc} in synchronous-comoving gauge by [95, 185]

$$\frac{\mathrm{K}_F}{H_0^2} = \frac{5}{3}\Omega_{m0}\frac{\delta_{\mathrm{sc}}(\tau)}{D_{\mathrm{md}}(\tau)}\,, \tag{2.97}$$

where $D_{\mathrm{md}}(\tau)$ is the linear growth factor normalized to $a(\tau)$ during matter domination (for $a(\tau) \ll 1$). δ_{sc} is in turn proportional to $\nabla^2\Phi$ in conformal-Newtonian gauge. This shows that, at linear order, the physical variable with respect to which galaxies are biased is δ_{sc} or equivalently $\nabla^2\Phi$ [190, 95, 96]. Note that the separate-universe picture holds fully nonlinearly for isotropic perturbations. For such configurations, the complete bias expansion thus consists of powers of K_F or equivalently ϱ_F (and spatial derivatives thereof). This provides another interpretation for the fact that we do not have to include convective time derivatives of δ in the bias expansion. On the other hand, the anisotropic part is encoded in the trace-free perturbations $A_{00,kl} = K_F^{kl}$ and $A_{ij,kl} = \delta_{ij}K_F^{kl}$, while $A_{0i,kl} = 0$.

Beyond linear order, there are nontrivial relativistic effects. First, relativistic effects due to the motion of galaxies enter at order $(v/c)^2 \sim 10^{-5}(v/[1000\,\mathrm{km\,s^{-1}}])^2$. Second, the galaxy density couples to vector and tensor metric perturbations, which in turn are produced by large-scale structure at nonlinear order [191, 192, 193] (and primordially, in case of tensor modes); one can show [150] that in the CFC frame, the nonlinearly generated vector and tensor modes only start to contribute at third order in perturbations.[9] These "genuine" relativistic effects in large-scale structure, while interesting, are most likely too small to be of observational relevance for current and upcoming surveys.

Apart from such genuine relativistic effects, the results of this section give a rigorous description of galaxy clustering in relativistic perturbation theory *in a specific frame*, the CFC. In order to connect to observations, we then have to transform the galaxy density in CFC to observed coordinates. This is straightforward to do, given that the galaxy number density transforms as the 0-component of a four-vector, namely the galaxy current vector j_g^μ. The coordinate transformation is determined by tracing photon geodesics from the galaxy's position to the observer. We shall describe this in Sec. 9.3.

2.10 Renormalization: bare vs. physical bias parameters*

2.10.1 Motivation

So far in Sec. 2, we have enumerated which bias terms to include to express the galaxy density in an expansion of the form

$$\delta_g(\boldsymbol{x}, \tau) = \sum_O [b_O(\tau) + \varepsilon_O(\boldsymbol{x}, \tau)]\, O(\boldsymbol{x}, \tau) + \varepsilon(\boldsymbol{x}, \tau)\,. \tag{2.98}$$

However, there is a subtlety when relating this bias expansion to the statistics of galaxies. To understand this issue, let us disregard nonlinear gravitational evolution for the moment, and go back to the case of LIMD bias in Lagrangian space considered in Sec. 2.1. We start from the LIMD bias expansion Eq. (2.10), now written as

$$\delta_g^L(\boldsymbol{q}) = c_1\delta_R(\boldsymbol{q}) + \frac{1}{2}c_2\left(\delta_R^2(\boldsymbol{q}) - \sigma^2(R)\right) + \frac{1}{6}c_3\delta_R^3(\boldsymbol{q}) + \cdots\,, \tag{2.99}$$

where we have dropped the superscript (1) on the matter density δ_R for clarity, and denoted the Lagrangian bias parameters with c_i, for a reason that will become clear shortly. Now, given that $\delta_R = \delta_R^{(1)}$ is Gaussian, we can easily work out the two-point function of δ_g^L in Lagrangian space, given Eq. (2.99):

$$\xi_g^L(r) = \left[c_1^2 + c_1c_3\sigma^2(R)\right]\xi_{\mathrm{L},R}(r) + \frac{1}{2}c_2^2[\xi_{\mathrm{L},R}(r)]^2 + \cdots\,. \tag{2.100}$$

---

[9]In case of galaxy shapes, *primordial* tensor modes actually lead to effects at linear order which can become relevant on large scales [194, 195].

* This section is of a more technical nature and is not essential for the remainder of the review. However, we encourage readers to go through the non-technical introduction, Sec. 2.10.1.

We immediately see a difference to the expansion Eq. (2.7) that we obtained from an explicit calculation of the pair probability in Sec. 2.1: while in Eq. (2.7), the coefficient of the leading term $\xi_{\mathrm{L},R}(r)$ is $(b_1^L)^2$, here we have a coefficient $c_1(c_1 + c_3\sigma^2(R))$. Moreover, if we were to continue the expansion in Eq. (2.99) to higher orders, we would obtain additional contributions to $\xi_g^L(r)$ that are proportional to $c_1 c_{2n+1}\sigma^{2n}(R)\xi_{\mathrm{L},R}(r)$ and thus contribute at leading order on large scales. This is clearly in contradiction with the spirit of the perturbative bias expansion, which is based on the principle that the contributions of higher-order bias terms are suppressed on large scales. Further, if we choose a small smoothing scale R such that $\sigma(R)$ is not much less than 1, then these higher-order contributions change the amplitude of $\xi_g^L(r)$ by order one. This contradicts Sec. 2.6, where we have argued that any smoothing scale involved should be irrelevant on large scales.

On the other hand, Eq. (2.7) shows exactly the desired behavior, with higher-order bias terms (as well as the effects of smoothing) being suppressed on large scales. The solution to the undesirable situation posed by Eq. (2.100) is that the coefficients c_n in the bias expansion Eq. (2.99) are not physical, but "bare" bias parameters. Instead, the physical bias parameters are introduced as coefficients of *renormalized operators*. In the present case,

$$c_1\delta_R \to b_1^L[\delta_R] \equiv b_1^L\delta_R\,; \quad c_3\delta_R^3 \to b_3^L[\delta_R^3] \equiv b_3^L(\delta_R^3 - 3\sigma^2(R)\delta_R)\,. \qquad (2.101)$$

Inserting these relations into Eq. (2.99), we see that this removes the undesired contribution to $\xi_g^L(r)$, and we recover $\xi_g^L(r) = (b_1^L)^2\xi_{\mathrm{L},R}(r)$ in the large-scale limit. An equivalent approach is to define $b_1^L \equiv c_1 + c_3\sigma^2(R)/2$, as done in [76]. Here, the scale R should really be seen as an arbitrary smoothing scale (denoted as Λ^{-1} below) whose value becomes irrelevant in the end. Thus, renormalization is an essential part of the connection between bias expansion and galaxy statistics.

In the case of a Gaussian density field, the renormalized LIMD operators are simply given at all orders by Hermite polynomials [99],

$$[\delta_N] = \sigma^N(R)H_N\left(\frac{\delta_R}{\sigma(R)}\right)\,. \qquad (2.102)$$

Suitable orthogonal polynomials can be defined for other operators, such as $(\boldsymbol{\nabla}\delta)^2$ [196]. This approach is discussed in Sec. 6.6.1 in the context of peak theory, but it is applicable to any Lagrangian bias scheme. However, this method cannot be directly applied to the evolved, Eulerian density field as the latter is highly non-Gaussian on small scales. Nevertheless, even in this case, renormalized operators can be systematically derived order by order in perturbation theory. We stress again that all physical statements about the values of bias parameters, for example a large bias for rare objects, refer to the *physical, renormalized bias parameters*, i.e. the coefficients of the renormalized operators.

The aim of this section is to describe in detail the renormalization procedure that connects the bias expansion derived so far for the *evolved* galaxy density field to the observable statistics of galaxies on large scales. In the course of this derivation, we make the physical arguments of the previous sections more formal and rigorous, and connect to the EFT language ([143, 83, 88]; see [144] for a review). Renormalization in the context of galaxy bias was first considered by [76], and later expanded on in [100, 197, 88, 131].

2.10.2 Equivalence principle and local gravitational observables

In essence, the perturbative bias expansion attempts to connect the proper rest-frame density of galaxies at some time τ and position $\boldsymbol{x}$ to the initial conditions (for example, produced by inflation) at time $\tau = 0$. In full generality, the proper comoving density n_g can be written as

$$n_g(\boldsymbol{x},\tau) = F_g[\Phi^{(1)}(\boldsymbol{y})](\boldsymbol{x},\tau)\,, \qquad (2.103)$$

where F_g is a nonlinear functional of the initial potential perturbations $\Phi^{(1)}(\boldsymbol{y}) = \Phi(\boldsymbol{y},\tau = 0)$. The assumption of statistical homogeneity of the Universe dictates that the functional be invariant under a constant spatial coordinate shift $\boldsymbol{x} \to \boldsymbol{x}+\boldsymbol{\xi}$. This implies that F_g only involves the combination $\boldsymbol{y} - \boldsymbol{x}$. More generally, n_g, being the 0-component of the current vector j_g^μ, has specific transformation properties under general coordinate (or gauge) transformations. In particular, it transforms as a 3-scalar under coordinate

rotations on a fixed time slice, a property we have already used from the beginning (e.g., when introducing the term $(K_{ij})^2$ in Sec. 2.3).

A particular *generalized Galilean* coordinate transformation, which corresponds to a time-dependent but spatially uniform coordinate shift, is of special relevance to large-scale structure [198, 199, 200]. In General Relativity, the equivalence principle holds, which implies that we can remove a pure-gradient metric perturbation by going to the free-falling frame of comoving observers, who move on trajectories $\boldsymbol{x}_{\mathrm{fl}}(\tau')$ with $\tau' \in [0, \tau]$. Specifically, for the conformal-Newtonian gauge metric Eq. (1.2), a pure-gradient potential perturbation,

$$\Phi(\boldsymbol{x}, \tau) = \Psi(\boldsymbol{x}, \tau) = \boldsymbol{A}(\tau) \cdot \boldsymbol{x}, \tag{2.104}$$

can be removed by performing a spatially constant but *time-dependent* spatial translation,

$$\boldsymbol{x} \to \boldsymbol{x} + \boldsymbol{\xi}(\tau); \quad \boldsymbol{v} \to \boldsymbol{v} + \partial_\tau \boldsymbol{\xi}(\tau). \tag{2.105}$$

Under this translation, the 00-component of the metric transforms as

$$\Phi \to \Phi - (\partial_\tau^2 \boldsymbol{\xi} + \mathcal{H} \partial_\tau \boldsymbol{\xi}) \cdot \boldsymbol{x}. \tag{2.106}$$

Thus, if we demand that $\boldsymbol{\xi}(\tau)$ solves

$$\frac{\partial}{\partial \tau^2} \boldsymbol{\xi} + \mathcal{H} \frac{\partial}{\partial \tau} \boldsymbol{\xi} = \boldsymbol{\nabla} \Phi(\tau) = \boldsymbol{A}(\tau), \tag{2.107}$$

then $\boldsymbol{\nabla}\Phi$ vanishes in the transformed coordinates at all times. We recognize Eq. (2.107) as the equation for (minus) the Lagrangian displacement $\boldsymbol{\xi} = -\boldsymbol{s}$, Eq. (2.25). That is, $\boldsymbol{x} + \boldsymbol{\xi}(\tau)$ corresponds to the Lagrangian coordinate of a fluid trajectory in the presence of a long-wavelength (pure-gradient) mode. An observer comoving with this trajectory will experience no gravitational force, as $\boldsymbol{\nabla}\Phi = 0$. Clearly however, we can only remove first spatial derivatives of the potential, while second spatial derivatives lead to locally observable gravitational effects. This is the physical content of the well-known *consistency relations* in large-scale structure [201, 202, 200, 203, 204, 205], which phrase the requirement that the local physics must be invariant under time-dependent translations as conditions relating certain limits of N-point functions to $(N-1)$-point functions. Following our discussion, the consistency relations hold for biased tracers as well [206, 207], on scales much larger than R_*.

Working in the fluid coordinates, we can then write F_g fully generally as

$$n_g(\boldsymbol{x}, \tau) = F_g\Big[\partial_i \partial_j \Phi(\boldsymbol{x}_{\mathrm{fl}}'(\tau'), \tau')\Big](\boldsymbol{x}, \tau), \tag{2.108}$$

where F_g now only involves the difference $\boldsymbol{x}_{\mathrm{fl}}' - \boldsymbol{x}_{\mathrm{fl}}$ at various times τ' (with $0 \le \tau' \le \tau$). We could have equivalently written Eq. (2.108) in terms of $\partial_{q,i}\partial_{q,j}\Phi(\boldsymbol{q}, \tau)$. This nonlinear functional of $\partial_i\partial_j\Phi$ is significantly more restrictive than the full functional $F_g[\Phi]$. Moreover, since $\partial_i\partial_j\Phi$ corresponds to the leading locally observable effect of the gravitational field, this way of writing the functional isolates the actual physical impact of long-wavelength perturbations on the galaxy density. On the other hand, $F_g[\Phi^{(1)}(\boldsymbol{y})]$ contains a large number of unphysical gauge modes. The price we have payed for this reduction in degrees of freedom is that we need to include the dependence on $\partial_i\partial_j\Phi$ along the entire fluid trajectory. Physically, this makes sense, since, as argued in Sec. 2.5.1, the formation of halos and galaxies is not an instantaneous process and thus depends on the local environment throughout cosmic time. We also see that the dependence is now on the full potential Φ rather than the initial potential $\Phi^{(1)}$, a point to which we will return below.

The requirement that n_g be a spatial scalar is then simply achieved by only allowing terms where all indices in the factors of $\partial_i\partial_j\Phi$ are contracted. We stress again that any dependence on other local observables, such as the matter density or velocity shear, and time derivatives thereof, are implicitly included in the arguments listed in Eq. (2.108), as these can themselves be expressed as integrals over $\partial_i\partial_j\Phi(\boldsymbol{x}_{\mathrm{fl}}')$ via the equations of motion. One might wonder whether the relation Eq. (2.108) depends on when the spatial derivatives are taken, e.g. in Eulerian frame ($\partial_{x,i}\partial_{x,j}$) or Lagrangian frame ($\partial_{q,i}\partial_{q,j}$). However, the different spatial derivatives can be transformed into each other (see Tab. 5), where the transformation itself just

depends on $\partial_i\partial_j\Phi$ along the fluid trajectory. Hence, the expansion is independent of the time slice chosen for the spatial derivatives.

Further, following our physical arguments in Sec. 2.5.1, there is a spatial scale R_* such that the functional kernel in F_g becomes negligible if $|\boldsymbol{x}'_{\mathrm{fl}} - \boldsymbol{x}_{\mathrm{fl}}| \gg R_*$. In general, R_* is a function of time τ'; in this case, what we refer to as R_* should always be seen as $\max_{\tau'}[R_*(\tau')]$.

We now expand the functional in Eq. (2.108) in time. For this we consider an operator $O(\boldsymbol{x}'_{\mathrm{fl}}(\tau'), \tau')$ which is composed of any nonlinear combination of $\partial_i\partial_j\Phi(\boldsymbol{x}'_{\mathrm{fl}}(\tau'), \tau')$ (we will deal with the functional dependence on $\boldsymbol{x}'_{\mathrm{fl}}$ below). The time dependence then formally is [as in Eq. (2.57)]

$$\int d\tau'\, F_O(\tau; \tau') O(\boldsymbol{x}'_{\mathrm{fl}}(\tau'), \tau') = \sum_{n=0}^{\infty} \frac{1}{n!} \left[\int d\tau' (\tau' - \tau)^n F_O(\tau; \tau') \right] \left(\frac{\mathrm{D}}{\mathrm{D}\tau} \right)^n O(\boldsymbol{x}'_{\mathrm{fl}}(\tau), \tau), \qquad (2.109)$$

where we have expanded O in terms of convective time derivatives around $\tau' = \tau$, and $\boldsymbol{x}'_{\mathrm{fl}}(\tau)$ denotes a fluid trajectory in the vicinity of $\boldsymbol{x}_{\mathrm{fl}}(\tau)$. Of course, we could have equivalently expanded around $\tau = 0$. The factors in brackets are coefficients which only depend on τ. With this, the functional Eq. (2.108) can be written *on a single time slice*, by including convective derivatives of all combinations of $\partial_i\partial_j\Phi$:

$$n_g(\boldsymbol{x}, \tau) = F_g\Big[\partial_i\partial_j\Phi(\boldsymbol{x}', \tau), \mathrm{D}\partial_i\partial_j\Phi(\boldsymbol{x}', \tau)/\mathrm{D}\tau, \ldots; \tau \Big](\boldsymbol{x}), \qquad (2.110)$$

which is understood to also include nonlinear mixed terms such as $(\partial_k\partial_l\Phi)\mathrm{D}(\partial^k\partial^l\Phi)/\mathrm{D}\tau$ [including these terms captures functionals involving products of operators such as $\int d\tau' \int d\tau'' F(\tau; \tau', \tau'') O_1(\tau') O_2(\tau'')$]. Eq. (2.110) is still a functional in terms of the spatial dependence $\boldsymbol{x}'$, but is local in time. Again, any reference time $\tilde\tau \neq \tau$ could be chosen for the expansion on the right-hand side, as long as $\boldsymbol{x}$ is replaced with $\boldsymbol{x}_{\mathrm{fl}}(\tilde\tau)$.

2.10.3 Coarse graining and bare bias expansion

In the context of the perturbative bias expansion, our goal is to isolate the dependence of $n_g(\boldsymbol{x}, \tau)$ on large-scale perturbations, with the ultimate goal of deriving statistics of n_g for some large scale r, or Fourier wavenumber $k \propto 1/r$. The theory itself, once applied to measurements, will tell us what "large" here means precisely.

Let us coarse-grain, or filter, Eq. (2.110) on a spatial scale Λ^{-1} which is much smaller than r, or, equivalently, $\Lambda \gg k$ for all wavenumbers k of interest (see Fig. 8). Since this is a crucial step, we carefully describe it here. We let $W_\Lambda(|\boldsymbol{x}|)$ denote the filtering kernel, assumed to be isotropic and normalized to unity (which is a natural assumption since any anisotropy would correspond to introducing preferred directions). We denote any filtered quantity with a subscript Λ, e.g. Φ_Λ. In EFT language, Λ provides our UV (ultraviolet; high-energy, or small-scale) cutoff, introduced to render loop integrals finite. We can think of $\delta_\Lambda(\boldsymbol{x}) \propto \nabla^2\Phi_\Lambda$ as the average density perturbation within a region $\mathcal{U}$ of size Λ^{-1} centered on $\boldsymbol{x}$ (dotted region in Fig. 8). Complementary to Φ_Λ, we define the small-scale potential perturbations as

$$\Phi_s(\boldsymbol{x}, \tau) \equiv \Phi(\boldsymbol{x}, \tau) - \Phi_\Lambda(\boldsymbol{x}, \tau). \qquad (2.111)$$

Note that the cutoff Λ is merely a computational tool whose value we are free to determine; any observable, including statistics of galaxies as well as the physical bias parameters, have to be independent of the value of Λ. This can in fact be used as a sanity check for any predictions for observables. Thus, Λ^{-1} must be distinguished from the *physical* scale R_* which controls the amplitude of the higher derivative bias parameters discussed in Sec. 2.6, and is, for halos, typically of order the Lagrangian radius. In fact, demanding that galaxy statistics be independent of Λ forces us to introduce precisely these higher derivative terms with physical renormalized coefficients, as we will see in Sec. 2.10.5. However, going step by step, we will again start by neglecting higher-derivative terms. In this case, we implicitly assume that Λ^{-1} is larger than R_*.

Note that we could have equally chosen to perform the coarse-graining in terms of the initial potential perturbations $\Phi^{(1)}$. The nonlinear gravitational potential in Eulerian coordinates, $\Phi(\boldsymbol{x}, \tau)$, and $\Phi^{(1)}(\boldsymbol{q})$ are nonlinearly related. Their relation, however, can be expanded in terms of the same operators as we will
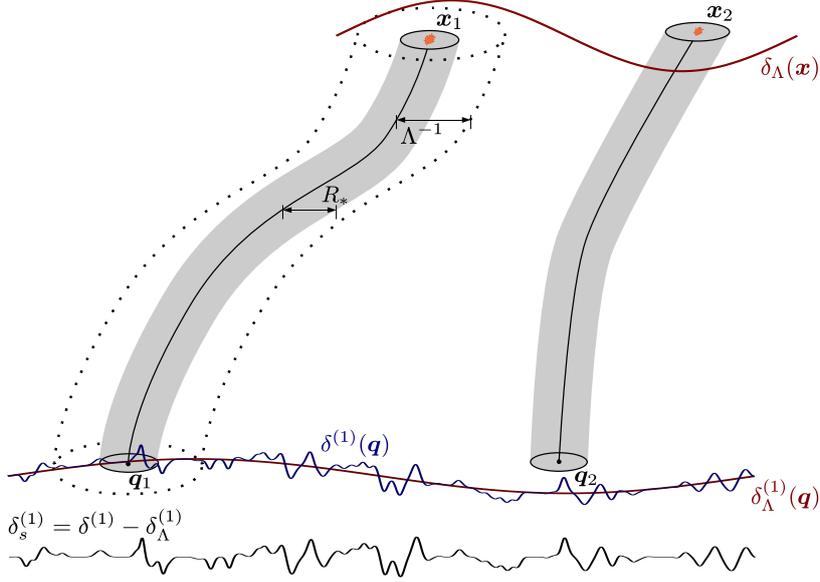
Figure 8: Sketch of the setup and quantities used in the derivation of bias renormalization. Time runs vertically as in Fig. 6. The observed galaxies reside at Eulerian positions $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ and each form within a spacetime region centered around the fluid trajectories to each point as in Fig. 6. The top of the plot corresponds to the epoch of observation, while the bottom denotes the initial condition (Lagrangian positions $\boldsymbol{q}_1$, $\boldsymbol{q}_2$). The blue line near the bottom shows the total initial (linear) density field. We coarse-grain the galaxy and matter fields on the scale Λ^{-1} (dotted region), resulting in a smoothed large-scale density field δ_Λ shown by red solid lines. The black line at the bottom shows the small-scale density field in the initial conditions, which is statistically the same everywhere. For clarity, we only represent the density field, although in reality the galaxy density is a function of all components of $\partial_i \partial_j \Phi$ as well as its convective time derivatives.

include in the bias expansion. Since coarse-graining (convolution) and multiplication do not commute, there are differences in the result obtained by coarse-graining in $\Phi^{(1)}$; however, these differences are at higher order in spatial derivatives, and will be absorbed by higher derivative operators (Sec. 2.10.5).

If we make Λ sufficiently small (that is, the smoothing scale sufficiently large), then the functional dependence on $\partial_i \partial_j \Phi$ in Eq. (2.110) separates into an *ordinary* [as in the local expansion in Eq. (2.56)] dependence on the variable $\partial_i \partial_j \Phi_\Lambda(\boldsymbol{x})$, while the functional dependence is restricted to $\partial_i \partial_j \Phi_s(\boldsymbol{x}')$:

$$n_{g,\Lambda}(\boldsymbol{x}) = F_{g,\Lambda}\left[\partial_i \partial_j \Phi_\Lambda(\boldsymbol{x}),\, \mathrm{D}\partial_i \partial_j \Phi_\Lambda(\boldsymbol{x})/\mathrm{D}\tau,\, \ldots\, ;\; \partial_i \partial_j \Phi_s(\boldsymbol{x}'),\, \mathrm{D}\partial_i \partial_j \Phi_s(\boldsymbol{x}')/\mathrm{D}\tau,\, \ldots\right], \qquad (2.112)$$

where here and in the following we suppress the time arguments τ for clarity, since they are the same on both sides of Eq. (2.112). We now prove this statement. First, Eq. (2.112) holds if and only if Λ^{-1} is much larger than R_*, the spatial scale over which the functional in Eq. (2.110) has significant support. We now expand Eq. (2.110) into a series of linear, quadratic, and higher-order functionals, in analogy with the expansion of the functional in time Eq. (2.109). By inserting $\Phi = \Phi_\Lambda + \Phi_s$ [Eq. (2.111)], any term without time derivatives in this expansion can be written as

$$\left[\prod_{i=1}^{n} \int d^3 \boldsymbol{x}_i\right] F^{(n)}_{i_1 j_1 \cdots i_n j_n}(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n) \partial_{i_1} \partial_{j_1} \Phi(\boldsymbol{x}_{\mathrm{fl}} + \boldsymbol{x}_1) \cdots \partial_{i_n} \partial_{j_n} \Phi(\boldsymbol{x}_{\mathrm{fl}} + \boldsymbol{x}_n) \qquad (2.113)$$

$$\simeq \sum_{k=0}^{n} \frac{n!}{(n-k)!} \partial_{i_1} \partial_{j_1} \Phi_\Lambda(\boldsymbol{x}_{\mathrm{fl}}) \cdots \partial_{i_k} \partial_{j_k} \Phi_\Lambda(\boldsymbol{x}_{\mathrm{fl}}) \left\{ \left[\prod_{i=1}^{n} \int d^3 \boldsymbol{x}_i\right] F^{(n)}_{i_1 j_1 \cdots i_n j_n}(\boldsymbol{x}_1, \cdots \boldsymbol{x}_n) \right.$$

$$\left. \times\, \partial_{i_{k+1}} \partial_{j_{k+1}} \Phi_s(\boldsymbol{x}_{\mathrm{fl}} + \boldsymbol{x}_{k+1}) \cdots \partial_{i_n} \partial_{j_n} \Phi_s(\boldsymbol{x}_{\mathrm{fl}} + \boldsymbol{x}_n) \right\},$$

53

where we have approximated $\Phi_\Lambda(\boldsymbol{x}_{\mathrm{fl}}[\tau])$ as constant over the region over which the kernel $F^{(n)}$ is nonzero, as implied by the condition $\Lambda^{-1} \gg R_*$. This allowed us to pull the factors of $\Phi_\Lambda(\boldsymbol{x}_{\mathrm{fl}} + \boldsymbol{x}_i)$ outside the integral. In Sec. 2.10.5 we will show that going beyond this approximation leads to the higher-derivative terms discussed in Sec. 2.6. Further, we have assumed without loss of generality that the kernels $F^{(n)}_{i_1 j_1 \cdots i_n j_n}$ are symmetrized over the indices $i_k j_k$ and the associated spatial positions $\boldsymbol{x}_k$. The same reasoning also goes through for terms with time derivatives. If we think of the spatial derivatives as being with respect to $\boldsymbol{q}$, i.e. Lagrangian derivatives, then $\mathrm{D}/\mathrm{D}\tau$ commutes with ∂_i. For Eulerian spatial derivatives, this is not the case, however the term $v^i \partial_i$ in the convective derivative cancels via the time dependence that appears in the spatial argument $\boldsymbol{x}_{\mathrm{fl}}(\tau)$ of all instances of the potential Φ.

The terms in curly brackets in Eq. (2.113) correspond to weighted integrals over the small-scale fluctuations. It is then clear that we can write Eq. (2.112) as

$$n_{g,\Lambda}(\boldsymbol{x}) = \sum_{1,\, O_\Lambda} c_{O,\Lambda} \left[\partial_i \partial_j \Phi_s, \, \mathrm{D}\partial_i \partial_j \Phi_s/\mathrm{D}\tau, \, \ldots \right] O_\Lambda(\boldsymbol{x}) \,, \tag{2.114}$$

where the coefficients $c_{O,\Lambda}[\partial_i \partial_j \Phi_s, \cdots]$ are still functionals of the small-scale modes $\partial_i \partial_j \Phi_s$. The sum runs over 1 and all scalar operators constructed out of $\partial_i \partial_j \Phi_\Lambda$ and its time derivatives, i.e. *precisely the set described in Sec. 2.5.1*, with a zeroth-order constant related to the mean density of the galaxies. However, here the operators are constructed out of the coarse-grained quantity $\partial_i \partial_j \Phi_\Lambda$. We have made this explicit through the subscript Λ.

Now, by construction, the short modes have no support at low k in Fourier space. Specifically, $\Phi_s(\boldsymbol{k}) = [1 - W_\Lambda(\boldsymbol{k})]\Phi(\boldsymbol{k}) \xrightarrow{k \to 0} \mathcal{O}(k^2 \Lambda^2)\Phi(\boldsymbol{k})$. The small-scale modes in the initial conditions $\Phi_s^{(1)}$ are then independent of long-wavelength modes $\Phi_\Lambda^{(1)}$ for Gaussian initial conditions (see lower panel of Fig. 8). At finite time, they do depend on long-wavelength modes through the gravitational influence of the latter. However, the gravitational effects on the evolution of the small-scale modes are also precisely captured by $\partial_i \partial_j \Phi_\Lambda$ and its convective time derivatives, i.e. the operators O_Λ that appear in the sum in Eq. (2.114). This can be shown for example by considering the Lagrangian evolution equations of the matter density field (Sec. 4 of [131]). Then, at any given time, we can expand the evolved small-scale statistics and their time derivatives in terms of the long wavelength operators O_Λ and the initial small-scale modes $\Phi_s^{(1)}$ in Eq. (2.114). Finally, we reorder the sum to write the coefficients $c_{O,\Lambda}$ as functionals of the small-scale modes in the *initial* conditions:

$$c_{O,\Lambda} \left[\partial_i \partial_j \Phi_s, \, \mathrm{D}\partial_i \partial_j \Phi_s/\mathrm{D}\tau, \, \ldots \right] \to c_{O,\Lambda}[\partial_i \partial_j \Phi_s^{(1)}] \,. \tag{2.115}$$

Note that we can only do this reordering consistently if the basis contains all operators that appear in the solution for the gravitational evolution of the small-scale modes Φ_s up to a given order in perturbation theory. Only in this case is the operator basis closed under renormalization, which we will discuss below. Since, on large scales, the small-scale initial conditions $\Phi_s^{(1)}$ are uncorrelated with each other as well as the coarse-grained operators, the dependence of $c_{O,\Lambda}$ on $\partial_i \partial_j \Phi_s^{(1)}$ can be completely described by stochastic fields $\varepsilon_{\Lambda,O}$; that is, fields that are entirely characterized by their one-point PDF. This is of course only valid on scales much larger than Λ^{-1}. We then obtain the expression

$$n_{g,\Lambda}(\boldsymbol{x}) = n_{0,\Lambda} \left[1 + \varepsilon_\Lambda(\boldsymbol{x}) + \sum_O \left\{ c_{O,\Lambda} + \varepsilon_{O,\Lambda}(\boldsymbol{x}) \right\} O_\Lambda(\boldsymbol{x}) \right] \,, \tag{2.116}$$

where the sum runs over the operators described in Sec. 2.5, and $\varepsilon_{O,\Lambda}$ are stochastic fields with zero expectation value which are uncorrelated with any other coarse-grained fields (in particular the O_Λ), and only have zero-lag correlations among themselves. We have pulled out an overall factor $n_{0,\Lambda}$ in Eq. (2.116), so that the quantities in brackets are dimensionless. Note that the term $c_{1,\Lambda}[\cdots]$ in Eq. (2.114) now corresponds to ε_Λ.

2.10.4 Renormalizing the bias parameters

We see that the bias expansion written in Eq. (2.116) contains exactly the operators of the general local bias expansion described in Sec. 2.5. However, now all terms in Eq. (2.116) depend on the coarse-graining

scale Λ, which is an arbitrary cutoff introduced in the calculation. Nevertheless, if we have included all necessary operators in the bias expansion up to a given order in perturbation theory, we know that we should be able to describe the statistics of n_g up to that order without making reference to an unphysical smoothing scale Λ. The goal of renormalization is to reorder the sum in Eq. (2.116) into a sum over Λ-independent operators multiplied by Λ-independent parameters:

$$n_g(\boldsymbol{x}) = \overline{n}_g \left[1 + [\varepsilon] + \sum_O \Big\{ b_O + [\varepsilon_O] \Big\} [O] \right] , \qquad (2.117)$$

where $[O]$ are renormalized operators which we will discuss below and b_O, $[\varepsilon_O]$ are the physical, renormalized bias coefficients and stochastic fields. Clearly, Eq. (2.117) is not unique, as different linear combinations of operators $[O]$ will yield an equivalent expression with different bias parameters. However, any complete renormalized bias expansion of the form Eq. (2.117) can be related unambiguously to any other (see Appendix C). In the following, we will focus on the operators $[O]$ and biases b_O, and will not further consider the stochastic contributions; their renormalization can be derived analogously.

The reason why we need to renormalize the bias coefficients is that, when calculating the large-scale statistics of n_g, we obtain results that are strongly dependent on the artificial coarse-graining scale Λ if we use the "bare" bias expansion Eq. (2.116). This already becomes apparent when taking the ensemble average of Eq. (2.116), which should yield the observed mean density of galaxies $\langle n_{g,\Lambda} \rangle = \overline{n}_g$. We obtain

$$\overline{n}_g = \langle n_{g,\Lambda} \rangle = n_{0,\Lambda} \left[1 + \sum_O c_{O,\Lambda} \langle O_\Lambda \rangle \right] = n_{0,\Lambda} \left[1 + c_{\delta^2,\Lambda} \langle \delta_\Lambda^2 \rangle + c_{K^2,\Lambda} \langle (K_\Lambda^{ij})^2 \rangle + \cdots \right] , \qquad (2.118)$$

where we have used the fact that $\langle \varepsilon_\Lambda \rangle = 0 = \langle \varepsilon_{O,\Lambda} \rangle$, and have written the leading terms at second order. These terms thus renormalize the "bare" mean number density $n_{0,\Lambda}$, which is physically meaningless due to its Λ-dependence. We can infer from this relation, and the corresponding ensemble average of Eq. (2.117), our first *renormalization condition* for the operators, namely that $\langle [O] \rangle = 0$. Consider $O = \delta^2$ as an example. We can define

$$[\delta^2] = \delta_\Lambda^2 - \sigma^2(\Lambda) + \cdots , \quad \text{where} \quad \sigma^2(\Lambda) \equiv \left\langle \left(\delta_\Lambda^{(1)} \right)^2 \right\rangle , \qquad (2.119)$$

and the ellipsis stands for higher-order contributions. Clearly, $\langle [\delta^2] \rangle = 0$ at lowest order, as desired. This trivial renormalization can clearly be performed for all other operators as well.

Next, let us consider the galaxy-matter cross-power spectrum, which is the simplest way to estimate the linear bias b_δ (Sec. 4.1). For demonstration, we include both the leading term and the contribution $\propto \delta_\Lambda^2$ in the bare bias expansion:

$$\langle \delta_g(\boldsymbol{k}) \delta(\boldsymbol{k}') \rangle = c_{\delta,\Lambda} \langle \delta_\Lambda(\boldsymbol{k}) \delta(\boldsymbol{k}') \rangle + c_{\delta^2,\Lambda} \langle (\delta_\Lambda^2)(\boldsymbol{k}) \delta(\boldsymbol{k}') \rangle + \cdots , \qquad (2.120)$$

where we have inserted Eq. (2.116) and the dots denote remaining second-order as well as higher-order terms. Our goal is to identify the bias b_δ on very large scales $k \to 0$ through

$$\langle \delta_g(\boldsymbol{k}) \delta(\boldsymbol{k}') \rangle \overset{k \to 0}{=} b_\delta \langle \delta^{(1)}(\boldsymbol{k}) \delta^{(1)}(\boldsymbol{k}') \rangle . \qquad (2.121)$$

Since, for fixed Λ, both δ_Λ and δ asymptote to the linear density field $\delta^{(1)}$ in the large-scale limit, we see that the first term in Eq. (2.120) already shows the desired behavior. That is, at this order there is no need to renormalize $O = \delta$, and we have $[\delta] = \delta$ (this does not imply however that $b_\delta = c_{\delta,\Lambda}$, as we will see below; only in the limit $\Lambda \to 0$ (infinite smoothing scale) does the bare bias parameter asymptote to the renormalized one).

This is not the end of the story however. Let us consider the second contribution in Eq. (2.120). Note that $(\delta_\Lambda^2)(\boldsymbol{k})$ involves a convolution in Fourier space. That is, small-scale (high-wavenumber) modes contribute to this term, no matter how small k is. Given the assumed Gaussian initial conditions, the leading contribution

to $\langle(\delta_\Lambda^2)\delta\rangle$ appears at second order in perturbation theory, and is obtained by replacing each instance of δ_Λ with $\delta_\Lambda^{(2)}$ (the third possible contribution vanishes). This yields

$$
\begin{aligned}
\langle(\delta_\Lambda^2)(\boldsymbol{k})\delta(\boldsymbol{k}')\rangle'\big|_{\mathrm{LO}} &= 2\left\langle\left(\delta_\Lambda^{(1)}\delta_\Lambda^{(2)}\right)(\boldsymbol{k})\delta^{(1)}(\boldsymbol{k}')\right\rangle' \\
&= 2\int_{\boldsymbol{p}_1}\int_{\boldsymbol{p}_2}\int_{\boldsymbol{p}_3}(2\pi)^3\delta_D(\boldsymbol{k}-\boldsymbol{p}_{123})W_\Lambda(p_1)W_\Lambda(|\boldsymbol{p}_{23}|)F_2(\boldsymbol{p}_2,\boldsymbol{p}_3) \\
&\qquad\times\langle\delta^{(1)}(\boldsymbol{p}_1)\delta^{(1)}(\boldsymbol{p}_2)\delta^{(1)}(\boldsymbol{p}_3)\delta^{(1)}(\boldsymbol{k}')\rangle' \\
&= \left[4\int_{\boldsymbol{p}}W_\Lambda(p)W_\Lambda(|\boldsymbol{p}+\boldsymbol{k}|)F_2(\boldsymbol{p},\boldsymbol{k})P_{\mathrm{L}}(p)\right]P_{\mathrm{L}}(k) \\
&\stackrel{p\gg k}{=}\left[4\frac{17}{21}\int_{\boldsymbol{p}}[W_\Lambda(p)]^2P_{\mathrm{L}}(p)\right]P_{\mathrm{L}}(k)=\frac{68}{21}\sigma^2(\Lambda)\,P_{\mathrm{L}}(k)\,,
\end{aligned}
\tag{2.122}
$$

where F_2 is the second-order perturbation theory kernel [Eq. (B.13)]. Here, a subscript "LO" denotes the leading-order expression of the correlator, that is, at the lowest non-vanishing order in perturbation theory. In the last line, we have assumed the large-scale limit $k \to 0$. Since the dominant contribution to the momentum integral comes from modes $p \sim \Lambda \gg k$, it is appropriate to assume $p \gg k$. In this limit, the angle-average of F_2 yields $17/21$. Thus, the second-order bare bias operator δ_Λ^2 leads to a contribution that is of the exact same shape on large scales as the linear bias term, namely $\propto P_{\mathrm{L}}(k)$, but depends on the variance $\sigma^2(\Lambda)$ of the density field at the cutoff scale. This is clearly unsatisfactory, and very reminiscent of the issue we encountered with the cubic bias term in Sec. 2.10.1. Fortunately, there is a simple remedy: we add δ_Λ as counter-term to $[\delta^2]$, extending Eq. (2.119) to

$$
[\delta^2] = \delta_\Lambda^2 - \sigma^2(\Lambda)\left[1+\frac{68}{21}\delta_\Lambda\right]+\cdots\,.
\tag{2.123}
$$

The same exercise can then be performed for the other bare bias operators, ensuring that Eq. (2.121) actually holds.

However, the large-scale limit of the two-point function Eq. (2.121) is not the only constraint. Consider for example the galaxy-matter-matter three-point function, which on large scales allows for a measurement of the two (physical, renormalized) second-order biases b_{δ^2}, b_{K^2}, as will be described in detail in Sec. 4.1. Specifically, the desired large-scale limit of the three-point function is

$$
\langle\delta_g(\boldsymbol{k})\delta(\boldsymbol{k}_1)\delta(\boldsymbol{k}_2)\rangle\stackrel{k_1,k_2\to 0}{=}b_\delta\langle\delta(\boldsymbol{k})\delta(\boldsymbol{k}_1)\delta(\boldsymbol{k}_2)\rangle_{\mathrm{LO}}+b_{\delta^2}\langle(\delta^2)(\boldsymbol{k})\delta(\boldsymbol{k}_1)\delta(\boldsymbol{k}_2)\rangle_{\mathrm{LO}}
$$
$$
+b_{K^2}\langle(K_{ij})^2(\boldsymbol{k})\delta(\boldsymbol{k}_1)\delta(\boldsymbol{k}_2)\rangle_{\mathrm{LO}}\,,
\tag{2.124}
$$

where, again, a subscript "LO" denotes the leading-order expression in perturbation theory. Note in particular the absence of the coarse-graining scale Λ in this expression; technically, this is because the right-hand side does not contain any loop integrals.

In the bare bias expansion, Eq. (2.124) involves $c_{\delta,\Lambda}\langle\delta_\Lambda(\boldsymbol{k})\delta(\boldsymbol{k}_1)\delta(\boldsymbol{k}_2)\rangle$. As in the case of the two-point function, this already asymptotes to the leading-order expression in Eq. (2.124) in the large-scale limit. Thus, we again do not need to renormalize the operator δ. This in fact continues at all orders in perturbation theory (but only at the level of the local bias expansion excluding higher-derivative operators), and is related to the fact that we did not need to include $\mathrm{tr}\,\Pi^{[n]}$ in the bias expansion in Sec. 2.5. On the other hand, now the expression Eq. (2.123) is not sufficient to renormalize δ_Λ^2, as it yields a cutoff-dependent result at next-to-leading (NLO, or 1-loop) order, which is fourth order in perturbations, when inserted into Eq. (2.124). In this case, we can cure the unphysical behavior by adding $(\delta_\Lambda)^2$ and $(K_\Lambda^{ij})^2$ as counter-terms. Specifically, Ref. [197] derived

$$
[\delta^2] = \delta_\Lambda^2 - \sigma^2(\Lambda)\left[1+\frac{68}{21}\delta_\Lambda+\frac{24032}{6615}\delta_\Lambda^2+\frac{254}{2205}(K_{ij,\Lambda})^2\right]\,.
\tag{2.125}
$$

Of course, going to higher order in perturbations requires the addition of more counter-terms, such as $\sigma^2(\Lambda)\delta_\Lambda^3$, as well as corrections to lower-order counter-terms such as $\sigma^4(\Lambda)\delta_\Lambda$, each multiplied by a coefficient determined through the renormalization conditions. Similarly, $(K_\Lambda^{ij})^2$ also needs to be renormalized. Crucially however, we do not need these relations to derive the large-scale galaxy statistics: at leading order, we can always be assured that the bias parameters that appear are the physical ones, thanks to renormalization. When including next-to-leading contributions, i.e. loop terms, we can simply drop those terms that are canceled by counter-terms to lower-order operators. They are easily identified by having the same scale dependence as the contribution from the lower-order operator, but with a cutoff-dependent amplitude (cf. Eq. (2.122); this will be illustrated in Sec. 4.1.4).

Finally, after these concrete examples, we can turn to systematically defining the conditions which the renormalized operators need to satisfy. A necessary condition for a properly renormalized operator O in the bias expansion is that *all N-point auto- and cross-correlations* of the galaxy density asymptote to their leading-order expressions on large scales. This is crucial, as it ensures that a single set of physical bias parameters describes all statistics of a given galaxy sample. Achieving this result might seem like a daunting task; however, many of the auto- and cross-correlations are redundant. Since, in our perturbative bias expansion, all operators are constructed out of $\partial_i \partial_j \Phi^{(1)}$, or equivalently in Fourier space $(k_i k_j/k^2)\delta^{(1)}(\boldsymbol{k})$, it is sufficient to enforce the following *renormalization conditions* for each operator in the bias expansion [197]:

$$\lim_{k_i \to 0} \langle [O](\boldsymbol{k})\delta^{(1)}(\boldsymbol{k}_1)\cdots\delta^{(1)}(\boldsymbol{k}_n)\rangle = \langle O(\boldsymbol{k})\delta^{(1)}(\boldsymbol{k}_1)\cdots\delta^{(1)}(\boldsymbol{k}_n)\rangle_{\text{LO}}\,, \tag{2.126}$$

where $n = 0, 1, 2, \ldots$. The $n = 0$ condition simply reads $\langle [O](\boldsymbol{x})\rangle = 0$, which is what we found after Eq. (2.118). At any order in perturbation theory, only a finite set of these conditions adds new constraints. Note that the constraints Eq. (2.126) are sufficient to ensure that all statistics where $[O]$ contributes at leading order, for example, the three-point function in case of $O = \delta^2$, approach the leading-order result on large scales. This is because Eq. (2.126) can be taken as building block to construct all other auto- and cross-correlations between any operators in the renormalized set. When $[O]$ appears in loop integrals, it can still lead to significant contributions at low k that depend on small-scale modes. However, these contributions can be expanded in powers of k^2 (they are analytic), and hence are absorbed by counter-terms to the renormalized stochastic fields $[\varepsilon]$, $[\varepsilon_O]$, which show precisely the same large-scale behavior [cf. Eq. (2.88)]. We will encounter an example of this kind in Sec. 4.1.4 for $O = \delta^2$.

In general, we can write the renormalized operators as a sum over bare operators with Λ-dependent coefficients,

$$[O] = \sum_{O'} Z_{OO'}(\Lambda)O'\,. \tag{2.127}$$

The relation between the bare bias parameters $c_{O,\Lambda}$ and their renormalized counterparts b_O is then given by matching coefficients in the respective expansions, Eq. (2.116) and Eq. (2.117) (see also Appendix C):

$$b_O = \sum_{O'} (Z^{-1})_{O'O}(\Lambda)c_{O',\Lambda}\,, \tag{2.128}$$

which involves the transpose of the inverse of the coefficient matrix $Z_{OO'}$. That is, in order to obtain an expression for the renormalized bias b_O, we need to know which renormalized operators $[O']$ contain O as counter-term. For this reason, even though δ is not renormalized, so that $Z_{\delta O'} = 0$ for any $O' \neq \delta$, the same is not true for b_δ, since δ appears as a counter-term for many other operators (including δ^2, as we have seen). Instead, we have [197][10]

$$b_\delta = c_{\delta,\Lambda} + \sigma^2(\Lambda)\left[\frac{68}{21}\left(c_{\delta^2,\Lambda} + \frac{2}{3}c_{K^2,\Lambda}\right) + 3c_{\delta^3,\Lambda} + \frac{2}{3}c_{\delta K^2,\Lambda} + \frac{32}{63}c_{\text{td},\Lambda}\right]\,, \tag{2.129}$$

---

[10]Note the slightly different normalization of bias coefficients there, and that $\mathcal{G}_2 = (K_{ij})^2 - (2/3)\delta^2$ in our notation (see Appendix C).

where $c_{\mathrm{td},\Lambda}$ is the bare bias coefficient multiplying $O_{\mathrm{td}}^{(3)}$. We will re-derive this expression from Eqs. (2.116)–(2.117) using the peak-background split approach in Sec. 3.4.

The renormalized operators and bias parameters defined through Eq. (2.126) and Eq. (2.128) describe the n-point functions of galaxies in the large-scale limit, and are thus particularly relevant to actual observations (e.g. from galaxy redshift surveys). It is however just as well possible to construct renormalized bias operators corresponding to other observables. Consider the joint moments of the smoothed halo density field and matter density field on a *physical* scale R_ℓ, $\langle[\delta_{h,\ell}(\boldsymbol{x})]^n[\delta_\ell(\boldsymbol{x})]^m\rangle$. As we will describe in detail in Sec. 4.2–4.3, these moments can be used to derive bias parameters, for example through the so-called hierarchical ansatz, or the "scatter plot" method (note that these moments are not sufficient to unambiguously measure all bias coefficients, and mostly restricted to the LIMD biases b_{δ^n}). The renormalized bias parameters inferred using these measurements are *not* the same as those introduced above. This is because the moments are defined with respect to a specific scale R_ℓ, and filtering kernel $W_{R_\ell}(k)$. The biases $b_O^{\mathrm{m}}(R_\ell)$ inferred from moments are renormalized bias parameters in analogy to the $b_O(R_\ell)$, i.e. they are coefficients of renormalized operators $[O]_\ell^{\mathrm{m}}$. However, the latter are defined with different renormalization conditions, specifically

$$\langle[O]_\ell^{\mathrm{m}}\rangle = 0 \quad \text{and} \quad \langle[O]_\ell^{\mathrm{m}}(\boldsymbol{x})[\delta_\ell^{(1)}(\boldsymbol{x})]^n\rangle = \langle O[\delta_\ell,\cdots\delta_\ell](\boldsymbol{x})[\delta_\ell^{(1)}(\boldsymbol{x})]^n\rangle_{\mathrm{LO}}, \qquad (2.130)$$

where $O[\delta_\ell,\cdots\delta_\ell]$ denotes the operator O constructed out of smoothed density fields; that is, in case of $O = \delta^n$, it stands for $O[\delta_\ell,\cdots\delta_\ell](\boldsymbol{x}) = (\delta_\ell(\boldsymbol{x}))^n$. On the other hand, on the left-hand side, O is smoothed *after* taking the nonlinear functional of the density field. It is straightforward to verify that the counter-terms for $[O]_\ell^{\mathrm{m}}$ scale in the same way with the cutoff as those for $[O]$; in other words, $[O]_\ell^{\mathrm{m}}$ and $[O]$ only differ by finite R_ℓ-dependent corrections. The same correspondingly holds for $b_O^{\mathrm{m}}(R_\ell)$ and b_O. The relation between the two sets of bias parameters can be derived at any given order in perturbation theory (Sec. 4.2). Note that the finite difference between $b_O^{\mathrm{m}}(R_\ell)$ and b_O does not necessarily vanish as $R_\ell \to \infty$, as one might think, because the finite terms involve different integrals over the linear power spectrum. These depend on the local spectral index of the power spectrum which, in turn, induces a logarithmic R_ℓ dependence as we will see in Sec. 4.2.

2.10.5 Higher-derivative operators

The considerations so far were limited to the operators in the local basis. This can be traced back to Eq. (2.113), where we approximated $\partial_i\partial_j\Phi_\Lambda$ as approximately constant over the scales over which the convolution kernel is nonzero. This corresponds to the lowest-order term in a Taylor series,

$$\partial_i\partial_j\Phi_\Lambda(\boldsymbol{x}_1) = \partial_i\partial_j\Phi_\Lambda(\boldsymbol{x}) + \partial_i\partial_j\partial_k\Phi_\Lambda(\boldsymbol{x})(\boldsymbol{x}_1 - \boldsymbol{x})^k + \frac{1}{2}\partial_i\partial_j\partial_k\partial_l\Phi_\Lambda(\boldsymbol{x})(\boldsymbol{x}_1 - \boldsymbol{x})^k(\boldsymbol{x}_1 - \boldsymbol{x})^l + \cdots, \quad (2.131)$$

where we have written $\boldsymbol{x} \equiv \boldsymbol{x}_{\mathrm{fl}}(\tau)$ for clarity. We can now include these sub-leading terms in Eq. (2.113). The leading correction to Eq. (2.113) is

$$\prod_{i=1}^{n}\int d^3\boldsymbol{x}_i\, F_{i_1j_1\cdots i_nj_n}^{(n)}(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n)\partial_{i_1}\partial_{j_1}\Phi(\boldsymbol{x}_1)\cdots\partial_{i_n}\partial_{j_n}\Phi(\boldsymbol{x}_n) = [\text{right-hand side of Eq. (2.113)}]$$

$$+ \sum_{k=0}^{n}\frac{n!}{(n-k)!}\partial_l\partial_m\partial_{i_1}\partial_{j_1}\Phi_\Lambda(\boldsymbol{x})\cdots\partial_{i_k}\partial_{j_k}\Phi_\Lambda(\boldsymbol{x})\prod_{i=1}^{n}\int d^3\boldsymbol{x}_i\,(\boldsymbol{x}_1 - \boldsymbol{x})^l(\boldsymbol{x}_1 - \boldsymbol{x})^m\, F_{i_1j_1\cdots i_nj_n}^{(n)}(\boldsymbol{x}_1,\cdots\boldsymbol{x}_n)$$

$$\times\,\partial_{i_{k+1}}\partial_{j_{k+1}}\Phi_s(\boldsymbol{x}_{k+1})\cdots\partial_{i_n}\partial_{j_n}\Phi_s(\boldsymbol{x}_n)$$

$$+\cdots. \qquad (2.132)$$

Here, we have written only one term with both derivatives acting on $\partial_{i_1}\partial_{j_1}\Phi_\Lambda(\boldsymbol{x}_1)$; the ellipsis in the last line stands for other terms which involve ∂_l and ∂_m each acting on one (i.e. not necessarily the same) of the other instances of $\partial_{i_k}\partial_{j_k}\Phi_\Lambda$ and contracted with one of the i_k, j_k or with each other. This is because neither the kernel $F^{(n)}$ nor the small-scale modes Φ_s have any preferred directions (apart from those induced by the long-wavelength modes themselves). For the same reason, the leading higher-derivative correction involves two powers of separations $(\boldsymbol{x}_i - \boldsymbol{x})(\boldsymbol{x}_j - \boldsymbol{x})$.

58

We see that the higher-derivative terms have the same structure as the lowest-order term, Eq. (2.113), with two differences: first, there are two more derivatives acting on the long-wavelength modes; second, the convolution over the small-scale modes now involves a modified kernel,

$$(\boldsymbol{x}_i - \boldsymbol{x})^l (\boldsymbol{x}_j - \boldsymbol{x})^m F^{(n)}_{i_1 j_1 \cdots i_n j_n} \,. \tag{2.133}$$

This modified kernel, given our assumptions about the scales over which $F^{(n)}$ is nonzero, scales as R_*^2 times the kernel appearing in Eq. (2.113). This then leads us to bare bias parameters that correspondingly scale as R_*^2 which multiply precisely the higher-derivative operators described in detail in Sec. 2.6. We can then let the smoothing scale $\Lambda \to 0$, leading to renormalized higher-derivative biases that scale as R_*^2. This reasoning continues correspondingly to higher orders.

2.11 Summary

In this section, we have described the general, perturbative bias expansion of arbitrary large-scale structure tracers. Clearly, while well-defined and systematic, this expansion is nontrivial and consists of a complex set of terms. For this reason, we provide a brief summary here before moving on to the following sections. The general bias expansion can be broken down into three ingredients:

- ***The deterministic local expansion, i.e. at leading order in derivatives (Sec. 2.5):*** this series of terms of the form $b_O\,O$, where O is an operator and b_O is its associated bias parameter, includes as operators powers of the density and tidal field, as well as convective time derivatives of the tidal field, as summarized in Eq. (2.64) (in Eulerian space) and Eq. (2.61) (in Lagrangian space). Each operator O has exactly two spatial derivatives acting on each occurrence of the gravitational potential Φ (where we count the differential operator $\partial_i\partial_j/\nabla^2$ as zero net derivatives).

- ***Stochastic contributions (Sec. 2.8):*** in addition to the leading stochastic field ε, there is an additional stochastic field ε_O associated with each operator O in the bias expansion. This can be interpreted as "scatter" in the deterministic bias parameter b_O.

- ***Higher-derivative terms (Sec. 2.6):*** For each operator O in the local bias expansion, there are higher spatial derivative terms such as $b_{\nabla^2 O}\nabla^2 O$ (and others; for the precise list of terms, see Sec. 2.6). Physically, these terms take into account the fact that galaxy formation is not perfectly local. The bias coefficients of these terms have units of length to some power, for example Mpc^2 in the case of $b_{\nabla^2 O}$. The length scale that sets the value of these coefficients is the physical "nonlocality scale" R_* of galaxy formation; on scales of order R_*, these terms are un-suppressed, and any perturbative description of galaxy bias necessarily breaks down. Note that the stochastic fields also have associated higher-derivative contributions, which effectively capture the fact that the stochastic fields are expected to be correlated over the scale R_* (see Sec. 2.8 for an example).

A further important result pertains to the relation between the galaxy velocity field and the matter velocity (***velocity bias***, Sec. 2.7): velocity bias is guaranteed to be a higher-derivative effect. That is, we can write at lowest order in perturbations and derivatives,

$$\boldsymbol{v}_g = \boldsymbol{v} + \beta_{\nabla^2\boldsymbol{v}}\nabla^2\boldsymbol{v} + \boldsymbol{\varepsilon}_v(\boldsymbol{x},\tau)\,, \tag{2.134}$$

where $\beta_{\nabla^2\boldsymbol{v}} \sim R_*^2$ is related to the nonlocality scale of galaxy formation (other possible terms such as $\propto \boldsymbol{\nabla}\delta$ are equivalent to $\nabla^2\boldsymbol{v}$ at linear order), and $\boldsymbol{\varepsilon}_v$ is a stochastic velocity contribution whose large-scale power spectrum scales as k^2 (Sec. 2.8).

The sections referenced above provide all the ingredients necessary to write down the general bias expansion at any desired order. To be specific, we now summarize the complete bias expansion of a general galaxy sample up to third order:

$$
\begin{aligned}
\delta_g =\ & b_1[\delta] + b_{\nabla^2\delta}[\nabla^2\delta] + [\varepsilon] \\
& + \frac{1}{2}b_2[\delta^2] + b_{K^2}\big[(K_{ij})^2\big] + [\varepsilon_\delta\delta] \\
& + \frac{1}{6}b_3[\delta^3] + b_{\delta K^2}[\delta(K_{ij})^2] + b_{K^3}\big[(K_{ij})^3\big] + b_{\mathrm{td}}\big[O_{\mathrm{td}}^{(3)}\big] + \big[\varepsilon_{\delta^2}\delta^2\big] + \big[\varepsilon_{K^2}(K_{ij})^2\big] \\
& + O(\delta^4) + \mathcal{O}\big[R_*^2\nabla^2(\delta^2),\, R_*^4\nabla^4\delta\big]\,.
\end{aligned}
\tag{2.135}
$$

The brackets denote renormalized operators, as defined in Sec. 2.10.4. As indicated in the last line, the terms neglected here are either fourth order in perturbation theory, or involve higher derivatives of nonlinear operators, or four powers of spatial derivatives (all of these include stochastic contributions as well). Note that the number of higher-derivative terms to be kept depends on the scale R_* (recall that $b_{\nabla^2\delta} \propto R_*^2$). Here, we have assumed that R_* is of order of the scale where the matter density field becomes nonlinear, in which case it is sufficient to keep only the leading higher-derivative term, $b_{\nabla^2\delta}\nabla^2\delta$ (see Sec. 4.1.4), when

going to third order in perturbations. In this case, Eq. (2.134) also provides the complete relation for galaxy velocities at this order. In practice, when analyzing an actual galaxy sample, as many higher-derivative terms should be included as the data are able to constrain (see the discussion in [128]).

In Sec. 4, we will describe how Eq. (2.135) makes predictions for galaxy (or halo) statistics. Depending on the precise statistic chosen, usually only a subset of the terms in the general bias expansion contribute at a given order. Before that however, in the next section (Sec. 3), we will discuss the physical interpretation of several of the bias parameters in Eq. (2.135) as responses of the galaxy number density to long-wavelength perturbations, an argument historically known as the "peak-background split."

3 Peak-background split: rigorous formulation and approximations

In the previous section, we have derived what bias parameters need to be included at any given order in the perturbative bias expansion to describe a general tracer of the large-scale structure. We now turn to physical arguments that can be used to derive the actual values of these bias parameters, as well as hierarchies and relations between them. We will begin with general physical arguments valid for general tracers such as galaxies, before making simplifying assumptions which mainly apply to dark matter halos. For this, we use the *peak-background split (PBS)* approach in a more general sense than frequently understood; this will lead to a well-defined physical interpretation of the bias parameters that are measured from the correlation functions of galaxies (i.e., the *renormalized* bias parameters).

The decomposition of the density field into the sum of a low-amplitude signal δ_ℓ (background) with a large coherence length and a high-amplitude, noisy component $\delta_s = \delta - \delta_\ell$ (peak) with a small coherence length was first introduced by [12], though the term "peak-background split" was coined by [13]. The PBS simply states that *a long-wavelength density perturbation acts like a local modification of the background density for the purposes of the formation of halos and galaxies*, since it can be considered constant over the spatial scale within which tracers form (that is, the scale R_* introduced in Sec. 2.5). This separation of scales between the long-wavelength perturbation—on the scales on which we measure correlations—and the small-scale perturbations—that are directly involved in the formation of tracers—is precisely the physical argument made for the general bias expansion in Sec. 2.5.1.

While the PBS argument was originally introduced for long-wavelength density perturbations, and thus only allowed for a derivation of the LIMD parameters b_N (e.g. [12, 14, 15, 110, 16]), similar arguments can be made for other bias terms:

- those with respect to the tidal field (see Sec. 5.9);

- higher-derivative operators such as $\nabla^2\delta$ or $(\boldsymbol{\nabla}\delta)^2$ (see Sec. 6.6.1); and

- bias contributions induced by primordial non-Gaussianity (see Sec. 7).

For the remainder of this section, we will focus on the bias with respect to powers of the density b_N.

After the initial proposition of [12], the physical argument regarding the separation of long- and short-wavelength modes has come to be associated with theoretical approaches such as excursion set (Sec. 5) and peaks (Sec. 6), that is, analytically tractable toy models of large-scale structure. However, as we describe in Sec. 3.1–3.2, the PBS argument is not specific to these models and is in fact *exact*. That is, when defined properly, the PBS predicts the exact bias parameters for halos and galaxies. One example of how the exact PBS can be implemented is given in Sec. 3.2; the results of implementing this approach with N-body simulations are presented later in Sec. 4.4. In Sec. 3.3, we specialize to the case of tracers which follow a so-called universal mass function. This yields the well-known expressions for the bias parameters commonly referred to as "peak-background split biases."

3.1 Bias parameters as responses

In Sec. 2, we arrived at a bias expansion of the galaxy number density perturbation of the form

$$\delta_g(\boldsymbol{x}, \tau) = \sum_O b_O(\tau)[O](\boldsymbol{x}, \tau), \tag{3.1}$$

where we have neglected stochastic terms, as they will not play a role in this section. As discussed in Sec. 2.10, the physical, measurable biases are really given as the coefficients of the *renormalized* operators

[O]. We will only deal with the renormalized bias parameters b_O in Sec. 3.1–3.3, although an interesting link to the bare bias parameters and renormalization will be presented in Sec. 3.4.

Now, we would like to ask: what is the physical interpretation of the bias coefficients appearing in Eq. (3.1)? Let us focus on the coefficients $b_N \equiv N! b_{\delta^N}$ of the operator δ^N [this is $\left(\operatorname{tr}\Pi^{[1]}\right)^N$ in the Eulerian basis Eq. (2.64)]. Consider a large region (much larger than the nonlocality scale R_* of the galaxy sample considered) characterized, at a given arbitrary time, by a mean fractional overdensity Δ. According to the peak-background split argument, the expectation value of the *physical* galaxy density n_g in such a region is given by the average abundance of tracers in a fictitious FRW spacetime with modified physical background density

$$\tilde{\varrho}_m = (1 + \Delta)\, \overline{\varrho}_m \,, \tag{3.2}$$

where $\overline{\varrho}_m$ is the fiducial background density. We shall further elaborate on exactly what this "fictitious spacetime" is in Sec. 3.2. Eq. (3.2) is equivalent to adding a uniform component $\overline{\varrho}_m \Delta$ to the matter density. Then, at a given point within this region where the density is $\varrho_m(\boldsymbol{x}) = \overline{\varrho}_m[1 + \delta(\boldsymbol{x})]$ (here δ is not necessarily small), the matter density is perturbed to

$$\varrho_m(\boldsymbol{x}) \to \varrho_m(\boldsymbol{x}) + \overline{\varrho}_m \Delta = \overline{\varrho}_m(1 + \delta(\boldsymbol{x}) + \Delta)\,, \quad \text{or} \quad \delta(\boldsymbol{x}) \to \delta(\boldsymbol{x}) + \Delta\,. \tag{3.3}$$

Thus, we *shift* all density perturbations by an amount Δ. Note that we add a fixed amount of *uniform* matter density everywhere; we do *not* rescale the local matter density ϱ_m by $1 + \Delta$, which would also amplify the fluctuations δ.

We now introduce the strict definition of *peak-background split bias parameters* b_N ($N \geq 1$) as the derivative of the mean physical number density of galaxies $\overline{\mathrm{n}}_g$ with respect to Δ:

$$b_N \equiv \frac{1}{\overline{\mathrm{n}}_g|_{\Delta=0}} \frac{\partial^N \overline{\mathrm{n}}_g|_\Delta}{\partial \Delta^N}\bigg|_{\Delta=0}. \tag{3.4}$$

Using Eq. (3.2), we can also write this as

$$b_N = \frac{\overline{\varrho}_m^N}{\overline{\mathrm{n}}_g} \frac{\partial^N \overline{\mathrm{n}}_g}{\partial \tilde{\varrho}_m^N}\bigg|_{\overline{\varrho}_m}\,, \tag{3.5}$$

where the derivatives are evaluated at the fiducial value of $\overline{\varrho}_m$. These relations show that, if we can predict the abundance of galaxies as a function of the modified background density $\tilde{\varrho}_m$, at fixed time t, we can predict the LIMD bias parameters b_N. In the next section, we will discuss precisely this approach of measuring biases, in particular for dark matter halos. In the course of that calculation, we will also clarify the physical meaning of the transformation Eq. (3.3).

3.2 Exact implementation of the PBS: separate-universe approach

To begin, let us define a bit more carefully what we mean by the transformation Eqs. (3.2)–(3.3). A uniform adiabatic density perturbation on an FRW background is equivalent to a different (curved) FRW background, as first shown by [186] (we continue to assume Gaussian initial conditions in this section). This idea, often referred to as "separate-universe" approach, has been used in many calculations and N-body simulations since then, first starting at linear order in the perturbation Δ [187, 188, 208, 209, 210, 211, 212, 95, 189, 213], and more recently generalized to fully nonlinear order [214, 185]. The following brief description follows the one laid out in [214]. We assume a ΛCDM cosmology throughout.

Consider a long-wavelength overdensity $\Delta(t)$ as in Eq. (3.3), distinguished there from the small-scale perturbations δ. Throughout, we will allow Δ to be fully nonlinear; the separate-universe approach is not restricted to linear order in the density perturbation, as emphasized in [185]. Moreover, t stands for the proper time of comoving observers throughout. We then proceed to absorb the overdensity Δ into the physical background matter density of a modified cosmology $\tilde{\varrho}_m(t)$ as

$$\overline{\varrho}_m(t)\,[1 + \Delta(t)] = \tilde{\varrho}_m(t)\,. \tag{3.6}$$

63

Thus, instead of embedding the region with overdensity Δ in a fiducial background Universe, one considers it as a separate universe with an altered cosmology. We will now derive the parameters of this modified cosmology in relation to Δ, specifically the *linearly extrapolated* present-day overdensity

$$\Delta_0^{(1)} \equiv \Delta(t_i) \frac{D(t_0)}{D(t_i)} \,, \tag{3.7}$$

where D is the linear growth function of the fiducial cosmology, t_0 is the present time, and t_i denotes an early time at which $\Delta(t_i)$ is still small. Again, we will not assume that $\Delta_0^{(1)}$ is small, although in most practical applications one will choose $|\Delta_0^{(1)}| \lesssim 1$.

Expressed in terms of the standard cosmological parameters, i.e. $\bar{\varrho}_m(a = 1) = \bar{\rho}_m = \Omega_{m0} \frac{3H_0^2}{8\pi G}$ and $H_0 = 100 \, h \, \mathrm{km \, s^{-1} Mpc^{-1}}$, Eq. (3.6) becomes

$$\frac{\Omega_{m0}h^2}{a^3(t)} [1 + \Delta(t)] = \frac{\tilde{\Omega}_{m0}\tilde{h}^2}{\tilde{a}^3(t)} \,, \tag{3.8}$$

where we used a tilde to denote quantities in the modified cosmology. For the fiducial cosmology, we adopt the standard convention for the scale factor $a(t_0) = 1$. In contrast, for the modified cosmology, it is convenient to choose the normalization such that $\lim_{t\to 0} \tilde{a}(t)/a(t) = 1$. These conventions lead to (as $\Delta(t \to 0) = 0$)

$$\Omega_{m0}h^2 = \tilde{\Omega}_{m0}\tilde{h}^2 \,, \quad \Omega_{\Lambda0}h^2 = \tilde{\Omega}_{\Lambda0}\tilde{h}^2 \,. \tag{3.9}$$

Note that this relation implies that the comoving matter densities defined using the respective scale factor are the same: $\tilde{\rho}_m \equiv \tilde{a}^3 \tilde{\varrho}_m = \bar{\rho}_m$. Introducing $\delta_a(t)$ through $\tilde{a}(t) = [1 + \delta_a(t)]a(t)$, we find from Eq. (3.6)

$$1 + \Delta(t) = [1 + \delta_a(t)]^{-3} \,. \tag{3.10}$$

This is just a statement of mass conservation. The Friedmann equation for $a(t)$ is, assuming a flat fiducial cosmology for simplicity, given by

$$H^2(t) = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3} \bar{\varrho}_m(t) + \frac{1}{3}\Lambda \,, \tag{3.11}$$

where Λ denotes the cosmological constant. The same equation, but including curvature $\tilde{\mathrm{K}}$ and modifying the matter density, holds for $\tilde{a}(t)$:

$$\tilde{H}^2(t) = \left(\frac{\dot{\tilde{a}}}{\tilde{a}}\right)^2 = \frac{8\pi G}{3} \tilde{\varrho}_m(t) + \frac{1}{3}\Lambda - \frac{\tilde{\mathrm{K}}}{\tilde{a}^2(t)} \,. \tag{3.12}$$

One can then combine Eqs. (3.11)–(3.12) with the corresponding second Friedmann equations to obtain an evolution equation for Δ:

$$\ddot{\Delta} + 2H\dot{\Delta} - \frac{4}{3}\frac{\dot{\Delta}^2}{1+\Delta} = 4\pi G\bar{\varrho}_m (1+\Delta)\Delta \,. \tag{3.13}$$

When linearizing this equation in Δ, one recovers the equation for the linear growth factor (given in Eq. (B.9) using τ rather than t as time variable). Beyond linear order, Eq. (3.13) is exactly the equation for the interior density of a spherical tophat perturbation in a ΛCDM Universe (see Sec. 5.2.1 and App. A of [215]). Taking the difference of Eq. (3.11) and Eq. (3.12) yields a relation for the curvature $\tilde{\mathrm{K}}$. One can verify that $\tilde{\mathrm{K}}$ is conserved [214], which is a necessary condition for Eq. (3.12) to describe a physical FRW solution [185].

One can also generalize the separate-universe picture to include a dark energy component instead of a cosmological constant. In that case, dark energy perturbations also need to be taken into account, and the conservation of curvature only holds outside the sound horizon of the dark energy component. Recently, Ref. [216] generalized the separate-universe approach to effectively include pressure and anisotropic stress perturbations, although this only holds strictly when following the evolution of non-relativistic fluids such

as baryons and CDM. We return to this issue in Sec. 8.3, and assume for the remainder of Sec. 3 that we are outside the sound horizon of all fluid components. In this case, we can evaluate the curvature at an early time t_i, when the perturbation δ_a is infinitesimal and the Universe is in matter domination. We then have $H^2 = H_0^2 \Omega_{m0} a^{-3}$, $\dot{\delta}_a = H \delta_a$, and $\delta_a = -\Delta/3$, with which one can derive

$$\frac{\tilde{K}}{H_0^2} = \frac{5}{3} \frac{\Omega_{m0}}{a(t_i)} \Delta(t_i) \,. \tag{3.14}$$

Alternatively, using the linear growth factor D_{md} normalized such that $D_{\mathrm{md}}(t_i) = a(t_i)$, we can write

$$\frac{\tilde{K}}{H_0^2} = \frac{5}{3} \frac{\Omega_{m0}}{D_{\mathrm{md}}(t_0)} \Delta_0^{(1)} \,. \tag{3.15}$$

We have thus recovered the relation between curvature and matter density in synchronous-comoving gauge given in Sec. 2.9 [Eq. (2.97)]. Now let us derive the parameters of the modified cosmology. They are defined through the Friedmann equation at time $\tilde{t}_0$ where $\tilde{a}(\tilde{t}_0) = 1$. Defining the fractional perturbation to the Hubble parameter δ_H through $\tilde{H}(\tilde{t}_0) = H_0[1 + \delta_H]$ and using Eq. (3.9), we obtain

$$\tilde{\Omega}_{m0} = \Omega_{m0}[1 + \delta_H]^{-2}; \quad \tilde{\Omega}_{\Lambda 0} = \Omega_{\Lambda 0}[1 + \delta_H]^{-2} \,. \tag{3.16}$$

Finally, in order to derive δ_H we can make use of the Friedmann equation at $\tilde{t}_0$, which yields, for a flat fiducial cosmology,

$$\tilde{\Omega}_{K0} = -\frac{\tilde{K}}{\tilde{H}_0^2} = 1 - \tilde{\Omega}_{m0} - \tilde{\Omega}_{\Lambda 0} = 1 - (1 + \delta_H)^{-2} \,. \tag{3.17}$$

Since $\tilde{K}$ is given by Eq. (3.15), we can use this relation to solve for δ_H:

$$\delta_H = \left(1 - \frac{\tilde{K}}{H_0^2}\right)^{1/2} - 1 \,. \tag{3.18}$$

Thus, $\delta_H > 0$ if $\tilde{K} < 0$, i.e. in case of an underdense region ($\Delta_0^{(1)} < 0$). There is no solution if $\tilde{K}/H_0^2 \geq 1$, or equivalently $\Delta_0^{(1)} \geq 3D_{\mathrm{md}}(t_0)/(5\Omega_{m0})$. This is because for such a large positive curvature, the Universe reaches turnaround at or before $\tilde{a} = 1$. This is not a physical problem, it is merely not possible to parametrize such a cosmology in the standard convention. For practical applications, smaller values of $\Delta_0^{(1)}$ are in any case sufficient.

With these relations, it is straightforward to perform N-body simulations which implement a uniform density perturbation as in Eqs. (3.2)–(3.3). The power spectrum used to generate the initial conditions is unmodified in shape and amplitude apart from the modification to the linear growth factor in the background $\tilde{a}(t)$. One remaining subtlety is that we want to output the data at a fixed physical time t_{out}, which in N-body codes is usually specified by the scale factor as $a(t_{\mathrm{out}}) = a_{\mathrm{out}}$. Therefore we need to determine the corresponding scale factor in the modified cosmology as $\tilde{a}(t_{\mathrm{out}}) = a_{\mathrm{out}}[1 + \delta_a(t_{\mathrm{out}})]$, which can be done easily numerically for any given value of $\Delta_0^{(1)}$.

It is now straightforward to give an operational procedure to derive the exact PBS biases Eq. (3.4) for any tracer whose formation can be simulated: we run simulations, of sufficient volume to contain a statistical sample of the tracers of interest, up to a given fixed proper time t, with different parameters following the prescription above, which implement various values of $\Delta(t)$. This allows for a measurement of $\overline{n}_g(t, \Delta(t))$. The Eulerian LIMD biases for the tracers are then given as derivatives of the physical density of tracers $\overline{n}_g$ with respect to Δ:

$$b_N^E(t) = N! \, b_{\delta^N}^E(t) = \frac{1}{\overline{n}_g(t,0)} \frac{\partial^N \overline{n}_g(t, \Delta(t))}{\partial [\Delta(t)]^N} \bigg|_{\Delta=0} \,, \tag{3.19}$$

which corresponds to the rigorous definition of Eq. (3.4). Note that a nonlinear implementation of the separate-universe rescaling is essential if one wants to derive the higher-order biases starting with b_2; that is, the implementation has to correctly take into account the nonlinear evolution of the long-wavelength mode.

We now specialize to the most commonly considered case of dark matter halos at fixed mass. We define $\overline{n}_h(M, t)$ as the mean *comoving* number density of halos per logarithmic mass interval. Since the comoving density is related to the physical density through $\overline{n}_h = a^3 \overline{\mathrm{n}}_h$, and correspondingly $\overline{n}_h(\Delta) = \tilde{a}^3 \overline{\mathrm{n}}_h(\Delta)$ in the modified cosmology, the quantity $\overline{n}_h(\Delta)/\overline{n}_h(0) - 1$ is the fractional *Lagrangian* overdensity of halos induced by a long-wavelength density perturbation Δ [see Eq. (3.10)]. Then, the Lagrangian LIMD biases b_N^L of halos are given as derivatives of $\overline{n}_h(M, t)$ with respect to $\Delta_0^{(1)}$, i.e.

$$b_N^L(M, t) = N!\, b_{\delta^N}^L(M, t) = \frac{1}{\overline{n}_h(M, t, 0)} \frac{\partial^N \overline{n}_h(M, t, \Delta(t))}{\partial[\Delta_0^{(1)}]^N}\bigg|_{\Delta_0^{(1)} = 0} . \tag{3.20}$$

Since we are considering a pure density perturbation, i.e. a spherically symmetric long-wavelength perturbation which does not induce a proper tidal field ($K_{ij} = 0$), the mapping between b_N^L and the Eulerian LIMD biases b_N^E is exactly given by the mapping described in Sec. 2.1, in particular Eq. (2.16). Note that it is essential that halos are identified in the separate-universe simulations with respect to a fixed physical density criterion in the fiducial cosmology; in particular, the overdensity or linking length passed on to the halo finder (see Appendix D) needs to be adjusted accordingly. We will discuss results of these measurements in Sec. 4.4.

3.3 PBS biases for universal mass functions

After having defined the rigorous peak-background split biases, we now present a well-known case where Eq. (3.20) can be evaluated analytically. Motivated by the excursion-set argument (which we will describe in Sec. 5), the halo mass function, the mean comoving number density of dark matter halos per logarithmic mass interval, is often parametrized in the form

$$\overline{n}_h(M) \equiv \frac{\partial^2 \bar{N}_h}{\partial V \partial \ln M} = \frac{\overline{\rho}_m}{M} \nu_c(M)\, f[\nu_c(M)]\, J(M) \tag{3.21}$$

$$\nu_c(M) \equiv \frac{\delta_{\mathrm{cr}}}{\sigma(M)}; \quad J(M) \equiv \left|\frac{d\ln \sigma(M)}{d\ln M}\right| = \frac{d\ln \nu_c(M)}{d\ln M}, \tag{3.22}$$

where $\sigma(M) \equiv \sigma[R(M)]$ is the variance of the linear matter density field smoothed on the Lagrangian radius R which is related to the mass M through $M = (4\pi/3)\overline{\rho}_m R^3(M)$, and δ_{cr} is the linearly extrapolated initial spherical overdensity that collapses at the time at which Eq. (3.21) is evaluated (see Sec. 5.2.1). The scale-independent collapse threshold δ_{cr} can be promoted to a mass-dependent *barrier* $B(M)$. We discuss different choices of barrier in Sec. 5.9. Furthermore, the so-called multiplicity function $f(\nu_c)$ generally is an arbitrary function of ν_c. Finally, the Jacobian J is included in order to convert from an interval in the variance $\sigma(M)$ to a mass interval.

Eq. (3.21) is referred to as "universal mass function" and was first introduced by [217]. More generally, an EdS Universe with a pure power-law matter power spectrum $P_{\mathrm{L}}(k) \propto k^n$ obeys a scaling symmetry (e.g., [218]). Then, the halo mass function as a function of M and z can be exactly written in the form Eq. (3.21), since ν_c is invariant under the scaling symmetry and all moments of the matter density are directly proportional to $\sigma(M)$ [16]. The fact that the linear matter power spectrum in the standard ΛCDM cosmology can be fairly well approximated as a power law over a range of scales (thus approximately obeys the scaling symmetry) is one of the main reasons why Eq. (3.21) gives fairly accurate results in ΛCDM. In fact, the approximate symmetry can be used to rescale simulations from one cosmology to another [219]. For precision estimates of the halo bias however, the departure of ΛCDM from a power-law EdS Universe must be taken into account.

In order to derive the bias parameters via Eq. (3.20), we need to know how $\overline{n}_h$ changes under a change in the background density of the Universe [Eq. (3.2)]. Since the comoving density $\tilde{\rho}_m \equiv \tilde{a}^3 \tilde{\varrho}_m$ is unchanged [see the discussion after Eq. (3.8)], the only nontrivial contribution to the LIMD bias comes from the response of

the barrier to a change in the background density.[11] Ref. [15] argued that, under a long-wavelength density perturbation $\Delta_0^{(1)}$, the collapse threshold shifts to

$$\delta_{\rm cr} \to \delta_{\rm cr} - \Delta_0^{(1)}(t_0) \ . \tag{3.23}$$

We now derive why and when this holds in the context of the exact separate-universe picture described in Sec. 3.2. The threshold $\delta_{\rm cr}$ is defined as the linearly extrapolated initial density contrast of a region collapsing at a given proper time t_0. Since General Relativity is scale-free, this threshold is independent of the size and enclosed mass of the perturbation, unless there is an additional scale in the matter sector, for example the sound horizon of one of the fluid components, which we will ignore here. In an EdS Universe, a spherical perturbation with a present-day linear fractional overdensity $\delta_{\rm cr} \approx 1.686$ collapses at $a = 1$ (this value is derived in Sec. 5.2.1). The same reasoning holds for more general expansion histories, where $\delta_{\rm cr}$ assumes other values (e.g., [220]). Since the evolution of a spherical perturbation is independent of the external Universe (by Birkhoff's theorem) a perturbation of the same physical density $\varrho_{\rm cr} \equiv \overline{\varrho}_m(1 + \delta_{\rm cr})$ will collapse at the same proper time in a Universe with modified background density following Eq. (3.2). We now derive what this implies for the significance $\nu_c = \delta_{\rm cr}/\sigma(M) = (\varrho_{\rm cr} - \overline{\varrho}_m)/\delta\varrho_{\rm RMS}$, which quantifies how rare fluctuations above a physical density threshold $\varrho_{\rm cr}$ are in the *linearly extrapolated initial density field*, given its root-mean-square fluctuation amplitude $\delta\varrho_{\rm RMS} = \sigma(M)\overline{\varrho}_m$. Following our discussion in Sec. 3.2, we add a uniform matter density component $\Delta_0^{(1)}\overline{\varrho}_m$ to the linear density field. Then, the critical overdensity changes to

$$\varrho_{\rm cr} - \tilde{\varrho}_m = (1 + \delta_{\rm cr})\overline{\varrho}_m - (1 + \Delta_0^{(1)})\overline{\varrho}_m = (\delta_{\rm cr} - \Delta_0^{(1)})\overline{\varrho}_m \ . \tag{3.24}$$

Thus, the significance is modified to

$$\tilde{\nu}_c = \frac{\varrho_{\rm cr} - \tilde{\varrho}_m}{\sigma(M)\overline{\varrho}_m} = \frac{\delta_{\rm cr} - \Delta_0^{(1)}}{\sigma(M)} \ . \tag{3.25}$$

For a mass function of the form Eq. (3.21), changing the background density is thus equivalent to changing $\delta_{\rm cr} \to \delta_{\rm cr} - \Delta_0^{(1)}$. Eq. (3.20) applied to Eq. (3.21) thus immediately yields

$$b_N^L(M) = \frac{(-1)^N}{[\sigma(M)]^N} \frac{1}{\nu_c f(\nu_c)} \frac{d^N[\nu_c f(\nu_c)]}{d\nu_c^N} \ . \tag{3.26}$$

This is the widely known expression for the peak-background split Lagrangian bias parameters, which really are a special case of Eq. (3.20). Furthermore, Eq. (3.26) also holds if $f(\nu_c)$ is given by an integral over other variables (such as, for example, the peak curvature), as long as these other variables are independent of $\delta_{\rm cr}$ [152, 221, 100]. Specifically, the result remains valid if $f(\nu_c)$ depends on ratios of spectral moments, e.g. $\sigma_1(M)/\sigma(M)$, since these are not changed by the long-wavelength density perturbation. Eq. (3.26) has been applied extensively to compute the Lagrangian and Eulerian LIMD bias parameters from analytic prescriptions of the multiplicity function; see [14, 15, 16] to only cite a few. We will review some of these prescriptions in Sec. 5 and Sec. 6.

Perhaps the most widely known calculation along these lines is the derivation of the LIMD bias parameters from the Press-Schechter (PS, [217]) and Sheth-Tormen (ST, [16]) mass functions. The Press-Schechter multiplicity function, which can be derived analytically starting from simple physical assumptions (see Sec. 5.2.2), is given by

$$\nu_c f_{\rm PS}(\nu_c) = \sqrt{\frac{2}{\pi}} \, \nu_c e^{-\nu_c^2/2} \ . \tag{3.27}$$

Applying Eq. (3.26) to this, we immediately obtain for the linear and second-order Lagrangian LIMD bias

---

[11]Note that since we are studying an infinite-wavelength perturbation, it does not contribute to the variance $\sigma(M)$; the effect of a finite wavelength of the perturbation is captured by higher-derivative biases.

parameters

$$(b_1^L)_{\text{PS}} = \frac{\nu_c^2 - 1}{\delta_{\text{cr}}}$$

$$(b_2^L)_{\text{PS}} = \frac{\nu_c^2}{\delta_{\text{cr}}^2} \left(\nu_c^2 - 3 \right) . \tag{3.28}$$

We will re-derive these well-known results using the excursion-set formalism in Sec. 5.6.2. We can now map these to Eulerian biases through Eq. (2.16), to obtain

$$(b_1^E)_{\text{PS}} = 1 + \frac{\nu_c^2 - 1}{\delta_{\text{cr}}}$$

$$(b_2^E)_{\text{PS}} = \frac{8}{21} \left(\frac{\nu_c^2 - 1}{\delta_{\text{cr}}} \right) + \frac{\nu_c^2}{\delta_{\text{cr}}^2} \left(\nu_c^2 - 3 \right) . \tag{3.29}$$

Now, the agreement between the Press-Schechter mass function and the abundance of halos in N-body simulations is far from perfect, as is apparent in the left panels of Fig. 9. The same is also true for the linear bias parameter b_1^E of simulated halos when compared with Eq. (3.29) (right panels in Fig. 9). Following the suggestion of [222] that this discrepancy primarily arises because of the limitations of the Press-Schechter model, and not the peak-background-split argument, Sheth and Tormen [16] derived an empirical expression for the multiplicity function which subsequently was shown to be of a functional form expected from a version of the ellipsoidal collapse model ([223, 224]; see Sec. 5.9),

$$\nu_c f_{\text{ST}}(\nu_c) = \sqrt{\frac{2}{\pi}} \left[1 + \frac{1}{2^p \sqrt{\pi}} \Gamma \left(\frac{1}{2} - p \right) \right]^{-1} \left[1 + (q\nu_c^2)^{-p} \right] \sqrt{q}\, \nu_c e^{-q\nu_c^2/2} , \tag{3.30}$$

where $p = 0.3$ and $q = 0.707$ were fitted to the measured halo mass function. Therefore, these parameters generally depend on the precise definition of halos employed in the simulation analysis; we describe the commonly used halo finding algorithms in Appendix D. The PS mass function corresponds to the special case $p = 0$, $q = 1$ in Eq. (3.30). Note that Eq. (3.30) is normalized to satisfy the constraint $\int M \bar{n}_h d \ln M = \bar{\rho}_m$ [as is Eq. (3.27)]. The Eulerian PBS bias parameters b_1^E and b_2^E inferred from the multiplicity function Eq. (3.30) on applying the same procedure as for the PS mass function, Eq. (3.26) and Eq. (2.16), are obtained as

$$(b_1^E)_{\text{ST}} = 1 + \frac{q\nu_c^2 - 1}{\delta_{\text{cr}}} + \frac{2p/\delta_{\text{cr}}}{1 + (q\nu_c^2)^p} \tag{3.31}$$

$$(b_2^E)_{\text{ST}} = \frac{8}{21} \left(\frac{q\nu_c^2 - 1}{\delta_{\text{cr}}} + \frac{2p/\delta_{\text{cr}}}{1 + (q\nu^2)^p} \right) + \frac{q\nu_c^2}{\delta_{\text{cr}}^2} \left(q\nu_c^2 - 3 \right) \tag{3.32}$$

$$+ \frac{2p}{\delta_{\text{cr}}^2} \left(-1 + 2p + 2q\nu_c^2 \right) \frac{1}{1 + (q\nu_c^2)^p} .$$

Ref. [16] found that Eqs. (3.30)–(3.31) provide a much better fit to the mass function and linear bias of halos than the Press-Schechter prediction. Fig. 9 shows the multiplicity function $\nu_c f(\nu_c)$ (left panels) and scaled linear Lagrangian bias b_1^L (right panels) for halos identified at redshift $z_{\text{obs}} = 0, 1, 2$ and 4. Clearly, the Sheth-Tormen mass function is a better fit to the data at all masses and redshifts, and so is the PBS bias prediction based on this mass function. As emphasized by [16], this demonstrates that knowledge of the mass function also means knowledge of clustering, through the peak-background split.

The accuracy of Eq. (3.26) was further quantified in [226, 227], who empirically calibrated the multiplicity function of the halos identified in their simulations. They found that the bias parameters derived from this multiplicity function via Eq. (3.26) are accurate at the 10–20% level. Following our discussion in this section, this limited accuracy is due to the limitations of the universal mass function prescription [Eq. (3.21)], and not due to the peak-background split argument itself, which is exact when defined rigorously.
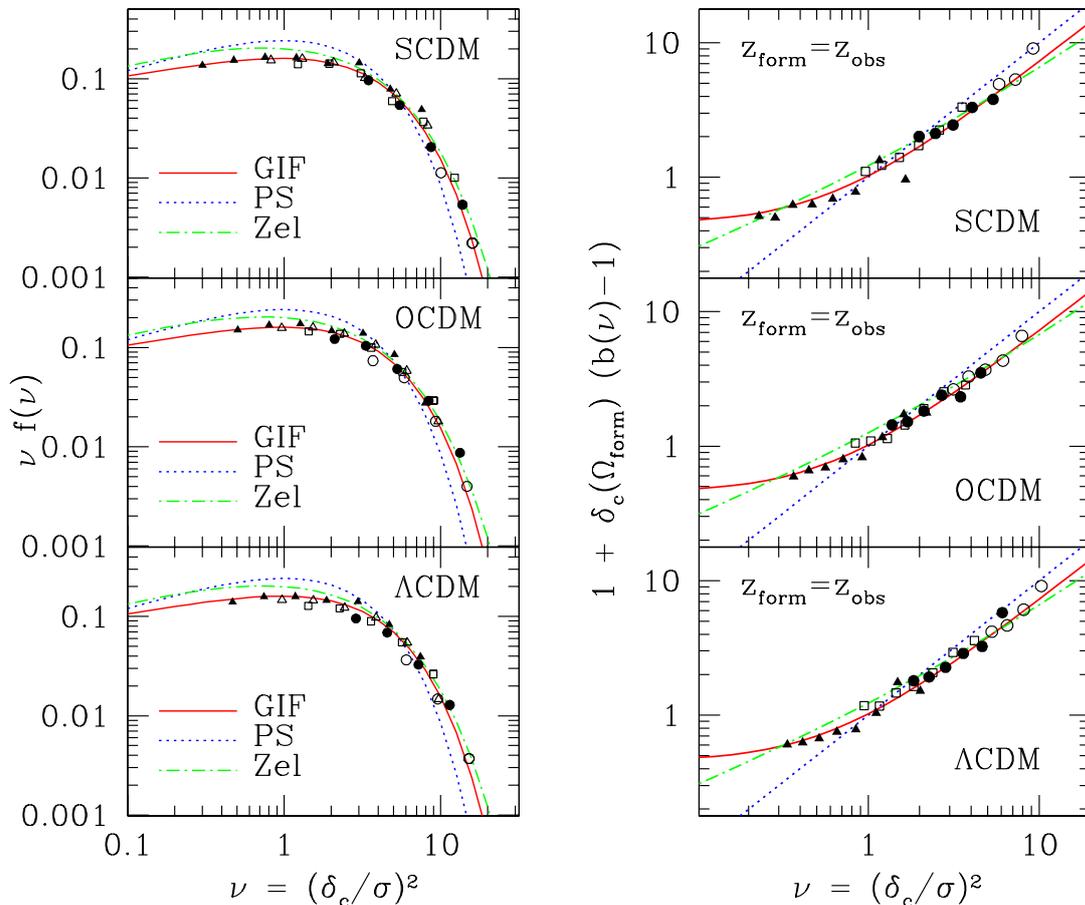
Figure 9: The multiplicity function $f(\nu) \equiv \nu_c f(\nu_c)$ *(left panels)* and the scaled linear Lagrangian LIMD bias $b(\nu) - 1 \equiv b_1^L$ *(right panels)* of halos identified using an FoF halo finder at the output redshifts $z_{\rm obs} = 0$ (filled triangle), 1 (empty square), 2 (filled circle) and 4 (empty circle). Note that $\nu \equiv (\nu_c)^2$ in our notation. The halo bias was measured using the halo power spectrum (Sec. 4.1). b_1^L has been rescaled as indicated by the axis label to show that, for a universal mass function, the dependence of the PBS bias parameters on cosmology arises only through ν_c. The solid (red) curves (labeled "GIF") represent the relation predicted by the Sheth-Tormen mass function [Eq. (3.30) and Eq. (3.31), respectively], whereas the dotted (blue) curves (labeled "PS") show the relations which follow from the Press-Schechter mass function [Eq. (3.27) and Eq. (3.29), respectively]. Finally, the (green) dot-dashed line (labeled "Zel") is the mass function in the Zel'dovich approximation [225]. From [16].

Eq. (3.26) is based on a mass-independent critical density, or *barrier*, $\delta_{\rm cr}$. One can generalize this to a general barrier $B(M)$, so that $\nu_c \equiv B(M)/\sigma(M)$, which is intended to capture other physical effects, such as ellipsoidal collapse and the scatter due to the specific realization of small-scale initial conditions (see Sec. 5.9). We can now make the *assumption* that Eq. (3.25) analogously holds for a general barrier, yielding

$$\tilde{\nu}_c = \frac{B(M) - \Delta_0^{(1)}}{\sigma(M)} \,. \tag{3.33}$$

Then, Eq. (3.26) remains valid even for a mass-dependent barrier. Let us thus summarize the two assumptions made in Eq. (3.21) and Eq. (3.26): (*i*) the abundance of halos depends on the amplitude of the matter power spectrum only through a single moment $\sigma(M)$; (*ii*) the barrier $B(M)$ is unchanged by the presence of a long-wavelength density perturbation, which allows us to derive the bias parameters from the mass

function. Birkhoff's theorem ensures that the second assumption is true for a spherically symmetric setup. However, real halos are not spherically symmetric, and tidal fields cannot be ignored, rendering the second assumption an approximation.

Note that, for a general barrier, we have two free functions at our disposal: $f(\nu_c)$ and $B(M)$. These can, for example, be estimated from the mass function $\overline{n}_h(M)$ and the linear bias $b_1(M)$. In that case, measurements of $b_1(M)$ can be matched exactly, and the PBS biases only become predictive for b_N with $N = 2, 3, \cdots$. We will discuss the comparison with N-body simulations in detail in Sec. 4.4.

3.4   Renormalized biases and the PBS*

We now make the connection between the PBS bias parameters defined as response to the transformation Eq. (3.3) and the renormalized bias parameters defined in Sec. 2.10. We can formally obtain the average number density of galaxies $\overline{n}_g = \langle n_g \rangle$ in a Universe with modified background density $\tilde{\varrho}_m = (1 + \Delta)\overline{\varrho}_m$, starting from the expansion in terms of the coarse-grained fields $\delta_\Lambda, K_{ij,\Lambda}, ...$, Eq. (2.118). For this, we will assume that the coarse-graining scale Λ^{-1} is sufficiently large so that NLO (1-loop) perturbation theory applies. This assumption is being made merely in order to be able to connect to the perturbative relation between bare and renormalized bias parameters. Expanding up to cubic order and keeping terms that depend on Δ, Eq. (2.118) yields

$$\langle n_g \rangle|_\Delta = n_{0,\Lambda} \left[1 + c_{\delta,\Lambda}\langle \delta_\Lambda \rangle_\Delta + c_{\delta^2,\Lambda}\langle \delta_\Lambda^2 \rangle_\Delta + c_{K^2,\Lambda}\langle (K_{ij,\Lambda})^2 \rangle_\Delta + c_{\delta^3,\Lambda}\langle \delta_\Lambda^3 \rangle_\Delta + c_{\delta K^2,\Lambda}\langle \delta_\Lambda (K_{ij,\Lambda})^2 \rangle_\Delta \right.$$
$$\left. + c_{\mathrm{td},\Lambda}\langle O_{\mathrm{td},\Lambda} \rangle_\Delta + \cdots \right]. \tag{3.34}$$

Note that in the presence of Δ, the expectation value of the overdensity $\langle \delta_\Lambda \rangle_\Delta$ no longer vanishes. Here, both $n_{0,\Lambda}$ and the $c_{O,\Lambda}$ refer to the quantities defined in the Universe with background density $\overline{\rho}_m$, i.e. $\Delta = 0$. Let us expand this expression to linear order in Δ around $\Delta = 0$. The expectation values on the right-hand side are easily calculated at cubic order in perturbation theory. For the linear term, we trivially have $\langle \delta_\Lambda \rangle_\Delta = \Delta$. In the cubic terms, we can also directly insert $\delta_\Lambda \to \delta_\Lambda + \Delta$ to obtain

$$\langle \delta_\Lambda^3 \rangle_\Delta = \langle \delta_\Lambda^3 \rangle + 3\langle \delta_\Lambda^2 \rangle \Delta + \mathcal{O}(\Delta^2)$$
$$\langle \delta_\Lambda (K_{ij,\Lambda})^2 \rangle_\Delta = \langle \delta_\Lambda (K_{ij,\Lambda})^2 \rangle + \langle (K_{ij,\Lambda})^2 \rangle \Delta + \mathcal{O}(\Delta^2)$$
$$\langle O_{\mathrm{td},\Lambda} \rangle_\Delta = \frac{16}{21}\langle (K_{ij,\Lambda})^2 \rangle \Delta + \mathcal{O}(\Delta^2), \tag{3.35}$$

where expectation values without subscripts denote those in the fiducial cosmology ($\Delta = 0$), and $\langle (K_{ij,\Lambda})^2 \rangle = (2/3)\langle \delta_\Lambda^2 \rangle$. The last line can be obtained from the definition of O_{td}, Eq. (2.50), and noting that $\mathcal{D}_{ij}[(\delta_\Lambda)^2 - 3/2(K_{ij,\Lambda})^2]$ becomes, upon replacing $\delta_\Lambda(\boldsymbol{x})$ with $\delta_\Lambda(\boldsymbol{x}) + \Delta$, $\mathcal{D}_{ij}[(\delta_\Lambda)^2 - 3/2(K_{ij,\Lambda})^2] + 2(K_{ij,\Lambda} + \delta_\Lambda \delta_{ij}/3)\Delta$.

In order to derive the vacuum expectation value of the quadratic operators, $\langle \delta_\Lambda^2 \rangle_\Delta$, $\langle (K_{ij,\Lambda})^2 \rangle_\Delta$, we need to insert the change in the amplitude of the small-scale power spectrum in the presence of a long-wavelength mode. This can be calculated directly in second-order perturbation theory, either using Eq. (B.14), taking the angle-averaged squeezed limit of the leading-order bispectrum, or calculating the growth factor in a modified background cosmology (Sec. 3.2) [228]. Any of these calculations leads to

$$\langle \delta_\Lambda^2 \rangle_\Delta = \langle \delta_\Lambda^2 \rangle \left(1 + \frac{68}{21}\Delta \right) + \mathcal{O}(\Delta^2) \quad \text{and} \quad \langle K_{ij,\Lambda}^2 \rangle_\Delta = \langle K_{ij,\Lambda}^2 \rangle \left(1 + \frac{68}{21}\Delta \right) + \mathcal{O}(\Delta^2). \tag{3.36}$$

Putting everything together, we obtain

$$\langle n_g \rangle|_\Delta = \langle n_g \rangle \left[1 + c_{\delta,\Lambda}\Delta + \Delta\, \sigma^2(\Lambda)\left\{ \frac{68}{21}\left(c_{\delta^2,\Lambda} + \frac{2}{3}c_{K^2,\Lambda} \right) + 3c_{\delta^3,\Lambda} + \frac{2}{3}c_{\delta K^2,\Lambda} + \frac{32}{63}c_{\mathrm{td},\Lambda} \right\} + \mathcal{O}(\Delta^2) \right], \tag{3.37}$$

---

* This section is of a more technical nature and is not essential for the remainder of the review.

where $\sigma^2(\Lambda) \equiv \langle \delta_\Lambda^2 \rangle$ and we have absorbed the Δ-independent terms into $\langle n_g \rangle$ (consistently at the order we are working in).

Following Eq. (3.4), we now introduce the strict definition of the PBS bias parameters b_N ($N \geq 1$) as the derivative of $\langle n_g \rangle$ with respect to Δ. Evaluating the derivative on Eq. (3.37), we immediately see that b_1 defined through Eq. (3.4) becomes

$$b_1 = c_{\delta,\Lambda} + \sigma^2(\Lambda) \left\{ \frac{68}{21} \left(c_{\delta^2,\Lambda} + \frac{2}{3} c_{K^2,\Lambda} \right) + 3 c_{\delta^3,\Lambda} + \frac{2}{3} c_{\delta K^2,\Lambda} + \frac{32}{63} c_{\mathrm{td},\Lambda} \right\} . \tag{3.38}$$

This exactly coincides with the renormalized bias parameter given in the previous section, Eq. (2.129). Thus, the peak-background split bias parameters understood in the sense of Eq. (3.4) are *exactly the renormalized bias coefficients* which describe large-scale correlations of tracers [79, 100]. This proof can be straightforwardly generalized to b_2, b_3, $\cdots$, which then necessitates higher-order perturbative solutions for δ, K_{ij}, and so on. When neglecting gravitational evolution and performing the renormalization purely in Lagrangian space, Ref. [100] has shown that this result holds to all orders and for all bias parameters b_N. Moreover, the same renormalized biases describe both tracer auto- and cross-correlations as well as higher N-point functions such as the bispectrum, as will be discussed in detail in the following Sec. 4. This is ensured by the renormalization conditions applied in Sec. 2.10.

It is worth emphasizing again the difference between the PBS bias parameters b_N and the bare $c_{\delta^N,\Lambda}$: 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 scale Λ. The $c_{\delta^N,\Lambda}$ on the other hand quantify the average response of the abundance of tracers within a region of size Λ^{-1} to a change in the average density δ_Λ within that region, evaluated at $\delta_\Lambda = 0$; they thus necessarily depend on the cutoff, i.e. the filtering kernel W_Λ and scale Λ^{-1}. On the other hand, after renormalization we only need a prediction for $\langle n_g \rangle$ as function of the background density $\bar{\varrho}_m$ to calculate the LIMD bias parameters b_N. Note also that the b_N are closely related to the resummed bias propagators defined in [113] [see Eqs. (83)–(84) there], while the bare bias parameters c_n correspond to the bare propagators [Eqs. (1)–(2) in that paper].

4 Measuring galaxy and halo bias

Having described the general framework of the bias expansion (Sec. 2), and the physical interpretation of bias parameters offered by the peak-background split argument (Sec. 3), we now turn to the connection with clustering statistics and actual measurements of the bias parameters. We will discuss various observables which involve bias parameters and can be used to measure them, including the auto-correlation functions of galaxies and cross-correlations between galaxies and matter, as well as moments-based and "scatter plot" methods. In each case, we define the range of scales over which the perturbative bias expansion holds, and derive how the parameters inferred in each method are precisely related to the bias parameters defined in Sec. 2. Since this is an important point, let us emphasize again: on the scales where perturbation theory applies, and barring any systematics in the measurement, *all of the various methods to estimate bias recover the unique, large-scale bias parameters b_O defined in Sec. 2.* This holds, of course, up to linear combinations of bias parameters at a given order, which are a matter of choice of basis; we give relations between various popular choices in Appendix C. In this sense, the bias parameters are *physical quantities rather than observable-dependent fitting parameters.* This is not the case for phenomenological fitting relations between, for example, the galaxy power spectrum and the matter power spectrum for a fiducial cosmological model (such as, for instance, the "Q model" of [25]). We will not discuss these fitting functions here.

The final goal of the bias expansion is, of course, to predict observed galaxy statistics given a set of bias parameters as well as a cosmological model with its associated predictions for the statistics of matter. In this section, we will simplify the treatment by ignoring complications that, while important for actual observed galaxy statistics, are not directly related to bias, the most important being *redshift-space distortions.* These, along with other issues affecting observed galaxy statistics, will be discussed in Sec. 9. The idealized relations given below are thus strictly valid only for tracers extracted from numerical simulations—such as dark matter halos—for which we have access to the real-space clustering. Correspondingly, we will denote the tracer overdensity as δ_h in this section, and mostly review measurements of halo bias here. These also provide the context for the models of halo bias discussed in the following sections Sec. 5–6. We emphasize again that *all results are applicable also to galaxy statistics* once the effects described in Sec. 9 are added in.

Furthermore, we continue to restrict ourselves to the case of Gaussian initial conditions in this section. Results for galaxy clustering with primordial non-Gaussianity are discussed in Sec. 7. We will focus on bias parameters up to third order, as derived in Sec. 2.4 and summarized in Sec. 2.11. This is essentially the current state-of-the-art in the published literature on galaxy and halo bias. The lowest-order biases are phenomenologically the most important ones. While linear- and second-order terms are sufficient to describe the galaxy power spectrum and bispectrum (three-point function) at leading order, respectively, cubic order terms contribute to the next-to-leading (1-loop) correction to the galaxy power spectrum.

At this order, the galaxy as well as halo density contrast can be expressed as

$$
\begin{aligned}
\delta_h = {} & b_1\delta + b_{\nabla^2\delta}\nabla^2\delta + \varepsilon \\
& + \frac{1}{2}b_2\delta^2 + b_{K^2}(K_{ij})^2 + \varepsilon_\delta\delta \\
& + \frac{1}{6}b_3\delta^3 + b_{\delta K^2}\delta(K_{ij})^2 + b_{K^3}(K_{ij})^3 + b_{\mathrm{td}}O^{(3)}_{\mathrm{td}} + \varepsilon_{\delta^2}\delta^2 + \varepsilon_{K^2}(K_{ij})^2 \\
& + \mathcal{O}(\delta^4) + \mathcal{O}[R_*^2\nabla^2(\delta^2),\, R_*^4\nabla^4\delta]\,,
\end{aligned}
\tag{4.1}
$$

where ε and ε_O are stochastic fields that are uncorrelated with δ, K_{ij} (see Sec. 2.8), and we have listed, in the last line, examples of higher-order terms that are neglected in this expression. Although not made explicit here, the operators appearing in Eq. (4.1) are strictly the *renormalized* operators, and the b_O are consequently renormalized bias parameters. Indeed, as we are dealing with measurements as well as physical models of bias in the following, all instances of the bias parameters appearing in the remainder of the review are renormalized (with the exception of Sec. 7.1.2, where, for the purposes of derivation, the bare bias parameters appear).

Note that all observables, fields, and bias parameters depend on the time τ of observation. For clarity, however, we will drop the time argument throughout this section. We divide the various procedures to measure bias parameters into four categories:

1. *n-point correlation functions:* two- and three-point correlation functions, or power spectrum and bispectrum in Fourier space (Appendix A);

2. *Moments:* one-point statistics of the halo and matter density smoothed on a large scale R_ℓ;

3. *Scatter plots* of the halo density as a function of the local matter density;

4. *Responses:* an implementation of the separate-universe approach (or exact PBS) described in Sec. 3.2.

We will successively describe each approach in Sec. 4.1–4.4. We then present a brief overview of published results for bias parameters (and stochastic amplitudes) obtained using these various methods for dark matter halos in simulations, and observed galaxies, in Sec. 4.5. Finally, in Sec. 4.6 we will discuss the phenomenon of *assembly bias*, which is specific to dark matter halos. This refers to the fact that the bias parameters of halos depend on other halo properties than just their mass.

4.1 n-point correlation functions

We begin with the measurement of bias parameters from halo n-point correlation functions. This approach is most closely related to the discussion of Sec. 2. The lowest-order statistics that allow us to unambiguously measure the first- and second-order bias parameters are, respectively, the two- and three-point functions. We consider both halo auto-correlations as well as cross-correlations with matter. Once projection effects are included, auto-correlations are readily measurable for galaxies in real survey data. In addition, measuring stacked weak gravitational lensing around galaxies (*galaxy-galaxy lensing*) yields the galaxy-matter cross-correlation function projected along the line of sight [229], as will be briefly discussed below.

Apart from the close connection to measurements, the advantage of n-point functions is that, upon restricting all scales r_i to be greater than a minimum scale $r_{\min}$, or, in Fourier space, all wavevectors k_i to be less than some prescribed $k_{\max}$, one has complete control over nonlinear and higher-derivative corrections. Moreover, the measurements can be made even more robust by marginalizing over the leading nonlinear corrections which have a known functional form, as we will see in Sec. 4.1.4.

On the other hand, the main practical disadvantage of this method is that large observed or simulated volumes are necessary to obtain a high signal-to-noise measurement of n-point statistics on scales above $r_{\min}$ and/or below $k_{\max}$. This is because we need a significant number of independent modes in order to reduce sample variance. In addition, measuring the bias parameters at cubic or higher order becomes increasingly

difficult, since measurements of higher-order statistics become necessary, for instance the trispectrum in the case of cubic-order bias parameters. The implementation and the required computational resources for higher-order statistics become increasingly demanding.

Since this is a substantial subsection, we provide a brief outline here. We begin with the leading two- and three-point functions in Eulerian space, both in the Fourier- and real-space representations (Sec. 4.1.1). We then briefly discuss the corresponding results in Lagrangian space (Sec. 4.1.2), which are relevant for estimating bias parameters from halos identified in N-body simulations. Sec. 4.1.3 then provides a quantitative, albeit simplified and idealized, forecast of the ability of current and future galaxy surveys to measure the bias parameters and amplitude of the matter power spectrum using the results of Sec. 4.1.1. Next, we derive the next-to-leading correction to the galaxy two-point function (1-loop power spectrum) in Sec. 4.1.4, illustrating how the predictions of Sec. 4.1.1 can be taken to higher order and what scalings the higher-order terms obey.

4.1.1 Two- and three-point functions at leading order

We begin with the leading-order (LO), or tree-level, predictions for the power spectrum and bispectrum of halos, that is, the two- and three-point correlation functions in Fourier space. The leading-order calculation of the halo power spectrum and bispectrum requires, respectively, linear- and second-order perturbation theory (see Appendix B). These leading-order predictions are accurate on sufficiently large scales, roughly at the level of 10% for $k \lesssim 0.03\,h\,\mathrm{Mpc}^{-1}$ in Fourier space at $z = 0$ (a more precise calculation is the subject of Sec. 4.1.4); the range increases at higher redshifts [101]. We will present the corresponding real-space results, the correlation functions, at the end of this section.

The halo auto-power spectrum and halo-matter cross-power spectrum are given by

$$
\begin{aligned}
P_{hh}^{\mathrm{LO}}(k) &\equiv \langle \delta_h(\boldsymbol{k}) \delta_h(\boldsymbol{k}') \rangle'_{\mathrm{LO}} = b_1^2 P_{\mathrm{L}}(k) + P_\varepsilon^{\{0\}} \\
P_{hm}^{\mathrm{LO}}(k) &\equiv \langle \delta_h(\boldsymbol{k}) \delta_m(\boldsymbol{k}') \rangle'_{\mathrm{LO}} = b_1 P_{\mathrm{L}}(k) \,,
\end{aligned}
\tag{4.2}
$$

where, here and throughout, a prime on an expectation value denotes that the momentum-conserving Dirac delta, $(2\pi)^3 \delta_D(\boldsymbol{k} + \boldsymbol{k}')$ in case of Eq. (4.2), is to be dropped (see Tab. 2). As mentioned in the introduction, we drop the time argument throughout this section for clarity. Again, we would obtain the same relation for galaxies if we were able to measure their proper rest-frame density at the true physical position, that is, without redshift-space distortions and other projection effects. $P_\varepsilon^{\{0\}} = \lim_{k\to 0} \langle \varepsilon(\boldsymbol{k}) \varepsilon(\boldsymbol{k}') \rangle'$ is the scale-independent large-scale stochastic contribution [see Eq. (2.83) in Sec. 2.8]. Note that this is a renormalized stochastic term which absorbs scale-independent terms from higher loop integrals (see Sec. 4.1.4). We will discuss $P_\varepsilon^{\{0\}}$ in more detail in Sec. 4.5.3. The next-to-leading-order corrections to $P_{hh}(k)$ as well as $P_{hm}(k)$ from nonlinear evolution of both matter and bias, and from higher-derivative biases, will be described in Sec. 4.1.4.

Since the halo stochasticity contributes to the halo auto-power spectrum $P_{hh}(k)$ but not to the halo-matter cross-power spectrum $P_{hm}(k)$, the latter offers the simplest and cleanest measurement of the linear bias parameter b_1 for halos (see e.g. [230, 231, 127]). This technique can also be applied to galaxies, by measuring the matter distribution through weak gravitational lensing, specifically, the cross-correlation ("galaxy-galaxy lensing") of the projected galaxy density with the tangential shear measured from source galaxies at higher redshifts [94, 232, 233, 234, 235, 236] (see [237] for a recent review). Briefly, for lens galaxies at a known comoving distance χ_L and source galaxies following a normalized redshift distribution $p(z)$, the stacked tangential shear around galaxies in angular multipole space corresponds to a projection of the real-space galaxy-matter power spectrum, $P_{gm} = b_1 P_{mm}$ at leading order, given in the Limber approximation [238] by

$$
C_{g\gamma}(l) = \frac{3}{2} \Omega_{m0} H_0^2 \left[\int dz\, p(z) \frac{\chi(z) - \chi_L}{\chi(z)} \right] \frac{1 + z(\chi_L)}{\chi_L} P_{gm}\left(k = \frac{\sqrt{l(l+1)}}{\chi_L}, z(\chi_L) \right) .
\tag{4.3}
$$

By itself, this observable suffers from a degeneracy between b_1 and the matter power spectrum normalization. This degeneracy can be broken by including the projected auto-correlation of galaxies $C_{gg}(l)$, and/or the cosmic shear power spectrum $C_{\gamma\gamma}(l)$, as recently applied in [239, 31].

74

Alternatively, the degeneracy between bias and amplitude of fluctuations can be broken by measuring the three-point correlation function (in real space; [240, 241]) or bispectrum (in Fourier space; [242]). Statistical homogeneity and isotropy dictate that the bispectrum depends on three parameters which describe the shape and scale of a triangle. The bispectrum thus encodes much more information than the two-point function, which is a function of one scale only. Moreover, the leading nonlinear (second-order) effects of bias become apparent in the bispectrum. Finally, the bispectrum contains interesting cosmological information in its own right [61]. The leading-order (tree-level) expressions for the matter-matter-halo, matter-halo-halo, and halo-halo-halo bispectra are respectively given by

$$
\begin{aligned}
B^{\mathrm{LO}}_{mmh}(k_1, k_2, k_3) &\equiv \langle \delta(\boldsymbol{k}_1)\delta(\boldsymbol{k}_2)\delta_h(\boldsymbol{k}_3)\rangle'_{\mathrm{LO}} \\
&= b_1 B^{\mathrm{LO}}_{mmm}(k_1, k_2, k_3) + \left[b_2 + 2b_{K^2}\left(\left[\hat{\boldsymbol{k}}_1 \cdot \hat{\boldsymbol{k}}_2\right]^2 - \frac{1}{3}\right)\right] P_{\mathrm{L}}(k_1)P_{\mathrm{L}}(k_2) \\
B^{\mathrm{LO}}_{mhh}(k_1, k_2, k_3) &\equiv \langle \delta(\boldsymbol{k}_1)\delta_h(\boldsymbol{k}_2)\delta_h(\boldsymbol{k}_3)\rangle'_{\mathrm{LO}} \\
&= b_1^2 B^{\mathrm{LO}}_{mmm}(k_1, k_2, k_3) + 2P_{\mathrm{L}}(k_1)P^{\{0\}}_{\varepsilon\varepsilon_\delta} \\
&\quad + \left\{ b_1 \left[b_2 + 2b_{K^2}\left(\left[\hat{\boldsymbol{k}}_1 \cdot \hat{\boldsymbol{k}}_2\right]^2 - \frac{1}{3}\right)\right] P_{\mathrm{L}}(k_1)P_{\mathrm{L}}(k_2) + (\boldsymbol{k}_2 \leftrightarrow \boldsymbol{k}_3)\right\} \\
B^{\mathrm{LO}}_{hhh}(k_1, k_2, k_3) &\equiv \langle \delta_h(\boldsymbol{k}_1)\delta_h(\boldsymbol{k}_2)\delta_h(\boldsymbol{k}_3)\rangle'_{\mathrm{LO}} \\
&= b_1^3 B^{\mathrm{LO}}_{mmm}(k_1, k_2, k_3) + B^{\{0\}}_{\varepsilon} \\
&\quad + \left\{ b_1^2\left[b_2 + 2b_{K^2}\left(\left[\hat{\boldsymbol{k}}_1 \cdot \hat{\boldsymbol{k}}_2\right]^2 - \frac{1}{3}\right)\right] P_{\mathrm{L}}(k_1)P_{\mathrm{L}}(k_2) + 2\,b_1 P^{\{0\}}_{\varepsilon\varepsilon_\delta}P_{\mathrm{L}}(k_2)\right\} + 2\text{ perm.}.
\end{aligned}
\tag{4.4}
$$

In the first two lines, we choose k_3, and k_2, k_3, respectively, to refer to halo overdensities. Note that the bispectra B_{mmh} and B_{mhh} are symmetric under interchange of k_1, k_2 and k_2, k_3, respectively, while B_{hhh} is symmetric under interchange of any of the k_i. The leading-order matter bispectrum entering Eq. (4.4) is given by

$$
B^{\mathrm{LO}}_{mmm}(k_1, k_2, k_3) = 2F_2(\boldsymbol{k}_1, \boldsymbol{k}_2)P_{\mathrm{L}}(k_1)P_{\mathrm{L}}(k_2) + 2\text{ perm.},
\tag{4.5}
$$

where the F_2 kernel is defined in Eq. (B.13). Eq. (4.4) contains the leading stochastic terms discussed in Sec. 2.8 [cf. Eq. (2.85)], where

$$
P^{\{0\}}_{\varepsilon\varepsilon_\delta} \equiv \lim_{k\to 0} \langle \varepsilon(\boldsymbol{k})\varepsilon_\delta(\boldsymbol{k})\rangle' \quad \text{and} \quad B^{\{0\}}_{\varepsilon} \equiv \lim_{k_1,k_2\to 0} \langle \varepsilon(\boldsymbol{k}_1)\varepsilon(\boldsymbol{k}_2)\varepsilon(\boldsymbol{k}_3)\rangle'
\tag{4.6}
$$

are the cross-power spectrum of the leading (ε) and next-to-leading (ε_δ) stochastic fields, and the bispectrum of ε, respectively; both of these are scale-independent in the large-scale limit. The significance of ε_δ was discussed in Sec. 2.3, and we will return to the stochastic contributions in Sec. 4.5.3 below. In analogy with P_{hm}, there are no stochastic contributions to B_{mmh}, making this the cleanest statistic to measure b_1, b_2, and b_{K^2} for halos. For galaxies, if lensing data are available, the projected, real-space matter-matter-galaxy bispectrum B_{mmg} can be measured by constructing the galaxy-shear-shear cross-correlation.

The scale and configuration dependence of the halo bispectrum allows for the degeneracy between the bias and amplitude of fluctuations to be broken. Fig. 10 shows the dependence of the different contributions to B^{LO}_{hhh} on (k_1, k_2, k_3). The filled region in each panel is defined by the triangle inequality ($k_1 \leq k_2 + k_3$) with the condition $k_1 \geq k_2 \geq k_3$ that we impose without loss of generality. We also show names that are commonly used to refer to special triangle configurations; these will again play a role when discussing primordial non-Gaussianity in Sec. 7. From top to bottom we plot the contributions to Eq. (4.4) that are proportional to, respectively, b_1^3, $b_1^2 b_2$, $b_1^2 b_{K^2}$ and $b_1 P^{\{0\}}_{\varepsilon\varepsilon_\delta}$ for two different scales: the left-hand side plots are for fixed $k_1 = 0.01\,h\,\mathrm{Mpc}^{-1}$, corresponding to a scale larger than the matter-radiation equality turn-over of the power spectrum, while the right-hand side plots are for fixed $k_1 = 0.05\,h\,\mathrm{Mpc}^{-1}$, showing the shape dependence on smaller scales ($k_1 > k_{\mathrm{eq}}$). In all panels, we normalize the bispectrum to have a maximum value of 1, in order to highlight the shape dependence. One can find a detailed explanation for the shape

75
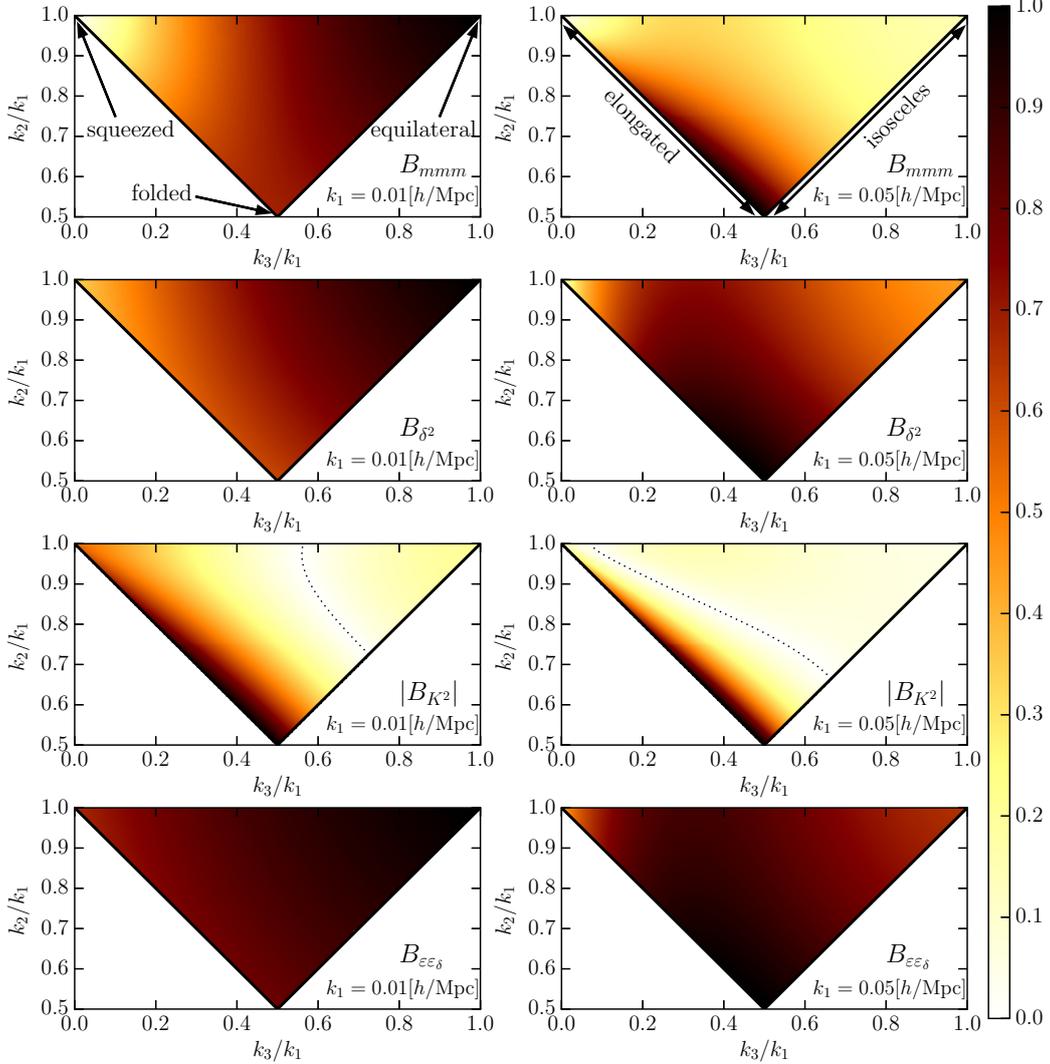
Figure 10: Illustration of the dependence of the different contributions to B_{hhh}^{LO} [Eq. (4.4)] on triangle configuration, where $k_1 \geq k_2 \geq k_3$. The color scale shows the magnitude of each contribution at fixed $k_1 = 0.01\, h\, \mathrm{Mpc}^{-1}$ (left column) and $k_1 = 0.05\, h\, \mathrm{Mpc}^{-1}$ (right column). We divide each amplitude by the maximum value attained. Commonly used designations for certain triangle configurations are indicated in the first row. Shown are, from top to bottom: (i) B_{mmm}^{LO} [Eq. (4.5)], (ii) $B_{\delta^2}(k_1, k_2, k_3) \equiv P_{\mathrm{L}}(k_1) P_{\mathrm{L}}(k_2) + 2$ perm., (iii) $B_{K^2}(k_1, k_2, k_3) \equiv ([\hat{\boldsymbol{k}}_1 \cdot \hat{\boldsymbol{k}}_2]^2 - 1/3) P_{\mathrm{L}}(k_1) P_{\mathrm{L}}(k_2) + 2$ perm., and (iv) $B_{\varepsilon\varepsilon\delta} \equiv P_{\varepsilon\varepsilon\delta}^{\{0\}}[P_{\mathrm{L}}(k_2) + 2$ perm.]. The prefactor $P_{\varepsilon\varepsilon\delta}^{\{0\}}$ is irrelevant in this representation. The dotted lines in the third row indicate the zero-crossing of B_{K^2}; B_{K^2} is positive (negative) on the left- (right)-hand side of the dotted lines, respectively.

dependence of B_{mmm}^{LO} in [243]. The different shape dependences are clearly visible by eye. Thus, a precise measurement of B_{hhh}^{LO} in principle yields clean measurements of b_1, b_2, b_{K^2} and $P_{\varepsilon\varepsilon\delta}^{\{0\}}$ that are independent of the power spectrum normalization. We will quantify these statements in Sec. 4.1.3.

An alternative approach to measuring and fitting Eqs. (4.2)–(4.4) was proposed in [244] and applied in [245], who considered the cross-power spectrum of $\delta_h(\boldsymbol{k})$ with the Fourier transform of squared smoothed fields $(\delta_{R_\ell}^2)(\boldsymbol{k}), [(K_{R_\ell}^{ij})^2](\boldsymbol{k})$, where R_ℓ is a smoothing scale that is sufficiently large to suppress the higher-order nonlinear contributions. This method is an extension of efficient higher-order correlation function estimators used in the analysis of CMB temperature and polarization fields [246, 247]. An advantage of this approach is that one can circumvent a direct measurement of all triangle configurations of the bispectrum,[12] and that it can be extended to extract information in higher-order statistics such as the trispectrum. The downside of this approach is that nonlinear and scale-dependent corrections are less under control than in the full n-point approach, as the nonlinear operation (squaring) performed in real space corresponds to a convolution in Fourier space, which leads to additional high-k contributions, although they are suppressed by the smoothing.

Finally, the bias parameters can equivalently be measured from real-space statistics. The leading-order halo two-point auto-correlation function and the halo-matter cross-correlation function are given by the Fourier transform of Eq. (4.2), or directly from the bias expansion Eq. (4.1):

$$\xi_{hh}^{\mathrm{LO}}(r) = b_1^2 \xi_{\mathrm{L}}(r), \quad \xi_{hm}^{\mathrm{LO}}(r) = b_1 \xi_{\mathrm{L}}(r), \tag{4.7}$$

where the linear matter two-point correlation function is given by

$$\xi_{\mathrm{L}}(r) = \int_{\boldsymbol{k}} P_{\mathrm{L}}(k) e^{i\boldsymbol{k}\cdot\boldsymbol{r}} = \frac{1}{2\pi^2} \int k^2 dk P_{\mathrm{L}}(k) j_0(kr). \tag{4.8}$$

Here and throughout, we use the shorthand $\int_{\boldsymbol{k}} \equiv \int d^3\boldsymbol{k}/(2\pi)^3$ (see Tab. 4). Note that a constant stochastic contribution $P_\varepsilon^{\{0\}}$ in the power spectrum Eq. (4.2) contributes a term $P_\varepsilon^{\{0\}}\delta_D(\boldsymbol{r})$ to the two-point correlation function, and we have not written it here as it only contributes at vanishing, or very small, separation (see Sec. 2.8 for a detailed discussion). Nevertheless, this contribution is important when evaluating the covariance of ξ_{hh}, which enters the likelihood that needs to be evaluated to obtain the best-fitting b_1 and its measurement uncertainty. It also enters the moments of halos (Sec. 4.2). The three-point functions are similarly obtained from the bias expansion Eq. (4.1), or alternatively from the Fourier transform of Eq. (4.4), yielding [251, 252, 253, 254]

$$\xi_{mmh}^{(3),\mathrm{LO}}(r_1, r_2, r_3) \equiv \langle \delta(\boldsymbol{x})\delta(\boldsymbol{x}+\boldsymbol{r}_1)\delta_h(\boldsymbol{x}+\boldsymbol{r}_2)\rangle_{\mathrm{LO}}$$
$$= b_1 \xi_m^{(3),\mathrm{LO}}(r_1, r_2, r_3) + b_2 \xi_{\mathrm{L}}(r_2)\xi_{\mathrm{L}}(r_3) + 2b_{K^2}\xi_2^{(0)}(r_2)\xi_2^{(0)}(r_3)\left[\mu_{23}^2 - \frac{1}{3}\right]$$

$$\xi_{mhh}^{(3),\mathrm{LO}}(r_1, r_2, r_3) \equiv \langle \delta(\boldsymbol{x})\delta_h(\boldsymbol{x}+\boldsymbol{r}_1)\delta_h(\boldsymbol{x}+\boldsymbol{r}_2)\rangle_{\mathrm{LO}}$$
$$= b_1^2 \xi_m^{(3),\mathrm{LO}}(r_1, r_2, r_3) + b_1 b_2 \Big\{ \xi_{\mathrm{L}}(r_1)\xi_{\mathrm{L}}(r_3) + \xi_{\mathrm{L}}(r_2)\xi_{\mathrm{L}}(r_3) \Big\}$$
$$+ 2b_1 b_{K^2} \left\{ \xi_2^{(0)}(r_1)\xi_2^{(0)}(r_3)\left[\mu_{13}^2 - \frac{1}{3}\right] + \xi_2^{(0)}(r_2)\xi_2^{(0)}(r_3)\left[\mu_{23}^2 - \frac{1}{3}\right] \right\}$$

$$\xi_{hhh}^{(3),\mathrm{LO}}(r_1, r_2, r_3) \equiv \langle \delta_h(\boldsymbol{x})\delta_h(\boldsymbol{x}+\boldsymbol{r}_1)\delta_h(\boldsymbol{x}+\boldsymbol{r}_2)\rangle_{\mathrm{LO}}$$
$$= b_1^3 \xi_m^{(3),\mathrm{LO}}(r_1, r_2, r_3) + b_1^2 b_2 \xi_{\delta^2}^{(3)}(r_1, r_2, r_3) + 2b_1^2 b_{K^2}\xi_{K^2}^{(3)}(r_1, r_2, r_3),$$
$$\text{where} \quad r_1 = |\boldsymbol{r}_1|, \ r_2 = |\boldsymbol{r}_2|, \ r_3 = |\boldsymbol{r}_1 - \boldsymbol{r}_2|, \ \text{and}$$
$$\xi_{\delta^2}^{(3)}(r_1, r_2, r_3) = \xi_{\mathrm{L}}(r_1)\xi_{\mathrm{L}}(r_2) + 2 \text{ perm.}$$
$$\xi_{K^2}^{(3)}(r_1, r_2, r_3) = \xi_2^{(0)}(r_1)\xi_2^{(0)}(r_2)\left[\mu_{12}^2 - \frac{1}{3}\right] + 2 \text{ perm.}, \tag{4.9}$$

---

[12]This can be a time-consuming process when done in a brute-force way. See [248, 249, 250] for efficient bispectrum estimators.
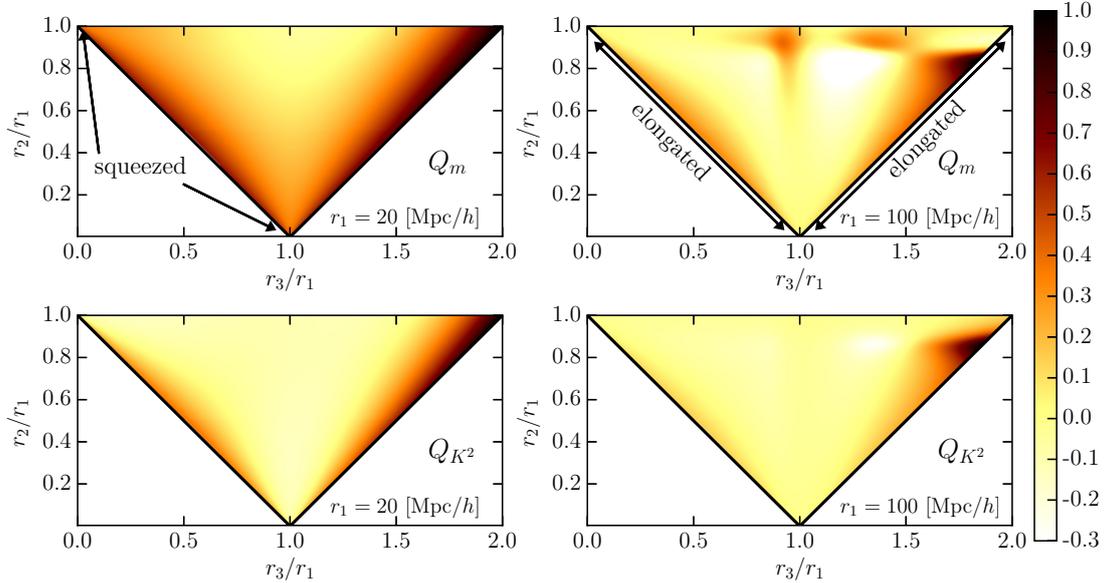
Figure 11: Shape dependence of two contributions to the reduced halo three-point correlation function: Q_m [Eq. (4.11)] (top panels), and $Q_{K^2} = \xi^{(3)}_{K^2}(r_1, r_2, r_3)/\xi^{(3)}_{\delta^2}(r_1, r_2, r_3)$ (bottom panels), for two different scales: $r_1 = 20\,[\text{Mpc}/h]$ (left panels) and $r_1 = 100\,[\text{Mpc}/h]$ (right panels). Here, we show all possible triangle configurations satisfying $r_2 \leq r_1$ for a given r_1. In order to highlight the shape dependence, we divide each panel by the maximum value. A horizontal cut $r_2/r_1 \equiv \text{const.}$ in the $\xi^{(3),\text{LO}}_m$ plot with $r_1 = 20\,[\text{Mpc}/h]$ reveals the well-known "U-shaped" curves of the three-point function. The features seen in the upper right panel (with $r_1 = 100\,[\text{Mpc}/h]$) are due to the BAO feature in the linear correlation function.

where $\mu_{ij} \equiv \hat{\boldsymbol{r}}_i \cdot \hat{\boldsymbol{r}}_j$ is the cosine between the two vectors $\boldsymbol{r}_i$ and $\boldsymbol{r}_j$, $\xi^{(0)}_2(r) \equiv \int_{\boldsymbol{k}} P_{\text{L}}(k) j_2(kr)$, which is is a special case of Eq. (6.4) without smoothing ($R \rightarrow 0$), and

$$\xi^{(3),\text{LO}}_m(r_1, r_2, r_3) = \left\{ \frac{34}{21} \xi_{\text{L}}(r_1)\xi_{\text{L}}(r_2) + \frac{4}{7}\left[\mu^2_{12} - \frac{1}{3}\right]\xi^{(0)}_2(r_1)\xi^{(0)}_2(r_2) \right.$$
$$\left. - \frac{1}{3}\mu_{12}\left[\left(\xi_{\text{L}}(r_1) + \xi^{(0)}_2(r_1)\right)r_1\frac{d\xi_{\text{L}}(r_2)}{dr_2} + (1 \leftrightarrow 2)\right] \right\} + 2 \text{ perm.} \qquad (4.10)$$

is the three-point correlation function of the matter density field at leading order (see, e.g., App. A of [254]; we have used that $j_1(x)/x = [j_0(x) + j_2(x)]/3$). Note that, similar to the case of the bispectrum, statistical homogeneity and isotropy demand that the three-point correlation functions only depend on the three separations r_i. Again, we did not include the stochastic terms that are non-zero only if at least one of the r_i are very small (i.e., non-perturbative).

In order to illustrate its shape dependence, we define the reduced halo three-point function $Q_{hhh}(r_1, r_2, r_3) \equiv \xi_{hhh}(r_1, r_2, r_3)/\xi^{(3)}_{\delta^2}$, and similarly for the individual contributions defined in Eqs. (4.9)–(4.10). Further,

$$Q_m(r_1, r_2, r_3) \equiv \frac{\xi^{(3,\text{LO})}_m(r_1, r_2, r_3)}{\xi_{\text{L}}(r_1)\xi_{\text{L}}(r_2) + 2 \text{ perm.}} = \frac{\xi^{(3,\text{LO})}_m(r_1, r_2, r_3)}{\xi^{(3)}_{\delta^2}(r_1, r_2, r_3)}. \qquad (4.11)$$

The nontrivial contributions are shown in Fig. 11. Here, we impose the condition $r_2 \leq r_1$, and show all regions that satisfy the triangle conditions. The configuration dependence of each term is clearly distinct, similar to the case in Fourier space. The interpretation of Fig. 11 is, however, complicated by the fact that the covariance of the three-point function is not diagonal, so that different triangles are significantly correlated.

78

On the other hand, the different configurations of the bispectrum shown in Fig. 10 are independent at leading order.

In this context, it is worth reiterating that the expressions for the power spectrum and bispectrum Eqs. (4.2)–(4.4) are only strictly correct on large scales. It is not consistent to extend these to smaller scales by inserting the nonlinear matter power spectrum and bispectrum calibrated with N-body simulations, as done in some references (e.g., [255, 256]). This is because many higher-order and higher-derivative bias parameters are implicitly set to zero in this procedure. While this might provide a good empirical match to halos in simulations, it is not guaranteed to be a good description of the statistics of galaxies. Moreover, this procedure will in general lead to inconsistent bias parameters when these are measured from different statistics (e.g., power spectrum and bispectrum), making it impossible to use the synergy between different statistics.

Finally, while we have restricted to the two- and three-point functions here, the four-point function (trispectrum) at leading order similarly allows for measurements of the cubic bias parameters written in Eq. (4.1). Unfortunately, the trispectrum has a small signal-to-noise ratio on large scales once taking into account the covariance with the two-point function as well as the survey geometry [60], and no full measurement for either galaxies or halos has been published so far. Hence we do not consider it further here.

4.1.2 Two- and three-point functions in Lagrangian space

In case of dark matter halos identified in N-body simulations, we can also measure n-point functions in Lagrangian space, by tracing the halos back to the initial conditions. Then, the contributions due to nonlinear evolution of the density field disappear in Eqs. (4.2)–(4.4), and the expressions for the n-point functions simplify. In particular, the leading-order two- and three-point functions become

$$P_{hh}^{L,\mathrm{LO}}(k) \equiv \langle \delta_h^L(\boldsymbol{k}) \delta_h^L(\boldsymbol{k}') \rangle_{\mathrm{LO}}' = (b_1^L)^2 P_{\mathrm{L}}(k) + P_{\varepsilon^L}^{\{0\}}$$

$$P_{hm}^{L,\mathrm{LO}}(k) \equiv \langle \delta_h^L(\boldsymbol{k}) \delta^{(1)}(\boldsymbol{k}') \rangle_{\mathrm{LO}}' = b_1^L P_{\mathrm{L}}(k)$$

$$B_{mmh}^{L,\mathrm{LO}}(k_1, k_2, k_3) = \left[b_2^L + 2b_{K^2}^L \left(\left[\hat{\boldsymbol{k}}_1 \cdot \hat{\boldsymbol{k}}_2 \right]^2 - \frac{1}{3} \right) \right] P_{\mathrm{L}}(k_1) P_{\mathrm{L}}(k_2) \tag{4.12}$$

$$B_{mhh}^{L,\mathrm{LO}}(k_1, k_2, k_3) = b_1^L \left[b_2^L + 2b_{K^2}^L \left(\left[\hat{\boldsymbol{k}}_1 \cdot \hat{\boldsymbol{k}}_2 \right]^2 - \frac{1}{3} \right) \right] P_{\mathrm{L}}(k_1) P_{\mathrm{L}}(k_2) + P_{\varepsilon^L \varepsilon_\delta^L}^{\{0\}} P_{\mathrm{L}}(k_1) + (\boldsymbol{k}_2 \leftrightarrow \boldsymbol{k}_3)$$

$$B_{hhh}^{L,\mathrm{LO}}(k_1, k_2, k_3) = (b_1^L)^2 \left[b_2^L + 2b_{K^2}^L \left(\left[\hat{\boldsymbol{k}}_1 \cdot \hat{\boldsymbol{k}}_2 \right]^2 - \frac{1}{3} \right) \right] P_{\mathrm{L}}(k_1) P_{\mathrm{L}}(k_2)$$
$$+ 2b_1^L P_{\varepsilon^L \varepsilon_\delta^L}^{\{0\}} P_{\mathrm{L}}(k_1) + 2 \text{ perm.} + B_{\varepsilon^L}^{\{0\}},$$

where Lagrangian statistics are denoted with a superscript L, and $P_{\varepsilon^L}^{\{0\}}$, $P_{\varepsilon^L \varepsilon_\delta^L}^{\{0\}}$ and $B_{\varepsilon^L}^{\{0\}}$ are the Lagrangian counterparts of the stochastic amplitudes defined in Eq. (4.6); again, they are constant on large scales. Moreover, for conserved tracers such as protohalos, they are the same as their Eulerian counterparts $P_\varepsilon^{\{0\}}$, $P_{\varepsilon \varepsilon_\delta}^{\{0\}}$, $B_\varepsilon^{\{0\}}$. In that case, Eq. (4.12) and Eqs. (4.2)–(4.4) also directly map onto each other through the relations between Eulerian and Lagrangian bias parameters given in Sec. 2.3. Thus, at each order in perturbation theory (but excluding higher-derivative terms), a knowledge of one set of bias parameters (either Eulerian or Lagrangian) determines the other, and describes both late time Eulerian statistics as well as Lagrangian statistics of proto-halos in the initial conditions; this includes the stochastic terms, as mentioned above. It is also worth noticing that the only corrections to the halo-matter cross-power spectrum $P_{hm}^{L,\mathrm{LO}}(k)$ and halo-matter-matter bispectrum $B_{hmm}^{L,\mathrm{LO}}(k)$ in Lagrangian space are due to higher-derivative terms [13, 257, 258]. This makes $P_{hm}^L(k)$ and $B_{hmm}^L(k)$ a convenient tool for measuring the Lagrangian higher-derivative biases of proto-halos (see e.g. [259, 174]).

4.1.3 A worked example: bias constraints from the leading-order power spectrum and bispectrum

As we have discussed in the previous section, the dependence on scale and triangle configuration of the halo bispectrum [Eq. (4.4) and Fig. 10] provides ample information on the bias parameters that can be

extracted from the observed halo sample itself. The bispectrum thus breaks the degeneracy between linear bias b_1 and the normalization of $P_{\mathrm{L}}(k)$ that is present in the leading-order halo power spectrum. The same is expected to hold for galaxies, once projection effects are included. We now investigate this information gain quantitatively, though in an idealized setting, by using the Fisher information matrix formalism [260] and applying it to upcoming galaxy surveys. Since we neglect redshift-space distortions (RSD, Sec. 9.3.2), this is clearly not a realistic, complete forecast. We will discuss this further below.

At leading order, i.e. at second order in perturbation theory, the galaxy power spectrum and bispectrum are statistically independent. Moreover, different wavenumbers are uncorrelated; that is, the covariance matrices of power spectrum and bispectrum are diagonal. Then, the full Fisher information matrix is given by

$$F_{ij} = F_{ij}^{(P)} + F_{ij}^{(B)}, \tag{4.13}$$

where the power spectrum and bispectrum Fisher matrices are respectively given by

$$F_{ij}^{(P)} = \sum_k \frac{1}{\mathrm{Var}\left[P_{gg}(k)\right]} \frac{\partial P_{gg}(k)}{\partial \theta_i} \frac{\partial P_{gg}(k)}{\partial \theta_j} \tag{4.14}$$

$$F_{ij}^{(B)} = \sum_{(k_1,k_2,k_3)} \frac{1}{\mathrm{Var}\left[B_{ggg}(k_1,k_2,k_3)\right]} \frac{\partial B_{ggg}(k_1,k_2,k_3)}{\partial \theta_i} \frac{\partial B_{ggg}(k_1,k_2,k_3)}{\partial \theta_j} . \tag{4.15}$$

Here, the sums run over wavenumber bins specified below, and $\mathrm{Var}\left[P_{gg}(k)\right]$ and $\mathrm{Var}\left[B_{ggg}(k_1,k_2,k_3)\right]$ denote the variances of the binned power spectrum and bispectrum estimators, respectively. For this example, our parameter vector $\vec{\theta}$ contains all first- and second-order bias parameters as well as stochastic amplitudes, in addition to the amplitude of the primordial power spectrum $\mathcal{A}_s$ (Tab. 1):

$$\vec{\theta} = \left\{ b_1,\, b_2,\, b_{K^2},\, \ln \mathcal{A}_s,\, P_\varepsilon^{\{0\}},\, P_{\varepsilon\varepsilon_\delta}^{\{0\}},\, B_\varepsilon^{\{0\}} \right\} . \tag{4.16}$$

For the biases b_O, we assume fiducial values as indicated in Tab. 6, while for the stochastic amplitudes we assume as fiducial the values predicted by Poisson statistics, $P_\varepsilon^{\{0\}} = (\bar{n}_g)^{-1}$, $P_{\varepsilon\varepsilon_\delta}^{\{0\}} = b_1/(2\bar{n}_g)$, and $B_\varepsilon^{\{0\}} = (\bar{n}_g)^{-2}$, as derived in Sec. 4.5.3.

Let us briefly discuss the effect of higher-order nonlinear corrections and RSD. These can lead to a degradation of the idealized bias constraints, since both effects introduce additional free parameters. On the other hand, RSD lead to anisotropies in the clustering statistics that can break parameter degeneracies. Furthermore, including nonlinear corrections allows us to include smaller scales, and hence many additional modes, in the analysis. These effects in turn can lead to an improvement over the idealized constraints presented here.

The leading-order variance of the binned galaxy power spectrum estimator $\hat{P}_{gg}(k)$ is given by [261]

$$\mathrm{Var}\left[P_{gg}(k)\right] = \frac{1}{N_k} \left[P_{gg}(k)\right]^2 , \tag{4.17}$$

where N_k is the number of independent Fourier modes in the wavenumber bin used to estimate the power spectrum $P_{gg}(k)$. Note that in our notation, $P_{gg}(k)$ includes the stochastic (shot-noise) contribution [Eq. (4.2)]. We do not take into account the effect of survey geometry and assume that N_k is the same as that of a cubic box with the survey volume V_{survey}. In that case, N_k is given by

$$N_k = \frac{1}{2} \frac{4\pi k^2 \Delta k}{(k_F)^3} = \frac{V_{\mathrm{survey}}}{4\pi^2} k^2 \Delta k , \tag{4.18}$$

where $k_F \equiv 2\pi/V_{\mathrm{survey}}^{1/3}$ is the fundamental wavenumber of the survey. Note that the constraint that $\delta_g(\boldsymbol{x})$ is a real field reduces the number of independent Fourier modes so that only one half of the total Fourier volume is counted in the number N_k.

Table 6: Projected uncertainties on the deterministic bias parameters, b_1, b_2, b_{K^2}, and the amplitude of the linear power spectrum, $\ln \mathcal{A}_s$, from current and upcoming galaxy surveys with listed specifications in an idealized setting. In all cases, we assume $b_1 = 1.5$, and calculate b_2 and b_{K^2} from the relation given in Tab. 7 ($b_2 \simeq -0.69$), and Lagrangian LIMD ($b_{K^2} \simeq -0.14$), respectively. The fiducial values of the stochastic amplitudes are given by Poisson sampling, $P_\varepsilon^{\{0\}} = 1/\overline{n}_g$, $P_{\varepsilon\varepsilon_\delta}^{\{0\}} = b_1/(2\overline{n}_g)$, $B_\varepsilon^{\{0\}} = 1/\overline{n}_g^2$.

survey	redshift $\bar{z}$	V_{survey} [$h^{-3}\mathrm{Gpc}^3$]	$10^4\,\overline{n}_g$ [$h^{-3}\mathrm{Mpc}^3$]	$k_{\max} = 0.1\,h/\mathrm{Mpc}$				$k_{\max} = 0.2\,h/\mathrm{Mpc}$			
				$100\,\sigma(b_O)$			$\sigma_{\ln \mathcal{A}_s}$	$100\,\sigma(b_O)$			$\sigma_{\ln \mathcal{A}_s}$
				b_1	b_2	b_{K^2}		b_1	b_2	b_{K^2}	
eBOSS (LRG)	0.8	6.1	4.4	32	45	30	0.43	7.0	4.5	5.9	0.093
eBOSS (QSO)	1.4	39	1.5	38	51	36	0.51	11	6.5	9.2	0.15
HETDEX	2.7	2.7	3.6	190	260	180	2.6	59	35	49	0.79
PFS	1.5	8.7	4.6	47	66	44	0.62	11	6.7	8.9	0.14
DESI	1.1	40	3.3	18	25	17	0.25	4.4	2.7	3.7	0.059
WFIRST	1.9	13	12	35	49	32	0.46	6.8	4.4	5.6	0.091
Euclid	1.4	63	5.2	15	20	14	0.20	3.3	2.1	2.7	0.044

The leading-order variance of the estimated galaxy bispectrum $\hat{B}_{ggg}(k_1, k_2, k_3)$ is given by [61, 62, 250]

$$\mathrm{Var}\left[B_{ggg}(k_1, k_2, k_3)\right] = s_B \frac{V_{\mathrm{survey}}}{N_t} P_{gg}(k_1) P_{gg}(k_2) P_{gg}(k_3) , \tag{4.19}$$

where s_B is a symmetry factor (6 for equilateral triangles, 2 for isosceles triangles, 1 for other triangles), and N_t is the number of triangle configurations in the bin considered; for simplicity, we use the expression in the continuum limit, valid if N_t is large:

$$N_t \simeq \frac{1}{k_F^6} \int_{|\boldsymbol{k}_1 - \boldsymbol{p}_1| < \Delta k_1/2} d^3 p_1 \int_{|\boldsymbol{k}_2 - \boldsymbol{p}_2| < \Delta k_2/2} d^3 p_2 \int_{|\boldsymbol{k}_3 - \boldsymbol{p}_3| < \Delta k_3/2} d^3 p_3\, \delta_D(\boldsymbol{p}_1 + \boldsymbol{p}_2 + \boldsymbol{p}_3)$$

$$= \left(\prod_{i=1}^3 \frac{k_i \Delta k_i}{k_F^2}\right) \times \begin{cases} 4\pi^2 , & k_i = k_j + k_k \\ 8\pi^2 , & \text{otherwise} \end{cases} , \tag{4.20}$$

where $i, j, k \in \mathrm{cyclic}(1, 2, 3)$. For this calculation, we use $\Delta k_i = k_F$ as bin width. We consider the following six galaxy surveys: HETDEX [50], eBOSS [49], DESI [53], PFS [57], Euclid [54] and WFIRST [58]. Their mean redshift $\bar{z}$ as well as volume and mean galaxy density assumed for this forecast are given in Tab. 6.

The results of the projected 1-sigma (68% confidence level) uncertainties on the bias parameters b_1, b_2, b_{K^2}, and the amplitude of the linear power spectrum, $\ln \mathcal{A}_s$, after marginalizing over the stochastic parameters $P_\varepsilon^{\{0\}}$, $P_{\varepsilon\varepsilon_\delta}^{\{0\}}$ and $B_\varepsilon^{\{0\}}$, are given in Tab. 6. First of all, as we have noted earlier, the bispectrum breaks the notorious degeneracy between the two quantities b_1 and $\ln \mathcal{A}_s$, and we can measure both parameters simultaneously by combining the power spectrum and bispectrum. Consequently, future galaxy surveys can constrain b_1, b_2 and b_{K^2} at the level of tens of percent when including Fourier modes with $k < k_{\max} = 0.1\,h/\mathrm{Mpc}$. However, we stress that the constraints depend sensitively on the range of wavenumbers included: going from $k_{\max} = 0.1\,h\,\mathrm{Mpc}^{-1}$ to $0.2\,h\,\mathrm{Mpc}^{-1}$ improves constraints by a factor of several, and the constraints are now at the few-percent level. This highlights the importance of an accurate theoretical model of nonlinear corrections to P_{gg}, B_{ggg}, as well as their covariance matrices. The precise value of $k_{\max}$ that leads to optimal unbiased constraints on cosmological parameters and the b_O is highly dependent on the survey considered, in particular its redshift range and the galaxy sample being targeted, and we refrain from attempting to estimate survey-specific values for $k_{\max}$ here. The constraints quoted in Tab. 6 should thus only be seen as illustrative figures.

4.1.4 Next-to-leading-order corrections to the two-point functions

So far, we have derived the leading contributions to the two- and three-point functions of galaxies and halos on large scales. In order to illustrate how higher-order corrections to the above results can be derived, we also present the next-to-leading-order (NLO, or 1-loop) correction to the two-point function. We only discuss real-space predictions without any projection effects. Hence our expressions mostly apply to halos, but the structure of the perturbative expansion remains the same even when including projection effects. Deriving the nonlinear correction to the two-point functions requires a third-order calculation in perturbation theory, since, for any operator O, contributions of the type $\langle O^{(1)}O'^{(3)} \rangle$ contribute at the same order as $\langle O^{(2)}O'^{(2)} \rangle$ [262, 128, 197]. Thus, the bias expansion in Eq. (4.1) is necessary and sufficient to derive this correction. Writing

$$
\begin{aligned}
P_{hm}(k) &= P_{hm}^{\mathrm{LO}}(k) + P_{hm}^{\mathrm{NLO}}(k) + \cdots \\
P_{hh}(k) &= P_{hh}^{\mathrm{LO}}(k) + P_{hh}^{\mathrm{NLO}}(k) + \cdots ,
\end{aligned}
\tag{4.21}
$$

and following the notation of [197], the NLO contributions to the halo-matter and halo-halo power spectrum are respectively given by[13]

$$
\begin{aligned}
P_{hm}^{\mathrm{NLO}}(k) &= b_1 \left[P_{mm}^{\mathrm{NLO}}(k) - 2C_{s,\mathrm{eff}}^2 k^2 P_{\mathrm{L}}(k) \right] + \hat{P}_{hm}^{\mathrm{NLO}}(k) \\
\hat{P}_{hm}^{\mathrm{NLO}}(k) &\equiv b_{\delta^2} \mathcal{I}^{[\delta^{(2)},\delta^2]}(k) + b_{K^2} \mathcal{I}^{[\delta^{(2)},K^2]}(k) + \left(b_{K^2} + \frac{2}{5} b_{\mathrm{td}} \right) f_{\mathrm{NLO}}(k) P_{\mathrm{L}}(k) \\
&\quad - b_{\nabla^2\delta} k^2 P_{\mathrm{L}}(k) + k^2 P_{\varepsilon\varepsilon_m}^{\{2\}} \\
P_{hh}^{\mathrm{NLO}}(k) &= (b_1)^2 \left[P_{mm}^{\mathrm{NLO}}(k) - 2C_{s,\mathrm{eff}}^2 k^2 P_{\mathrm{L}}(k) \right] + 2b_1 \hat{P}_{hm}^{\mathrm{NLO}}(k) + \sum_{O,O' \in \{\delta^2,K^2\}} b_O b_{O'} \mathcal{I}^{[O,O']}(k) + k^2 P_{\varepsilon}^{\{2\}} ,
\end{aligned}
\tag{4.22}
$$

where $P_{mm}^{\mathrm{NLO}}(k)$ is the NLO correction to the matter power spectrum [Eq. (B.18)], $C_{s,\mathrm{eff}}^2 \equiv (2\pi)c_{s,\mathrm{eff}}^2/k_{\mathrm{NL}}^2$ is the scaled effective sound speed of the *matter* fluid [83, 84] [cf. Eq. (B.28)] and k_{NL} is defined in Eq. (4.25), while ε_m is the effective stochastic contribution to the matter density (Appendix B.3). Further,

$$
f_{\mathrm{NLO}}(k) = 4 \int_{\boldsymbol{p}} \left[\frac{[\boldsymbol{p} \cdot (\boldsymbol{k} - \boldsymbol{p})]^2}{p^2 |\boldsymbol{k} - \boldsymbol{p}|^2} - 1 \right] F_2(\boldsymbol{k}, -\boldsymbol{p}) P_{\mathrm{L}}(p)
$$

$$
\mathcal{I}^{[O,O']}(k) = 2 \left[\int_{\boldsymbol{p}} S_O(\boldsymbol{k} - \boldsymbol{p}, \boldsymbol{p}) S_{O'}(\boldsymbol{k} - \boldsymbol{p}, \boldsymbol{p}) P_{\mathrm{L}}(p) P_{\mathrm{L}}(|\boldsymbol{k} - \boldsymbol{p}|) \right.
$$

$$
\left. - \int_{\boldsymbol{p}} S_O(-\boldsymbol{p}, \boldsymbol{p}) S_{O'}(-\boldsymbol{p}, \boldsymbol{p}) P_{\mathrm{L}}(p) P_{\mathrm{L}}(p) \right] ,
\tag{4.23}
$$

$$
\text{where} \quad S_O(\boldsymbol{k}_1, \boldsymbol{k}_2) =
\begin{cases}
F_2(\boldsymbol{k}_1, \boldsymbol{k}_2), & O = \delta^{(2)} \\
1, & O = \delta^2 \\
(\hat{\boldsymbol{k}}_1 \cdot \hat{\boldsymbol{k}}_2)^2 - 1/3, & O = K^2
\end{cases}.
\tag{4.24}
$$

We see that, in addition to[14] $b_{\delta^2} = b_2/2$ and b_{K^2}, which also enter the leading-order halo three-point function, one new local bias term appears in the NLO halo power spectra, namely $O_{\mathrm{td}}^{(3)}$ defined in Eq. (2.50), with associated bias parameter b_{td}. Thus, only one of four cubic-order local bias parameters contributes to the next-to-leading-order halo power spectra. We have also included the leading higher-derivative bias $b_{\nabla^2\delta} \propto R_*^2$ in Eq. (4.22), where R_* is the nonlocality scale of the halos or galaxies considered.

Correspondingly, we have also included the scale-dependent stochastic contributions to P_{hm}^{NLO} and $P_{hh}^{\mathrm{NLO}}(k)$. The latter is expanded following Eq. (2.88), and is expected to scale as $|P_{\varepsilon}^{\{2\}}| \sim R_*^2 P_{\varepsilon}^{\{0\}}$ [127, 177, 178, 179].

---

[13] We include the relevant stochastic terms which were not considered in [197].

[14] We use the bias parameter b_{δ^2} here to make the notation in Eqs. (4.22)–(4.24) more compact.
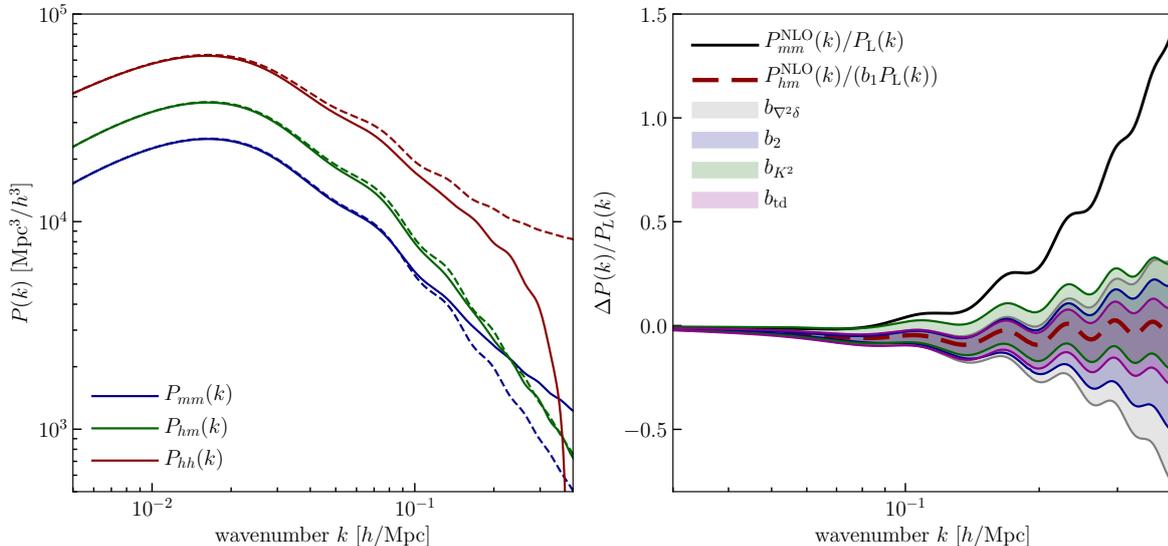
Figure 12: *Left panel:* illustration of halo auto- (red, top line) and cross-power spectra (green, middle line), and the matter power spectrum (blue, bottom line) at $z = 0$. The solid lines show the total LO plus NLO result, while the dashed curves show the LO (linear) prediction only. The bias parameters used here are $b_1 = 1.50$, $b_2 = -0.69$, and $b_{K^2} = -0.14$, as in Tab. 6, while $b_{\nabla^2\delta} = R_*^2$ with $R_* = 2.61\,h^{-1}\,\mathrm{Mpc}$. $b_{\mathrm{td}} = 23/42(b_1 - 1)$ is taken from the Lagrangian LIMD prediction (Sec. 2.4). The stochastic amplitudes are taken from the Poisson expectation, $P_\varepsilon^{\{0\}} = 1/\overline{n}_h$ and $P_\varepsilon^{\{2\}} = -R_*^2/\overline{n}_h$, with $\overline{n}_h = 1.41 \cdot 10^{-4}(h^{-1}\,\mathrm{Mpc})^{-3}$. We have set $P_{\varepsilon\varepsilon_m}^{\{2\}} = 0$ in $P_{hm}^{\mathrm{NLO}}(k)$. *Right panel:* fractional size of the NLO contributions to the matter and halo-matter cross-power spectrum at $z = 0$. The red dashed line shows the result for $P_{hm}(k)$ for the fiducial bias parameters given above. The different shaded areas around P_{hm}^{NLO} show the effect of rescaling the various bias parameters by a factor in the range $[0.5, 2]$. Clearly, the contributions from different bias parameters exhibit similar dependencies on k, and are in general difficult to disentangle using only the power spectrum. The perturbative description is expected to fail for $k \gtrsim 0.25\,h\,\mathrm{Mpc}^{-1}$, where $P_{mm}^{\mathrm{NLO}}(k)$ becomes as large as the LO prediction $P_{\mathrm{L}}(k)$.

We will return to this in Sec. 4.5.3. It is often assumed that there is no stochastic contribution to the halo-matter cross-power spectrum. However, this is only true at lowest order. The nonlinear small-scale modes of the density field are responsible for both the halo stochasticity ε and the stochastic contribution to the matter density field ε_m, which, as discussed in Appendix B.3, is due to the effective pressure of the nonlinear matter fluctuations and scales as k^2 in the low-k limit. Hence, one has to allow for a correlation between the two stochastic fields, leading to the term $k^2 P_{\varepsilon\varepsilon_m}^{\{2\}}$ in P_{hm}^{NLO}, which is comparable to the other NLO contributions. Note that it could be either positive or negative.

The magnitude and scale dependence of the NLO corrections to the halo and matter power spectra is shown in Fig. 12. As expected, we see that the corrections become increasingly important towards smaller scales (higher k). We see a particularly steep suppression of $P_{hh}(k)$, which, for our fiducial parameters, is dominated by the higher-derivative stochastic contribution $k^2 P_\varepsilon^{\{2\}}$. The right panel of Fig. 12 shows the fractional size of the NLO correction to $P_{mm}(k)$ and $P_{hm}(k)$. Depending on the value of the various bias and stochastic parameters, the NLO correction could be either positive or negative (shaded regions), and cancellations between the different NLO contributions can occur. In any case, as soon as the fractional size of the NLO correction approaches order unity, we expect that higher-order loop contributions which we have not included become comparable to $P_{hm}^{\mathrm{NLO}}(k)$ as well, and hence the perturbative expansion ceases to converge.

The NLO halo-matter power spectrum adds five additional free parameters to the ones present at leading order (b_1, $P_\varepsilon^{\{0\}}$). These can, in principle, be disentangled due to the different scale dependence of each term. However, as illustrated in Fig. 12, these scale dependences are sufficiently similar that it is difficult to

disentangle the various higher-order bias parameters in practice; note that there is only a limited range in wavenumbers that can be used for the parameter estimation, due to the presence of higher loop and derivative corrections (see below). Nevertheless, the leading-order bispectrum can be used to determine b_2 and b_{K^2}, leaving only $b_{\rm td}$, $b_{\nabla^2\delta}$, and $P_{\varepsilon\varepsilon_m}^{\{2\}}$ to be constrained from the NLO correction to the halo-matter cross-power spectrum.

In order to gain a more detailed understanding of the magnitude of the corrections in Eq. (4.22), let us approximate the matter power spectrum by a power law,

$$P_{\rm L}(k) \approx \frac{2\pi^2}{k_{\rm NL}^3} \left(\frac{k}{k_{\rm NL}} \right)^n , \qquad (4.25)$$

where $k_{\rm NL}$ is the *nonlinear scale* at which the dimensionless matter power spectrum $\Delta^2(k) = k^3 P_{\rm L}(k)/(2\pi^2)$ becomes unity. This yields, for example,

$$\frac{\mathcal{I}^{[\delta^2,\delta^2]}}{P_{\rm L}(k)} = 2 \left(\frac{k}{k_{\rm NL}} \right)^{3+n} \int_{-1}^1 \frac{d\mu}{2} \int_0^\infty x^2 dx \left[\left(x\sqrt{1 + x^2 - 2x\mu} \right)^n - x^{2n} \right] . \qquad (4.26)$$

While other NLO loop-integral terms have different angular integrands, the scaling $\propto (k/k_{\rm NL})^{3+n}$ is common to all (see also Fig. 12). Note that, depending on the value of n, the integral over x might need to be regularized in the UV (ultraviolet, small-scale, or large-x, limit of the integral), while the integral is safe from divergence in the IR (infrared, large-scale, or small-x, limit of the integral), because of the term subtracted in Eq. (4.23); in any case, this does not affect the scaling with $k/k_{\rm NL}$. This scaling allows us to estimate the importance of higher-order terms. For example, 2-loop corrections correspondingly scale as $(k/k_{\rm NL})^{2(3+n)}$ for a scale-free power spectrum [218]. For our reference ΛCDM cosmology, we have approximately[15] $k_{\rm NL}(z = 0) = 0.25 \, h \, {\rm Mpc}^{-1}$ and $n = d\ln P_{\rm L}/d\ln k|_{k_{\rm NL}} = -1.7$, so that the one-loop terms scale approximately as $(k/k_{\rm NL})^{1.3}$. Of course, this is only a rough approximation as $P_{\rm L}(k)$ cannot be approximated as a power law over the entire relevant range of scales. In particular, since n becomes positive for $k \lesssim 0.02 \, h \, {\rm Mpc}^{-1}$, the NLO terms eventually scale as k^2 for sufficiently small values of k. Nevertheless, such estimates are important as they allow us to marginalize over higher-loop corrections and rigorously take into account the uncertainty in the prediction of Eq. (4.22) [263].

The higher-derivative term $\propto b_{\nabla^2\delta}$ obeys a scaling with k ($\propto k^2$) that is in general different from that of the NLO corrections ($\propto k^{3+n}$). Further, the former involves an additional scale, R_*. Thus, we have *two independent expansion parameters*,

$$\epsilon_{\rm loop} \equiv \left(\frac{k}{k_{\rm NL}} \right)^{3+n} \approx \left(\frac{k}{0.25 \, h \, {\rm Mpc}^{-1}} \right)^{1.3} , \quad \text{and} \quad \epsilon_{\rm deriv.} \equiv k^2 R_*^2 . \qquad (4.27)$$

Thus, depending on the halo or galaxy sample, the leading higher-derivative term could be negligible compared to the NLO corrections on the scales of interest, e.g. $0.01 \lesssim k \, [h \, {\rm Mpc}^{-1}] \lesssim 0.2$, or could be significantly larger. If $\epsilon_{\rm deriv.}$ is comparable to $\epsilon_{\rm loop}$ on the scales considered, then both NLO and leading higher-derivative corrections should be included. This is what we have assumed in Eq. (4.22). More generally, when going to higher orders, one would then include terms that involve the same powers of $\epsilon_{\rm loop}$ and $\epsilon_{\rm deriv.}$. For example, at 2-loop order, these are the terms of order $\epsilon_{\rm loop}^2$, $\epsilon_{\rm loop}\epsilon_{\rm deriv.}$, and $\epsilon_{\rm deriv.}^2$. On the other hand, if the two expansion parameters are substantially different, then it is necessary to retain terms that are higher order in the larger parameter. For example, if $\epsilon_{\rm deriv.} \gg \epsilon_{\rm loop}$, one should allow for additional higher-derivative terms, which leads to contributions $\propto \{k^4 R_*^4, \, k^6 R_*^6, \cdots\} P_{\rm L}(k)$ in Eq. (4.22) [128, 264, 265]. The cutoff of the perturbative approach then is at $k \approx 1/R_*$. All of this applies analogously to the bispectrum and higher n-point functions.

Finally, the higher-derivative stochastic contributions, which scale as k^2 (as opposed to $k^2 P_{\rm L}(k)$ as the higher-derivative bias contribution), are higher order in terms of their k scaling, but the amplitude

---

[15]This was obtained by fitting a power law to $P_{\rm L}(k)$ over the range $k \in [0.1, 0.25] \, h \, {\rm Mpc}^{-1}$.

could be large for very rare high-mass halos. Note that, in the definition of $\mathcal{I}^{[O,O']}$ [Eq. (4.23)], we have subtracted any possible constant term in the $k \to 0$ limit, since such a constant term is absorbed by the renormalized stochastic term $P_\varepsilon^{\{0\}}$; this corresponds to the term $P_0 \equiv b_2^2 \int_{\boldsymbol{q}} |P_{\mathrm{L}}(q)|^2$ in [266], and evaluates to a finite but significant value for a ΛCDM power spectrum. In principle, one should also subtract the term $k^2 \partial^2 \mathcal{I}^{[O,O']}/\partial k^2|_{k=0}$ (and corresponding higher-order terms), since it is absorbed by the higher-derivative stochastic term. However, as long as the amplitude of $P_\varepsilon^{\{2\}}$ is allowed to be free, it can absorb the contribution from the local bias loop integrals (note that new terms with the same scaling appear at each higher loop order).

The prediction for P_{hm}^{NLO}, P_{hh}^{NLO} in Eq. (4.22) is unambiguous; however, the various contributions can be broken down in a variety of different ways [128, 137, 267], of which Eq. (4.22) is only one option. Also, when using Eq. (4.22) in conjunction with Eq. (4.2) to fit the data, one should, strictly speaking, include the leading connected 4-point function (trispectrum) in the covariance of P_{hh} and P_{hm}, as this is also a third-order contribution in perturbation theory. In practice, this leads to fairly small corrections to the uncertainties of inferred bias parameters [268].

Finally, the standard NLO result given in Eq. (4.22) does not provide a very good description of the power spectrum and correlation function around the scale of the BAO feature ($0.05 \, h \, \mathrm{Mpc}^{-1} \lesssim k \lesssim 0.2 \, h \, \mathrm{Mpc}^{-1}$). This is because large-scale modes introduce displacements which lead to a significant smoothing of the BAO feature. This deficiency can be improved by resumming these displacement terms [269, 270, 271, 272]. We briefly describe this resummation, which applies in the same way to biased tracers as to matter, in Appendix B.4.

Previous derivations have obtained a subset of the terms in Eq. (4.22); for example, [79] derived the analogous result in LPT for LIMD Lagrangian bias, where b_{K^2} and b_{td} are set to zero at the initial time (see Sec. 2.4). Ref. [266, 273] assumed LIMD Eulerian bias, setting b_{K^2} and b_{td} to zero at the final time. Ref. [255] used renormalized bias parameters, but also set $b_{K^2} = 0 = b_{\mathrm{td}}$, and replaced $b_1(P_{\mathrm{L}} + P_{mm}^{\mathrm{NLO}})$ with b_1 times the matter power spectrum given by the closure theory prescription [274], finding fairly good agreement with simulation measurements of $P_{hm}(k)$ up to $k = 0.3 \, h \, \mathrm{Mpc}^{-1}$ at $z = 0$. Note however that this does not prove that b_{K^2} and b_{td} vanish; as emphasized above, using the correct values for all bias parameters is essential in order to obtain a consistent prediction for all halo statistics (including auto- and cross-, two- and three-point functions). Ref. [141] derived the one-loop power spectrum in LPT with general Lagrangian bias functions $c_N^L(\boldsymbol{k}_1, \cdots, \boldsymbol{k}_N)$. This approach is somewhat different from the renormalized, perturbative bias expansion discussed in Sec. 2 which leads to Eq. (4.22), in that one needs a "microscopic" model for the free functions $c_N^L(\boldsymbol{k}_1, \cdots, \boldsymbol{k}_N)$. The example of inserting a bias parameter multiplied by N powers of a filtering kernel $W_R(k)$ corresponds to introducing a finite physical cutoff scale $\Lambda \sim R^{-1}$ in the loop integrals. This is again in contrast to Eq. (4.22), which is explicitly independent of any cutoff (see Sec. 6.9.3 for a more detailed comparison). More generally, it is worth noting that predictions from Eulerian SPT and Lagrangian LPT, while they agree by definition on large scales, begin to diverge when extrapolated beyond their regime of validity.

4.2 Moments

Historically, the first studies of bias considered the moments of the one-point distribution, i.e. number counts, of biased tracers within volumes of a fixed size R_ℓ, also referred to as counts-in-cells. Specifically, the moments are defined as $\langle [\delta_{g,\ell}(\boldsymbol{x})]^n [\delta_\ell(\boldsymbol{x})]^m \rangle$, where $n, m = 0, 1, \cdots$ and $\delta_{g,\ell}$, δ_ℓ denote the galaxy and matter density fields, respectively, both smoothed on the scale R_ℓ. These moments have been estimated on low-redshift galaxy catalogs as early as the 1970s [5, 275, 276, 277, 278, 279, 280, 281]. The moments can be conveniently rescaled to define the *hierarchical amplitudes* in terms of the skewness, kurtosis, and higher reduced moments of the filtered galaxy density field $\delta_{g,\ell}$:

$$\frac{\langle \delta_{g,\ell}^3 \rangle_c}{\langle \delta_{g,\ell}^2 \rangle^2} \, , \qquad \frac{\langle \delta_{g,\ell}^4 \rangle_c}{\langle \delta_{g,\ell}^2 \rangle^3} \, , \tag{4.28}$$

and so on, where a subscript c denotes the connected part of the moment. The skewness captures the asymmetry in the volume fraction of underdense and overdense regions of the galaxy density field. The

original motivation for the definitions in Eq. (4.28) was the fact that, in "hierarchical models" of clustering, the amplitudes $\langle \delta_{g,\ell}^N \rangle_c / \langle \delta_{g,\ell}^2 \rangle^{N-1}$ are independent of time and scale R_ℓ [6, 282, 283], as confirmed by gravity-only simulations with power-law initial conditions [284, 285, 286, 287, 288]. For realistic CDM power spectra, however, there are deviations from the hierarchical clustering prediction, which result in significant R_ℓ-dependent corrections to $\langle \delta_{g,\ell}^N \rangle_c / \langle \delta_{g,\ell}^2 \rangle^{N-1}$ [289, 286, 290]. A generalization of the moments in Eq. (4.28), the two-point moments [102, 291] employs the joint expectation value of powers of $\delta_{g,\ell}$, δ_ℓ at two different locations. Ref. [292, 293] considered rare excursions of the smoothed density field, which are related to moments and amenable to analytical statistical techniques.

For a LIMD bias expansion and assuming hierarchical clustering, the second and third moments can be combined to yield estimators of b_1 and b_2 [294]. Note that, since both numerator and denominator have some non-vanishing noise, care must be taken in constructing unbiased estimators of the hierarchical amplitudes [295]. Early applications of this method to simulations were presented in [296, 110, 297, 298]. More recent analyses can be found in [299]. Ref. [300] present a method to use the moments of weak lensing convergence and projected galaxy density to estimate the linear bias, which is the analog in moments to the galaxy-galaxy lensing angular power spectrum Eq. (4.3).

The advantage of moments-based methods for measuring bias is that they are simple to implement on galaxy catalogs or simulation outputs. Namely, we merely have to throw spheres randomly in the simulated volume, or divide the simulation volume into a grid, and compute fractional densities. The disadvantages are that large smoothing scales need to be chosen to ensure that nonlinear corrections are under control ($\gtrsim 30\,h^{-1}\,\mathrm{Mpc}$ [301, 254], see also below), and that more complicated moments than those in Eq. (4.28) need to be included in the analysis to disentangle the different bias parameters at a given order, for example b_2 and b_{K^2}. The tidal bias b_{K^2} and stochastic contributions such as ε_δ, which appear in the halo bispectrum Eq. (4.4) and, thus, contribute to the skewness, were neglected in the references listed above. Finally, application to observational data sets is further complicated by the usually non-trivial survey geometry and mask.

In the following, we again neglect redshift-space distortions and other projection effects to discuss moments of the smoothed halo density field $\delta_{h,\ell}$ as extracted from N-body simulations. We will derive the leading-order prediction for the moments in the general perturbative bias expansion, and show how the bias parameters estimated using this method are related to those obtained from halo n-point functions on large scales. For this, we introduce a slightly different bias expansion to Eq. (4.1) by writing

$$\delta_{h,\ell}(\boldsymbol{x}) = b_1^{\mathrm{m}}(R_\ell)\delta_\ell(\boldsymbol{x}) + \frac{1}{2}b_2^{\mathrm{m}}(R_\ell)\left([\delta_\ell(\boldsymbol{x})]^2 - \sigma^2(R_\ell)\right) + b_{K^2}^{\mathrm{m}}(R_\ell)\left(K_\ell^2(\boldsymbol{x}) - \langle K_\ell^2 \rangle\right)$$
$$+ \varepsilon_\ell^{\mathrm{m}}(\boldsymbol{x}) + \varepsilon_{\delta,\ell}^{\mathrm{m}}(\boldsymbol{x})\delta_\ell(\boldsymbol{x}) + \cdots, \tag{4.29}$$

where O_ℓ denotes an operator O smoothed on the scale R_ℓ, and $K_\ell^2 \equiv K_{ij,\ell}K_\ell^{ij}$. Note the difference to the expression obtained if one were to smooth Eq. (4.1): there, one would obtain $[\delta^2]_\ell$ while here we have $[\delta_\ell]^2$. Correspondingly, we have introduced a different set of bias parameters here, the *moments biases* $b_O^{\mathrm{m}}(R_\ell)$ which depend on the scale R_ℓ. The relation between the b_O^{m}, the moments, and the b_O of Eq. (4.1) will become clear momentarily.

We begin with the variance:

$$\sigma_{hm}^2(R_\ell) \equiv \langle \delta_\ell(\boldsymbol{x})\delta_{h,\ell}(\boldsymbol{x}) \rangle \qquad \overset{\mathrm{LO}}{=} b_1^{\mathrm{m}}(R_\ell)\sigma^2(R_\ell) \qquad \overset{\mathrm{LO}}{=} b_1\sigma^2(R_\ell)$$

$$\sigma_{hh}^2(R_\ell) \equiv \langle \delta_{h,\ell}(\boldsymbol{x})\delta_{h,\ell}(\boldsymbol{x}) \rangle \qquad \overset{\mathrm{LO}}{=} [b_1^{\mathrm{m}}(R_\ell)]^2\sigma^2(R_\ell) + \langle (\varepsilon_\ell^{\mathrm{m}})^2 \rangle \qquad \overset{\mathrm{LO}}{=} b_1^2\sigma^2(R_\ell) + \langle \varepsilon_\ell^2 \rangle, \tag{4.30}$$

where, for each line, the first relation in the LO limit is that obtained using the moments-bias expansion Eq. (4.29), while the second is what is obtained in the large-scale bias expansion Eq. (4.1). Further,

$$\langle (\varepsilon_\ell^{\mathrm{m}})^2 \rangle = \langle \varepsilon_\ell^2 \rangle = P_\varepsilon^{\{0\}} V_\ell^{-1}, \tag{4.31}$$

where $V_\ell^{-1} = [\int d^3\boldsymbol{x}\, W_\ell^2(x)]$ is the inverse of the volume of the filtering kernel $W_\ell \equiv W_{R_\ell}$. As emphasized in Eq. (4.30), these results hold only for sufficiently large smoothing scales so that nonlinear and higher-derivative corrections become unimportant; specifically, these scale as $\sigma^4(R_\ell)$ and $R_*^2\sigma_1^2(R_\ell)$, respectively, where $\sigma_n(R_\ell)$ is defined in Tab. 4.

Eq. (4.30) shows that we have $b_1^m(R_\ell) = b_1$ at this order, and that the variance of the stochasticity agrees as well. Note that these relations also receives corrections of order $R_*^2 \sigma_1^2(R_\ell)$, and $R_*^2 \int_{\mathbf{k}} k^2 W_\ell^2(k)$, respectively. These corrections not only depend on the filter scale R_ℓ, but the shape of the filter as well, in particular its behavior at high k.

We now consider third-order moments. Specifically, we focus on halo-matter-matter moments as these do not receive stochastic contributions at leading order and are thus convenient to single out the deterministic bias parameters; more generally, this holds for all moments involving only one power of the halo field. We will see that the differences between the (renormalized) bias parameters appearing in the n-point functions of Sec. 4.1 and those defined through Eq. (4.29) now become apparent. Consider the halo-matter-matter moment, which via Eq. (4.29) becomes at leading order

$$\langle \delta_\ell^2(\boldsymbol{x}) \delta_{h,\ell}(\boldsymbol{x}) \rangle \stackrel{\text{LO}}{=} b_1^m(R_\ell) \langle \delta_\ell^3 \rangle + \frac{1}{2} b_2^m(R_\ell) \Big\langle \delta_\ell^2 \left[\delta_\ell^2 - \sigma^2(R_\ell) \right] \Big\rangle + b_{K^2}^m(R_\ell) \Big\langle \delta_\ell^2 \left[K_\ell^2 - \langle K_\ell^2 \rangle \right] \Big\rangle$$
$$= b_1^m(R_\ell) \langle \delta_\ell^3 \rangle + b_2^m(R_\ell) \sigma^4(R_\ell) \,. \tag{4.32}$$

Note that in our choice of bias expansion Eq. (4.29), $b_{K^2}^m(R_\ell)$ does not contribute to $\langle \delta_\ell^2 \delta_{h,\ell} \rangle$. This changes for example if one replaces the operator K^2 with $\mathcal{G}_2 \equiv K^2 - (2/3)\delta^2$ [254]. In order to measure both b_2^m and $b_{K^2}^m$, we also need to measure an independent third moment, in particular one that includes the smoothed tidal field [118, 245],

$$\langle K_\ell^2(\boldsymbol{x}) \delta_{h,\ell}(\boldsymbol{x}) \rangle \stackrel{\text{LO}}{=} b_1^m(R_\ell) \langle K_\ell^2 \delta_\ell \rangle + \frac{1}{2} b_2^m(R_\ell) \Big\langle \left[\delta_\ell^2 - \sigma^2(R_\ell) \right] K_\ell^2 \Big\rangle + b_{K^2}^m(R_\ell) \Big\langle \left[K_\ell^2 - \langle K_\ell^2 \rangle \right] K_\ell^2 \Big\rangle$$
$$= b_1^m(R_\ell) \langle K_\ell^2 \delta_\ell \rangle + \frac{8}{45} b_{K^2}^m(R_\ell) \sigma^4(R_\ell) \,, \tag{4.33}$$

which now yields an estimate of $b_{K^2}^m(R_\ell)$. Here, we have used that $\langle K_\ell^2 \rangle = (2/3)\sigma^2(R_\ell)$. Similarly to the n-point functions, where higher-order terms in perturbation theory lead to corrections of order $(k/k_{\text{NL}})^{3+n}$, higher-order corrections to the moments enter as higher powers of $\sigma^2(R_\ell)$, in particular $\sigma^6(R_\ell)$ in the case of Eqs. (4.32)–(4.33). We will now show that there is a well-defined procedure for mapping the $b_2^m(R_\ell)$, $b_{K^2}^m(R_\ell)$ to parameters b_2, b_{K^2} measured from n-point functions in the large-scale limit; however, they are *not* equal, a fact which has been ignored in nearly all of the literature on moments of biased tracers (see e.g. [226, 299, 254], with the notable exception of [245]). The precise relation depends on the matter power spectrum shape as well as the scale R_ℓ and type of smoothing filter used.

Let us first briefly make a technical note. The moments of biased tracers on some physical scale R_ℓ are an observable just like the n-point functions discussed in Sec. 4.1. Thus, the moments biases appearing in Eqs. (4.32)–(4.33) are also physical renormalized bias parameters, but defined using different *renormalization conditions* than those employed for the biases b_O defined through n-point correlation functions, as described in Sec. 2.10.4. That is, Eq. (4.29) should strictly be written in terms of renormalized operators $[\delta^2]^m(R_\ell)$, $[K^2]^m(R_\ell)$, and so on, which are defined through the renormalization conditions given in Eq. (2.130), rather than Eq. (2.126) which defines the operators $[\delta^2]$, $[K^2]$ that are employed in Sec. 4.1. This means that the two sets of bias parameters can be mapped onto each other, via a transformation that is calculable at a given order in PT.

To obtain the relation between the $b_O^m(R_\ell)$ and the b_O, we evaluate Eqs. (4.32)–(4.33) as integrals over the leading-order halo-matter-matter bispectrum B_{mmh}^{LO},

$$\langle \delta_\ell^2(\boldsymbol{x}) \delta_{h,\ell}(\boldsymbol{x}) \rangle = b_1 \langle \delta_\ell^3 \rangle + \frac{1}{2} b_2 \langle \delta_\ell^2 [\delta^2]_\ell \rangle + b_{K^2} \langle \delta_\ell^2 [K^2]_\ell \rangle + \cdots$$
$$\stackrel{\text{LO}}{=} \int_{\boldsymbol{k}_1} \int_{\boldsymbol{k}_2} W_\ell(k_1) W_\ell(k_2) W_\ell(k_{12}) \tag{4.34}$$
$$\times \left\{ b_1 B_{mmm}^{\text{LO}}(k_1, k_2, k_{12}) + \left[b_2 + 2 b_{K^2} \left(\left[\hat{\boldsymbol{k}}_1 \cdot \hat{\boldsymbol{k}}_2 \right]^2 - \frac{1}{3} \right) \right] P_{\text{L}}(k_1) P_{\text{L}}(k_2) \right\} \,.$$

and similarly for $\langle K_\ell^2(\boldsymbol{x}) \delta_{h,\ell}(\boldsymbol{x}) \rangle$. We now see the source of the differences between the n-point biases b_O and the moments biases $b_O^m(R_\ell)$: the former are defined without making reference to any scale R_ℓ, which
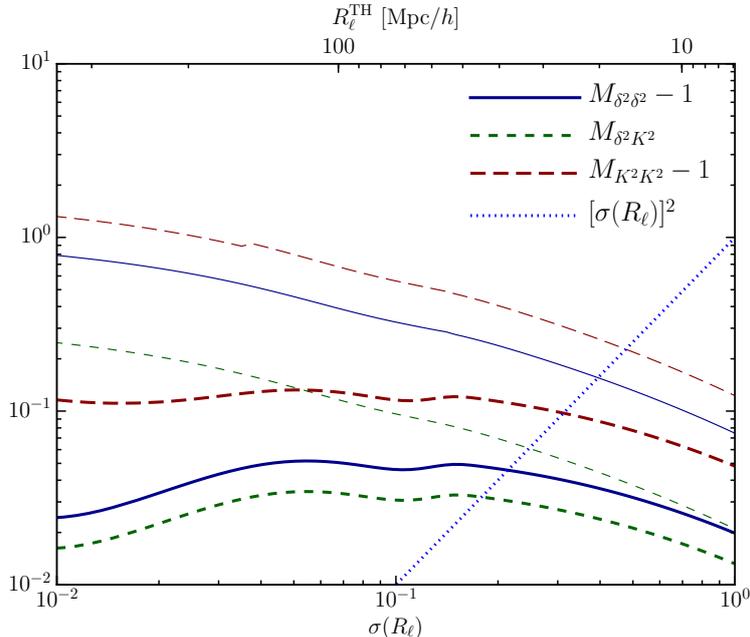
Figure 13: Elements of the bias transformation matrix $M_{OO'}$ [Eq. (4.36)] as a function of $\sigma(R_\ell)$, for two filters: real-space tophat (lower, thick set of lines) and Gaussian (upper, thin set of lines). Specifically, we show the departure of the linear map from the identity. The top axis shows the corresponding smoothing scale R_ℓ^{TH} for the tophat filter. The dotted line shows the order of magnitude of higher-order corrections $\propto \sigma^2(R_\ell)$.

only enters when smoothing the halo density field $\delta_h \rightarrow \delta_{h,\ell}$ *after* the bias expansion; hence, the correlators contain $[O]_\ell$. On the other hand, the operators $[O]^{\mathrm{m}}(R_\ell)$ corresponding to the $b_O^{\mathrm{m}}(R_\ell)$ make reference to the *smoothed linear density and tidal field* [Eqs. (4.32)–(4.33)], and thus depend explicitly on the scale R_ℓ. Note that an exactly analogous result, with $b_O \rightarrow b_O^L$, is obtained when considering moments of the proto-halo density field in Lagrangian space. In that case, the skewness of δ_ℓ vanishes.

The leading-order relation in PT between the different bias definitions can now be read off by comparing Eq. (4.32) and Eq. (4.34). For the linear bias parameter, we obtain simply

$$b_1^{\mathrm{m}}(R_\ell) \stackrel{\mathrm{LO}}{=} b_1 \,, \tag{4.35}$$

in agreement with Eq. (4.30). Thus, as long as NLO corrections of order $\sigma^2(R_\ell)$ and $R_*^2 \sigma_1^2(R_\ell)$ can be neglected, the linear moments bias is exactly equal to the linear bias inferred from the large-scale two- or three-point function. For the second-order biases, there is a nontrivial mapping already at leading order:

$$b_O^{\mathrm{m}}(R_\ell) \stackrel{\mathrm{LO}}{=} \sum_{O'=\delta^2, K^2} M_{OO'}(R_\ell) b_{O'} \,, \quad \text{where} \quad O, O' \in \{\delta^2, K^2\} \quad \text{and}$$

$$M_{OO'}(R_\ell) = \sigma^{-4}(R_\ell) \begin{pmatrix} \frac{1}{2}\langle[\delta^2]_\ell \delta_\ell^2\rangle & \frac{1}{2}\langle[K^2]_\ell \delta_\ell^2\rangle \\ \frac{45}{8}\langle[\delta^2]_\ell K_\ell^2\rangle & \frac{45}{8}\langle[K^2]_\ell K_\ell^2\rangle \end{pmatrix}_{\mathrm{LO}}. \tag{4.36}$$

Note that $b_{\delta^2} = b_2/2$, and $\langle[\delta^2]_\ell K_\ell^2\rangle = \langle[K^2]_\ell \delta_\ell^2\rangle$ at leading order. Fig. 13 shows the three independent elements of $M_{OO'}(R_\ell)$ as function of $\sigma(R_\ell)$ and for two different filters; the fourth element $M_{K^2\delta^2}$ is simply proportional to $M_{\delta^2 K^2}$. More specifically, we show the departure from the identity matrix. Clearly, the biases inferred from moments through Eq. (4.32) in general differ significantly from the biases inferred from the three-point function in a strongly R_ℓ-dependent way. The R_ℓ-dependence of $M_{OO'}(R_\ell)$ that is evident from Fig. 13 is due to the fact that $P_{\mathrm{L}}(k)$ is not a pure power law (for which the coefficients would be

R_ℓ-independent), but instead the effective power law index depends on the smoothing scale R_ℓ. Again, these results hold equally for proto-halo moments in Lagrangian space.

Nevertheless, there is an unambiguous transformation between the two bias definitions. Note that it is essential that all bias parameters that are relevant at a given order are included (only 2 at second order, but, for example, 4 at third order) in order to obtain the correct transformed biases. Moreover, if $\sigma^2(R_\ell)$ is not sufficiently small, there are significant nonlinear corrections to the relation Eq. (4.36). The order of magnitude of higher-order contributions, i.e. $\sigma^2(R_\ell)$, is shown by the dotted line in Fig. 13; if this quantity is not small, then the $b_O^{\rm m}(R_\ell)$ cannot be related to the large-scale biases b_O within the framework of perturbation theory.

For the real-space tophat filter, the bias transformation matrix $M_{OO'}$ is numerically close to the identity (within 3–15% depending on R_ℓ). Still, for precision bias measurements the difference between say b_2 and $b_2^{\rm m}(R_\ell)$ is not negligible. For the Gaussian filter, the differences between b_2 and $b_2^{\rm m}(R_\ell)$ are significant and cannot be neglected for any interesting filter scales. It might seem surprising that $b_{\delta^2}^{\rm m}(R_\ell)$ is fairly close to b_{δ^2} (assuming $\sigma^2(R_\ell)$ is negligible) for the tophat filter, given that they are related to different moments. To see why, let us write

$$W_\ell(k_{12}) = W_\ell(k_1)W_\ell(k_2)f_\ell(k_1, k_2, \hat{\boldsymbol{k}}_1 \cdot \hat{\boldsymbol{k}}_2)\,. \tag{4.37}$$

If $f_\ell = 1$, then the moment Eq. (4.34) reduces to the hierarchical result, and we recover $b_{\delta^2}^{\rm m}(R_\ell) = b_{\delta^2} + \mathcal{O}(\sigma^2(R_\ell))$. For a Gaussian filter, we simply have $f_\ell = \exp(-k_1 k_2 \mu R_\ell^2)$. A real-space tophat filter on the other hand satisfies $W_\ell^2(\boldsymbol{x}) = W_\ell(\boldsymbol{x})/V_\ell$, where $V_\ell = 4\pi R_\ell^3/3$ is the volume of the filter. This implies that

$$\int_{\boldsymbol{k}_1} |W(k_1)|^2 f(k_1, k, \mu) = \int_{\boldsymbol{k}_1} |W(k_1)|^2 = 1\,, \tag{4.38}$$

so that $f(k_1, k, \mu)$ is close to unity. This explains why, for the tophat filter, $\langle [\delta^2]_\ell \delta_\ell^2 \rangle \approx \sigma^4(R_\ell)$ while $\langle [K^2]_\ell \delta_\ell^2 \rangle \ll \sigma^4(R_\ell)$, so that the moments biases for this filter are close to the large-scale renormalized biases.

In order to determine the leading corrections of order $\sigma^2(R_\ell)$ in Eq. (4.36), one has to evaluate the NLO corrections to P_{hm} and B_{hmm}, respectively, which include higher-order and higher-derivative biases. Specifically, the leading correction to $b_1^{\rm m}(R_\ell)$ [Eq. (4.30)] is given by the integral over Eq. (4.22). The contribution proportional to b_1 captures the nonlinear evolution of matter. However, there are additional contributions at the same order which involve four additional bias parameters. The number of additional biases entering $b_2^{\rm m}(R_\ell)$ due to the NLO contribution to the halo-matter-matter bispectrum is even larger. Thus, in order to obtain rigorous, predictive measurements for the biases from moments, one has to restrict to the regime $\sigma(R_\ell) \ll 1$. At low redshifts, significant nonlinear corrections already appear for cell sizes $R_\ell \lesssim 50\,h^{-1}\,{\rm Mpc}$ [302].

A related, but subtly different method to measure LIMD bias parameters was proposed in [303, 304], where halos identified at a given redshift were traced back to the initial conditions, yielding the Lagrangian density field of halos $\delta_h^L(\boldsymbol{q})$. The Lagrangian bias parameters—including those induced by higher-derivative operators—can be obtained by cross-correlating the Lagrangian halo field δ_h^L with nonlinear transformations of the matter density field [305, 306]. The Hermite polynomials H_N provide a suitable basis to measure the Lagrangian LIMD bias parameters. Namely, one measures the joint moments

$$M_N^{\rm LCC}(R_\ell) = \left\langle [1 + \delta_h^L(\boldsymbol{q})]H_N\left[\frac{\delta_\ell^{(1)}(\boldsymbol{q})}{\sigma(R_\ell)}\right]\right\rangle\,, \tag{4.39}$$

where LCC stands for "Lagrangian cross-correlation." Crucially, unlike in the moments method discussed above, the halo density field itself is *not smoothed*. This method can be implemented simply by summing $H_N[\delta_\ell^{(1)}(\boldsymbol{q}_i)/\sigma(R_\ell)]$ evaluated at the Lagrangian halo positions $\boldsymbol{q}_i$. Since the Hermite polynomials are orthogonal with respect to a Gaussian weight, they have the useful property that $\langle H_N(\nu)H_M(\nu)\rangle = \delta_{NM}\langle\nu^2\rangle^{N/2}$

for a zero-mean Gaussian variate ν. At leading order, the first two bias parameters can be related directly to the large-scale renormalized biases b_N:[16]

$$M_1^{\mathrm{LCC}}(R_\ell) = b_1^L \frac{\langle \delta^{(1)} \delta_\ell^{(1)} \rangle}{\sigma(R_\ell)} + \mathcal{O}\left(R_*^2 \frac{\langle \nabla^2 \delta^{(1)} \delta_\ell^{(1)} \rangle}{\sigma(R_\ell)} \right)$$

$$M_2^{\mathrm{LCC}}(R_\ell) = b_2^L \left(\frac{\langle \delta^{(1)} \delta_\ell^{(1)} \rangle}{\sigma(R_\ell)} \right)^2 + \mathcal{O}\left(R_*^2 \frac{\langle \nabla^2 \delta^{(1)} \delta_\ell^{(1)} \rangle}{\sigma(R_\ell)} \right) . \tag{4.40}$$

Since these are Lagrangian moments, corrections to these expressions arise only from higher-derivative bias terms, starting with $b_{\nabla^2 \delta}^L$, just as is the case for P_{hm}^L, Eq. (4.12) [156, 305, 100, 303, 304]. Thus, in the large-R_ℓ limit where $R_*^2 \langle \nabla^2 \delta^{(1)} \delta_\ell^{(1)} \rangle \ll \sigma^2(R_\ell)$, the $b_N^{\mathrm{LCC}}(R_\ell)$ are simply related to the large-scale renormalized Lagrangian biases. This is a significant difference from the moments biases discussed above, and stems from the fact that the halo density field is not smoothed in this method. For reference, $\langle \delta^{(1)} \delta_\ell^{(1)} \rangle / \sigma^2(R_\ell)$ evaluates to $\sim 1.1 - 1.2$ for a tophat filter, depending on R_ℓ.

For smaller smoothing scales, corrections induced by higher-derivative terms $\propto \nabla^2 \delta$, and so on, become important. From the point of view of the general bias expansion, the perturbative expansion breaks down unless R_ℓ is much larger than the nonlocality scale $R_* \sim R(M)$ for the halos considered. On the other hand, in the context of the peak approach (Sec. 6), all higher-derivative contributions to Eq. (4.40) are controlled by the peak constraint and the filtering kernel, and measurements for any value of R_ℓ can be interpreted in the context of this model [see, e.g., 304, 306]. This method can be generalized to measure several different Lagrangian bias parameters, as was shown in [306] who measured the bias coefficient $b_{(\nabla \delta)^2}$. Sec. 6.7 describes this approach in more detail in the context of the peak formalism.

4.3 Scatter-plot method

We now turn to a frequently employed method to measure bias which, as we will see, is an application of the moments discussed in Sec. 4.2. Consider the conditional probability distribution of the smoothed halo density field $\delta_{h,\ell}(\boldsymbol{x})$ at an arbitrary location, given a fixed value of the smoothed matter density $\delta_\ell(\boldsymbol{x})$ at that location:

$$P(\delta_{h,\ell} | \delta_\ell) = \frac{P(\delta_{h,\ell}, \delta_\ell)}{P(\delta_\ell)} . \tag{4.41}$$

Note that we are not choosing any particular location, and statistical homogeneity then allows us to drop the argument $\boldsymbol{x}$ on the PDFs. In the following, we will drop the subscript ℓ for clarity, as all moments in the following will be defined for a fixed scale R_ℓ, and correspondingly denote $\sigma \equiv \sigma(R_\ell)$. The conditional PDF Eq. (4.41) has been studied extensively in the literature [176, 307, 308, 309]. The biases can be inferred by applying Eq. (4.41) to simulations as follows [17, 302, 307, 310]. First, one measures the halo and matter fractional density perturbations, $\delta_{h,\ell}$ and δ_ℓ, for example by counting halos in randomly thrown spheres of radius R_ℓ in the simulation volume. One then obtains the mean relation between $\delta_{h,\ell}$ and δ_ℓ by creating a "scatter plot" (see Fig. 14), and taking the mean of $\delta_{h,\ell}$ in slices of δ_ℓ. Mathematically, in the limit of infinite statistics, this corresponds to measuring

$$\langle \delta_h | \delta \rangle = \int d\delta_h \, \delta_h P(\delta_h | \delta) . \tag{4.42}$$

Then, one fits a quadratic or higher-order polynomial to $\langle \delta_h | \delta \rangle$ as function of δ. The coefficients of the polynomial then yield the estimates of the scatter-plot bias parameters. The order of the polynomial that is necessary depends on R_ℓ: at large smoothing radii R_ℓ, the matter fluctuations are sufficiently small that

---

[16]R_ℓ corresponds to R_0 in [304]. Note that they use $\langle \delta_{R(M)}^{(1)} \delta_\ell^{(1)} \rangle$ instead of $\langle \delta^{(1)} \delta_\ell^{(1)} \rangle$ to convert their results to the large-scale bias, where $R(M)$ is the Lagrangian radius of halos. This however has negligible numerical impact for the scales $R_0 \geq 50 \, h^{-1}$ Mpc used there (see Sec. 6.7 for a discussion). Further, $\langle \delta_{R(M)}^{(1)} \delta_\ell^{(1)} \rangle / \sigma^2(R_\ell) = S_\times / S_0$ in the notation of [304].

a truncation at third order is appropriate. In the following, we thus assume that the smoothing scale R_ℓ is sufficiently large so that a perturbative treatment is applicable, just as we have done for the moments derived in the previous section.

Let us begin with linear theory, where δ follows a Gaussian distribution with variance $\sigma^2 \equiv \sigma^2(R_\ell)$, and the halo overdensity is given by $\delta_h \equiv \delta_{h,\ell} = b_1\delta + \varepsilon$ where ε is a Gaussian stochastic variable. The corresponding variance of δ_h is given in Eq. (4.30), and hence we have

$$P(\delta_h, \delta)\Big|_{\text{linear}} \propto \exp\left(-\frac{1}{2}(\delta_h, \delta)\boldsymbol{C}^{-1}(\delta_h, \delta)^\top\right), \quad \text{where} \quad \boldsymbol{C} = \begin{pmatrix} b_1^2\sigma^2 + \langle\varepsilon^2\rangle & b_1\sigma^2 \\ b_1\sigma^2 & \sigma^2 \end{pmatrix}, \tag{4.43}$$

and $\langle\varepsilon^2\rangle \equiv \langle(\varepsilon_\ell^{\text{m}})^2\rangle = P_\varepsilon^{\{0\}}V_\ell^{-1}$ at this order. This yields

$$P(\delta_h|\delta)\Big|_{\text{linear}} \propto \exp\left[-\frac{(\delta_h - b_1\delta)^2}{2\langle\varepsilon^2\rangle}\right]. \tag{4.44}$$

The expectation value of δ_h conditioned on δ is then simply obtained as

$$\langle\delta_h|\delta\rangle\Big|_{\text{linear}} = b_1\delta. \tag{4.45}$$

Thus, the derivative of $\langle\delta_h|\delta\rangle$ with respect to δ, or equivalently the coefficient of the linear term of a polynomial fit to $\langle\delta_h|\delta\rangle$, yields the (renormalized) bias parameter $b_1 = b_1^{\text{m}}(R_\ell)$. Eq. (4.44) clearly shows that the scatter in δ_h at fixed δ is related to the stochasticity ε (see Eq. (4.30) and [176]). However, both of these statements strictly hold in the limit of $R_\ell \to \infty$; for a finite scale R_ℓ, Eq. (4.44) receives corrections of order $\sigma^2 = \sigma^2(R_\ell)$ from NLO terms involving the quadratic and cubic bias parameters, just as was discussed for the quadratic moments in Sec. 4.2. Note that, in case of scatter-plot bias estimates, there are further corrections of the same order from the non-Gaussianity of $P(\delta_h, \delta)$.

We now turn to the leading non-Gaussian, i.e. second-order, contributions to the conditional PDF. At this order, the PDF acquires a skewness and, using the Edgeworth expansion, can be written as

$$P(\delta_h, \delta)\Big|_{\text{2nd}} = \left\{1 - \frac{1}{6}\left[\langle\delta^3\rangle\frac{\partial^3}{\partial\delta^3} + 3\langle\delta^2\delta_h\rangle\frac{\partial^2}{\partial\delta^2}\frac{\partial}{\partial\delta_h} + 3\langle\delta\delta_h^2\rangle\frac{\partial}{\partial\delta}\frac{\partial^2}{\partial\delta_h^2} + \langle\delta_h^3\rangle\frac{\partial^3}{\partial\delta_h^3}\right]\right\}P(\delta_h, \delta)\Big|_{\text{linear}}. \tag{4.46}$$

Upon inserting this into Eq. (4.42), the second derivative with respect to δ of $\langle\delta_h|\delta\rangle$ evaluated at $\delta = 0$ yields

$$\frac{\partial^2}{\partial\delta^2}\langle\delta_h|\delta\rangle\Big|_{\text{2nd}, \delta=0} = \frac{1}{\sigma^4}\left[\langle\delta^2\delta_h\rangle - b_1\langle\delta^3\rangle\right] + \mathcal{O}(\sigma^2) = b_2^{\text{m}}(R_\ell) + \mathcal{O}(\sigma^2), \tag{4.47}$$

where we have used Eq. (4.32). This is precisely (1/2 of) the quadratic coefficient of a polynomial fit to $\langle\delta_h|\delta\rangle$. Thus, in the large-scale limit, the second-order scatter plot bias measurement yields exactly the second-order moments bias discussed in Sec. 4.2 for the same filter W_{R_ℓ}. The two differ by NLO corrections that scale as σ^2, but which are in principle calculable. Similar results hold if one were to construct the conditional PDF with respect to the tidal field smoothed on scale R_ℓ, $P(\delta_h|K^2)$. In order to connect to the biases measured from n-point functions, one needs to apply the transformation in Eq. (4.36). For measurements which use $R_\ell \lesssim 40\,h^{-1}\,\text{Mpc}$ and tophat filters, the higher-order contributions which scale as $\sigma^2(R_\ell)$ are however still larger than the effects of the transformation Eq. (4.36), as shown in Fig. 13.

Fig. 14 shows typical scatter plots of $\delta_{h,\ell}$ vs. δ_ℓ for different cell sizes and redshifts, and polynomial fits to the mean relation [307]; again, the coefficients of the polynomial fits correspond to the bias parameters inferred from this method. The variance of the matter density in cells σ^2 for each case is indicated in the caption. For the larger cells shown, $R_\ell \simeq 80\,h^{-1}\,\text{Mpc}$, the mean relation is well fitted by a second-order polynomial. The curvature of this fit corresponds to the quadratic moments bias $b_2^{\text{m}}(R_\ell)$; specifically, we see a negative ($z = 0$) or positive ($z = 0.5$) value of $b_2^{\text{m}}(R_\ell)$. In the excursion set and peaks models (Sec. 5–6), this is explained by the different peak significance of halos with fixed mass $M \gtrsim 10^{13}\,h^{-1}\,M_\odot$ at $z = 0$ and $z = 0.5$, respectively.
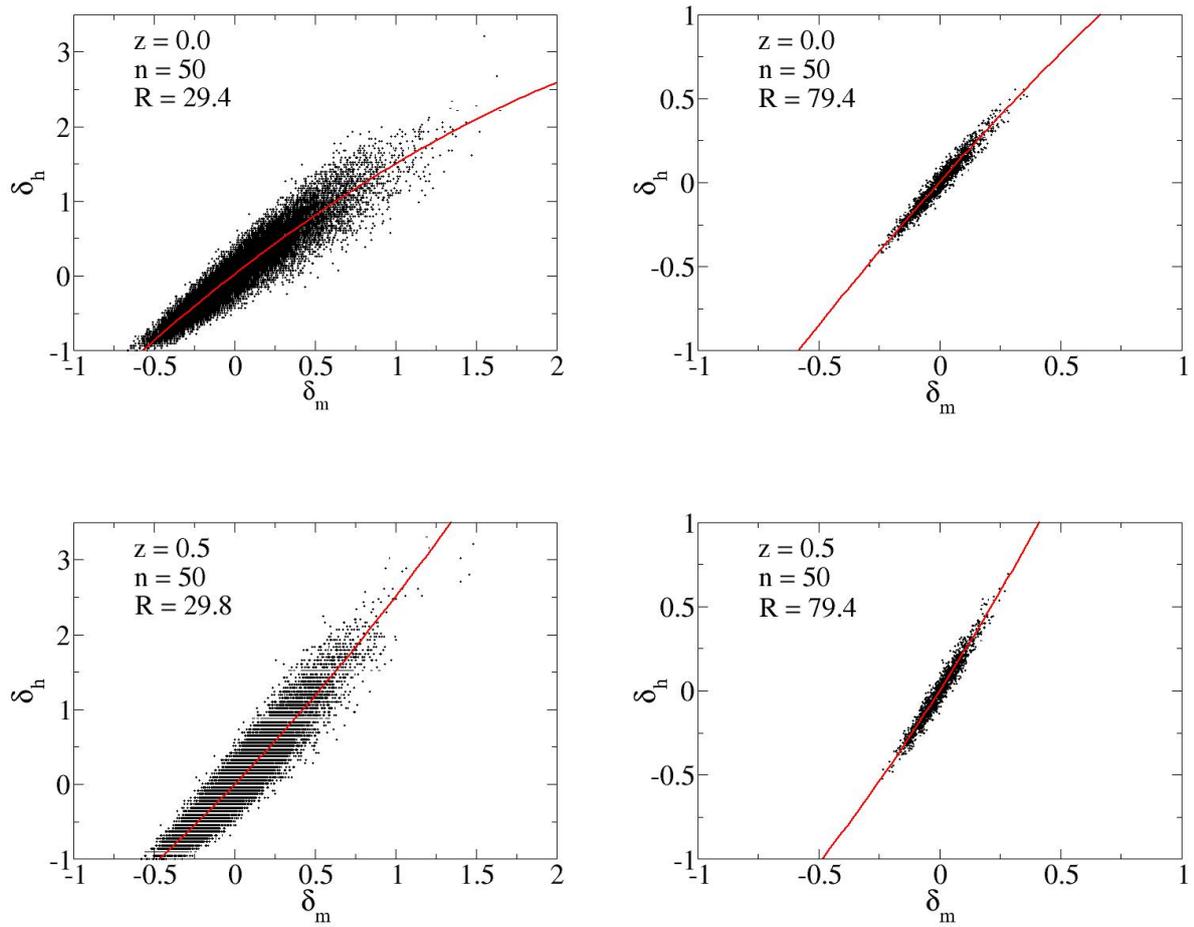
Figure 14: Scatter plots of the halo density contrast $\delta_{h,\ell}$ vs. matter density contrast δ_ℓ in spherical tophat cells of different radii and redshifts, as indicated in the caption. Only halos with $n = 50$ or more particles are included; this corresponds to a minimum halo mass of $1.17 \times 10^{13} \, h^{-1} \, M_\odot$. The solid lines represent the least-squares fits of a quadratic polynomial yielding estimates of the first two LIMD moments bias parameters. The variance of the matter density field in each case, clockwise from the top left, is $\sigma^2(R_\ell) = 0.085$, 0.010, 0.006, 0.051. From [307].
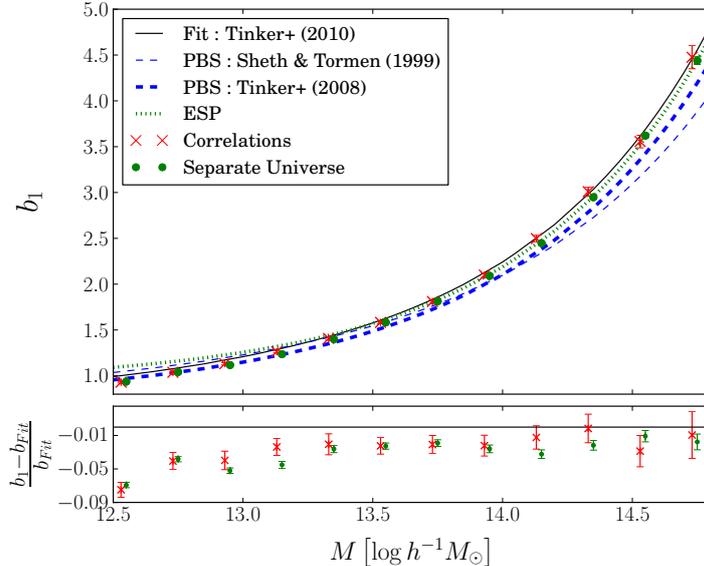
Figure 15: *Top panel:* comparison between the linear halo bias from separate-universe simulations (green dots), and from clustering (red crosses; displaced slightly horizontally for clarity). Error bars that are not visible are within the marker size. The solid black curve is the Tinker et al. (2010) [312] best fit curve for b_1 given in Tab. 7, while the dot-dashed green curve is the excursion set peak (ESP) prediction [derived from Eq. (6.75) via the PBS approach, Sec. 3.3]. Also shown are the results obtained by applying Eq. (3.26) for a constant barrier $B = \delta_{cr}$ to the Tinker et al (2008) [313] and Sheth-Tormen [16] mass functions (blue dashed curves; see also Fig. 9). *Bottom panel:* relative difference between the measurements and the fitting formula from Tinker et al. (2010). From [314].

For the smaller cell result at $z = 0.5$ (bottom-left panel), discreteness effects, which arise from the fact that there can only be an integer number of halos per cell, are apparent. This implies that the cell size is not much larger than the mean inter-halo separation, which is of order $R(M)$. For cells this small, non-perturbative effects such as exclusion become relevant. This points to the fact that scatter plot results for small R_ℓ cannot be rigorously connected with perturbative predictions on large scales.

The stochasticity between $\delta_{h,\ell}$ and δ_ℓ, which is responsible for the scatter around the mean relation in Fig. 14, is taken into account in the perturbative approach by the stochastic fields $\varepsilon, \varepsilon_\delta, \cdots$. This is appropriate as long as the scatter is small [note that both $\langle \varepsilon^3 \rangle$ and $\langle \varepsilon \varepsilon_\delta \rangle$ contribute to Eq. (4.46) at leading order, although they do not enter the second derivative at $\delta = 0$ in Eq. (4.47)]. This observations has led some authors to propose that, in the non-perturbative regime, the deterministic bias expansion be replaced by a more general bias distribution function [175, 176]. Another approach is to construct the scatter plot in terms of the log-transformed fields $\ln(1 + \delta)$ and $\ln(1 + \delta_h)$ [311]. However, these small-scale measurements strictly apply to counts-in-cells only. Namely, their results cannot be applied to the computation of the large-scale n-point correlations estimated using direct pair counting.

4.4   *Response approach*

We now turn to an alternative approach to measuring bias which is motivated by the rigorous definition of the peak-background split derived in Sec. 3.2. Specifically, the LIMD bias parameters of halos are measured through the response of the halo abundance to an infinite-wavelength density perturbation [Eq. (3.4)], implemented by performing a series of N-body simulations with different amounts of spatial curvature following the separate-universe approach. Moreover, it is possible to generalize this idea to other bias parameters, for example the tidal bias or higher-derivative biases, by measuring the response of the halo abundance to modified initial conditions that contain certain configurations of amplified long-wavelength modes. We use the term "response bias" to encompass all of these methods. Here, we will focus on the LIMD bias parameters b_N of dark matter halos.
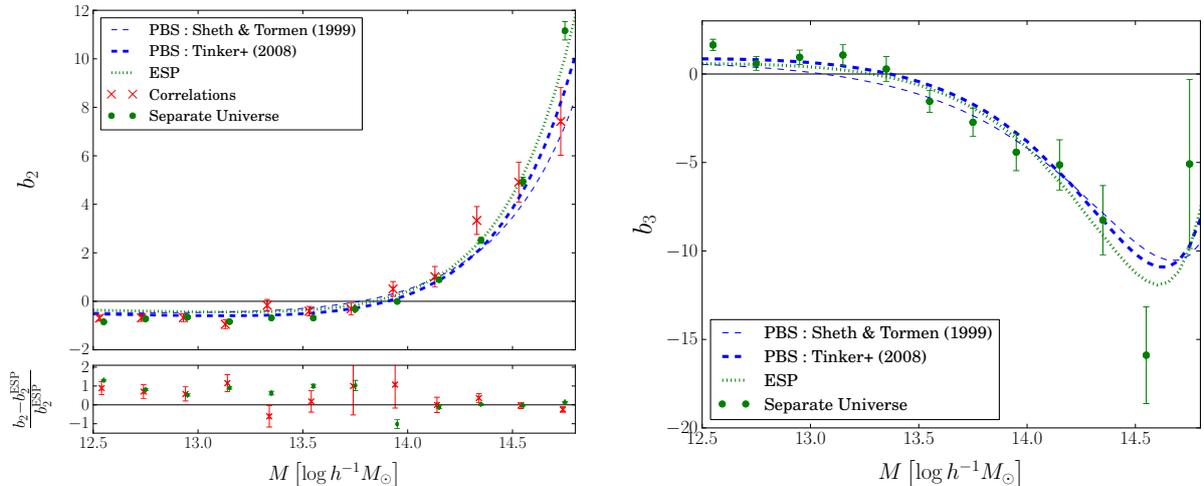
Figure 16: *Left figure, top panel:* same as Fig. 15, but for the quadratic Eulerian bias b_2. *Bottom panel:* relative difference between measurements and the theoretical prediction of the excursion set peaks (ESP). In each panel, the clustering points have been displaced horizontally as in Fig. 15. *Right figure:* as left, but for the cubic Eulerian bias b_3. From [314].

Clearly, this method is restricted to simulated tracers, and cannot be applied to actual observed galaxies. However, for simulations it offers several advantages. First, this approach directly isolates the desired bias parameters without needing to fit several parameters at the same time. Second, as we are dealing with an infinite-wavelength perturbation, we are not limited by the theoretical uncertainty of the perturbation theory predictions for finite-wavelength modes (Sec. 4.1.4). Finally, by using the same initial seeds for simulations with different long-wavelength modes, we can cancel cosmic variance to a large extent. In principle, this method is able to achieve a higher statistical precision on bias parameters for a given simulation volume than n-point functions. Whether this is attained in practice depends on the precise estimator used in measuring the response (for example, binned halo counts in [314, 315], and the cumulative mass function in [316]). Generally, one expects the statistical advantage of the response approach to be more important for higher-order biases, as can be seen when comparing the error bars in the lower panels of Fig. 15 and Fig. 16 (left panel).

This novel method has been applied in three recent papers [314, 315, 316], where [314] went to nonlinear order while [315, 316] only measured the linear bias b_1. All three references found consistency between the response approach and the n-point function measurements of b_1, as well as b_2 in the case of [314]. Fig. 15 shows a comparison of the linear bias of dark matter halos measured from the halo-matter cross-power spectrum on large scales, and that measured from the separate-universe response [314]. Fig. 16 shows the analogous result for the quadratic and cubic biases b_2, b_3. This shows explicitly that, if carefully measured, both techniques unambiguously measure the same bias parameters (the same is true for the moments-based approaches, provided sufficiently large smoothing scales are chosen and the transformation Eq. (4.36) is applied). Ref. [314] also provided fitting formulas for b_2, b_3 in terms of b_1 (Fig. 17), combining results at various redshifts since the relation was found to be redshift-independent. The fitting formulas are given in Tab. 7. The fitting formula for $b_2(b_1)$ given in [310], obtained from fits to the two- and three-point correlation functions in real space, agrees with the one given in Tab. 7, with differences much smaller than the error bars on the measurements.

Figs. 15–16 also compare the simulation results to predictions based on the peak-background split applied to universal mass functions (Sec. 3.3). Once a specific barrier B is assumed, the PBS allows for a derivation of all LIMD bias parameters b_N from a given halo mass function. The simplest and most common choice is $B = \delta_{\rm cr}$. Figs. 15–16 show the result of applying this to the Sheth-Tormen ([317], see Eq. (3.31) for the explicit expressions) and Tinker et al (2008) [313] mass function prescriptions. Even though the latter provides a very accurate mass function, the linear bias derived via the PBS and simple collapse threshold is

94
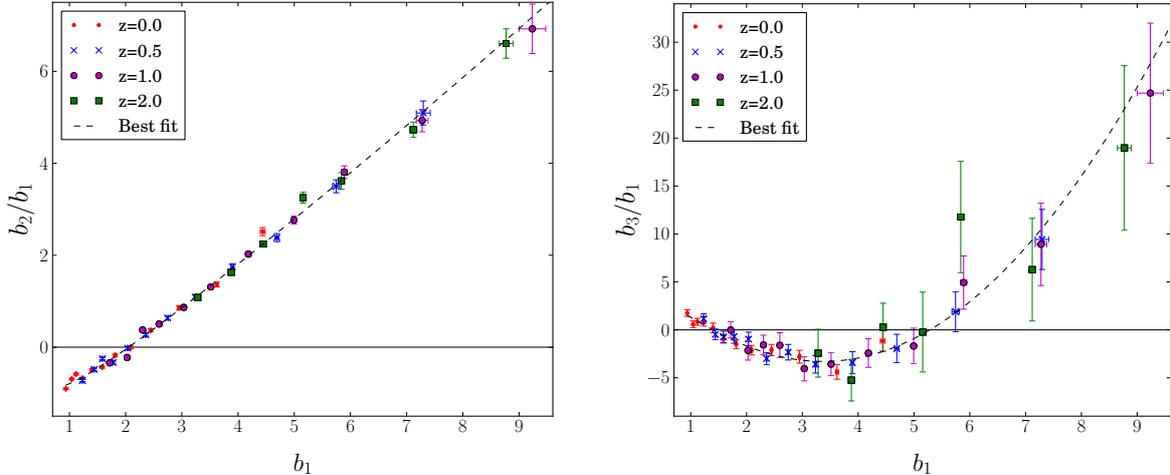
Figure 17: Eulerian bias parameters b_2 *(left panel)* and b_3 *(right panel)*, each divided by b_1, as a function of b_1 obtained from separate-universe simulations and for different redshifts. The dashed curves show the third-order best-fit polynomials which are summarized in Tab. 7. From [314].

only accurate at the $\sim 10\%$ level, as shown previously in [226, 312, 227]. The agreement is even worse for b_2, with up to 50% discrepancy at low mass, although the absolute difference between the PBS predictions and the measurements is similar to that in b_1.

Finally, Figs. 15–16 also show the biases predicted by the excursion set peak (ESP) approach, described in detail in Sec. 6.8, which includes a stochastic moving barrier motivated by simulation results. At high mass, this performs much better, at least for b_1, showing that, in the context of the PBS applied to universal mass functions, the choice of barrier is a key ingredient in deriving accurate bias parameters. In this context, it is important to note that previous results on the inaccuracy of PBS bias parameters [226, 227] relied on the simple constant threshold $B = \delta_{\rm cr}$. This shows that the cause of theses inaccuracies is not the peak-background split itself. Interestingly, the results of Fig. 17 show that there is still some universal scaling behavior in the higher-order bias parameters of halos: the relations $b_2(b_1)$ and $b_3(b_1)$ are found to be independent of redshift to within the uncertainties. For this reason, the fitting formulas of these relations (Tab. 7) are expected to depend only weakly on cosmology.

We conclude this discussion by reiterating that the "inaccuracy of the peak-background split" depends on what one defines PBS to mean, and summarize it as follows:

- As shown in Sec. 3, local bias parameters derived using the PBS implemented via the separate-universe response approach are *physically the same parameters* as those obtained from large-scale n-point correlations. Implementations of both approaches in N-body simulations have clearly confirmed this (Figs. 15–16).

- PBS biases derived using universal mass function based on a simulation-calibrated stochastic moving barrier [318, 303] are accurate to a few percent for b_1, at least at high masses [314].

- The PBS using the constant spherical collapse barrier is no better than 10% for b_1, and worse for b_2 [226, 227, 314].

4.5 An overview of bias measurements

We now briefly review actual measurements of bias parameters for dark matter halos in simulations and observed galaxies. We then review measurements and physical models of halo stochasticity, i.e. the stochastic amplitudes $P_\varepsilon, B_\varepsilon, \cdots$ which enter halo n-point statistics.

95

4.5.1 Halo bias

Early measurements of the large-scale halo bias using two-point functions are presented in, e.g., [296, 16, 319, 320, 321]. Some of the most accurate recent measurements of the linear halo bias can be found in [312], who used $P_{hh}(k)$ on large scales to infer b_1 for spherical overdensity (SO) halos in a large suite of N-body simulations (see Appendix D for a brief description of halo finding algorithms). For this, they assumed that $P_\varepsilon^{\{0\}} = 1/\overline{n}_h$. They also cross-checked the results with the linear bias parameter measured from the halo-matter cross-power spectrum, and find that, within the uncertainty of the measurements, the deviation from perfect Poisson stochasticity (Sec. 4.5.3) does not significantly affect the bias measurement. The resulting fitting function, which is accurate to 5–10% for a wide range of SO halo definitions and ΛCDM cosmologies, is summarized in Tab. 7. Note that the universality of b_1, i.e. that its mass and cosmology dependence are completely described by the parameter $\nu_c = \delta_{\rm cr}/\sigma(M)$, is not perfect (see Fig. 1 in [312]). As discussed in detail in Sec. 4.4, Ref. [314] showed that the bias fitting function agrees very well with the response bias measurements (see Fig. 15). It is well established that the bias b_1 of halos, at fixed halo mass, also depends on other halo properties. This is known as *assembly bias* and will be discussed in Sec. 4.6.

Robust measurement of the large-scale nonlinear halo bias parameters from n-point functions have only been published fairly recently. Refs. [322, 323] combined measurements of halo two- and three-point functions to infer b_1, b_2 assuming Eulerian LIMD bias, i.e. setting $b_{K^2} = 0$ in Eqs. (4.2)–(4.4). Ref. [323] did not find complete consistency between two- and three-point function measurements (see also [308]), which could be an indication that b_{K^2} is an important ingredient [118, 129] (see below). Refs. [137, 267] included the NLO corrections to P_{hm} in addition to the leading-order bispectrum $B_{hmm}^{\rm LO}(k_1, k_2, k_3)$ of friends-of-friends (FoF) halos, to measure b_1, b_2, b_{K^2}, and $b_{\rm td}$. Note that they used different parametrizations for the operators K^2 and $O_{\rm td}$ (see Appendix C for the relations between various different parametrizations). Ref. [137] set $b_{\nabla^2\delta}$ to zero, and performed a fit including modes with $k \lesssim 0.1\,h\,{\rm Mpc}^{-1}$ at $z = 0$, while Ref. [267] fit up to scales of $0.15\,h\,{\rm Mpc}^{-1}$ in the bispectrum and $0.3\,h\,{\rm Mpc}^{-1}$ in the power spectrum. The linear and quadratic bias parameters obtained from the two-and three-point functions were all found to be consistent.

Regarding real-space measurements, Ref. [301] analyzed N-body simulations using the LIMD bias assumption and found a 10% deviation of the linear growth factor estimated by combining the two-point and three-point correlation functions. Later, the same group [254] presented measurements of b_1, b_2 and b_{K^2} from the two-point halo-halo and halo-matter correlation functions, as well as the halo-matter-matter three-point function in real space. They found consistency in the measured value of b_1 from the two- and three-point functions, and, furthermore, confirmed the necessity of including b_{K^2} to correctly fit the latter.

Let us now briefly discuss the significance of the numerical results obtained for b_{K^2} and $b_{\rm td}$. Early studies of b_{K^2} found consistency with the Lagrangian LIMD prediction of $b_{K^2} = -2/7(b_1 - 1)$. This includes Refs. [129], who used Eulerian measurements of B_{mmh}, the same observable used by [118]. The latter, however, found evidence for a departure from the Lagrangian LIMD prediction at low halo mass, with $b_{K^2} > -2/7(b_1 - 1)$. Similar deviations were also reported for low mass halos in [254] from an analysis of the three-point correlation function. Importantly however, these authors obtained $b_{K^2} < -2/7(b_1 - 1)$. By contrast, the results of [121], obtained using the Lagrangian bispectrum [Eq. (4.12)], indicate a departure from Lagrangian LIMD at high masses, whereas LIMD provides a good fit at low masses. Subsequently, Ref. [137] presented combined constraints from the cross-power spectrum P_{hm} and bispectrum B_{mmh}, which are shown in Fig. 18. While they follow the general trend of the Lagrangian LIMD prediction, there are some indications for a deviation at lower values of b_1, corresponding to a Lagrangian tidal bias $b_{K^2}^L < 0$ for low-mass halos. These results were very recently confirmed by two studies [324, 325] who used the cross-correlation of halos with quadratic operators constructed from the smoothed density field, first proposed by [244]. Another recent study [326] measured $b_{K^2}^L$ directly using the method described in Sec. 4.2 (see Sec. 6.7 for the theoretical background), and found results consistent with Lagrangian LIMD, albeit with large error bars. Finally, Ref. [245] presented different estimates of the moments bias $b_{K^2}^{{\rm m},L}$ using halos in Lagrangian space. Their "PBS estimator" corresponds to the scatter-plot method described in Sec. 4.3 applied in Lagrangian space using a cell size $R_\ell \approx 150\,h^{-1}\,{\rm Mpc}$, which strongly suppresses higher-order corrections. Note that this yields the moments biases in our notation, which they correct to obtain the large-scale $b_{K^2}^L$ using the results of Sec. 4.2. They found strong evidence for $b_{K^2}^L \neq 0$, with $b_{K^2}^L$ positive
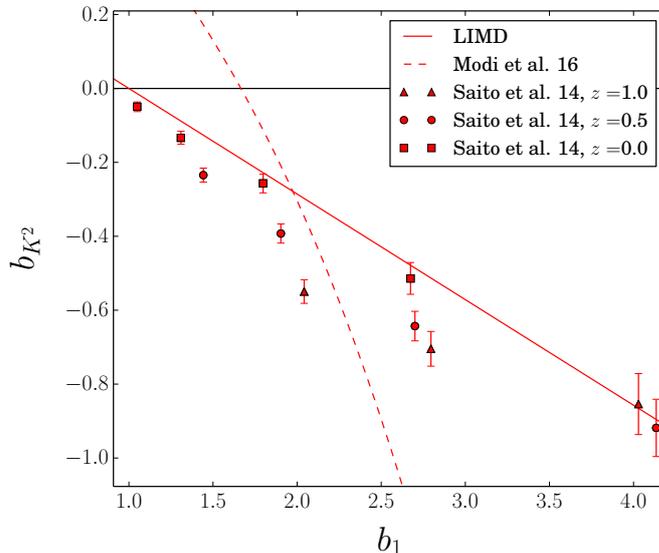
Figure 18: Eulerian tidal bias parameter b_{K^2}, measured for FOF-identified dark matter halos, as a function of their linear Eulerian bias b_1. The points show the measurements from [137] at the redshifts indicated, while the dashed line shows the fitting formula derived by [245] from their measurements (at $z = 0$ and $z = 1$) using halo moments in Lagrangian space, and converted to the large-scale Eulerian bias parameter. The solid line shows the prediction of the Lagrangian LIMD ansatz, $b_{K^2} = (-2/7)(b_1 - 1)$ [Sec. 2.3].

(negative) for low (high) mass halos. The fitting function for $b_{K^2}(b_1)$ they obtain from their measurements is also shown in Fig. 18. There is significant disagreement with the results from the halo bispectrum by [137]. This disagreement, which could possibly be due to the nontrivial conversion from the measured halo moments to the large-scale bias b_{K^2}, clearly warrants further investigation.

Only few measurements of the cubic-order bias parameter b_{td} have been reported [137, 267], which were based on the NLO halo power spectrum. Ref. [137] found broad consistency with the Lagrangian LIMD prediction for the combination [see Eq. (C.23)]

$$b_{3\text{nl}} \equiv -\frac{32}{21}\left(b_{K^2} + \frac{2}{5}b_{\text{td}}\right) \overset{\text{Lagr. LIMD}}{=} \frac{32}{315}b_1^L \,, \tag{4.48}$$

which is proportional to the combination of bias parameters that multiplies $f_{\text{NLO}}(k)P_{\text{L}}(k)$ in the NLO halo power spectrum [Eq. (4.22)]. The second equality follows from the Lagrangian LIMD prediction for b_{K^2} and b_{td}. Note, however, that the constraints on the combination $b_{3\text{nl}}$ from [137] are based on its contribution to the NLO halo power spectrum, and are thus strongly degenerate with the contribution from the higher-derivative bias $b_{\nabla^2\delta}$ (see Fig. 12), for which Ref. [137] assume $b_{\nabla^2\delta} = 0$. Thus, further measurements are needed to break the degeneracy between $b_{3\text{nl}}$ (or b_{td}) and $b_{\nabla^2\delta}$.

Turning to the higher-derivative bias $b_{\nabla^2\delta}$, the first constraints have been placed by studies testing the scale dependence of bias on large scales [105, 327]. More recent measurements include those of [264]. For their mass bin "2", which follows the definition of Table 1 in [328], they obtain a value of $b_{\nabla^2\delta}$ that is of order $3[R(M)]^2$, although $b_{\nabla^2\delta} = 0$ is only ruled out at the $\sim 1.3\sigma$ level, and no constraint is found for the other mass bins. On the other hand, Ref. [267] quote values for $b_{\nabla^2\delta}$ that are much smaller than $[R(M)]^2$ when using either of the two different definitions of b_{c_s} given there [Eq. (18) and Eq. (31)]. Thus, there is still large uncertainty in the magnitude of higher-derivative biases for halos. Note that there is a strong degeneracy in shape between the contribution from $b_{\nabla^2\delta}$ and the NLO term controlled by $b_{K^2} + (2/5)b_{\text{td}}$ (e.g., [197]; see Fig. 12). Thus, without an independent constraint on (or assumption about) the third-order

97

Bias		Fitting formula	Mass range	Halo finder	Reference
b_1	$=$	$1 - A\nu_c^\alpha[\nu_c^\alpha + \delta_{\rm cr}^\alpha]^{-1} + B\nu_c^\beta + C\nu_c^\gamma$ $A = 1 + 0.24\,y\,e^{-(4/y)^4},\ \alpha = 0.44y - 0.88$ $B = 0.183,\ \beta = 1.5$ $C = 0.019 + 0.107y + 0.19e^{-(4/y)^4},\ \gamma = 2.4$	$-0.4 \lesssim \nu_c \lesssim 0.6$	SO	[312]
b_2	$=$	$0.412 - 2.143\,b_1 + 0.929\,(b_1)^2 + 0.008\,(b_1)^3$	$1 \lesssim b_1 \lesssim 9$	SO	[314]
b_3	$=$	$-1.028 + 7.646\,b_1 - 6.227\,(b_1)^2 + 0.912\,(b_1)^3$	$1 \lesssim b_1 \lesssim 9$	SO	[314]

Table 7: Published fitting formulas for bias parameters of dark matter halos in ΛCDM N-body simulations. These phenomenological fitting functions should only be trusted within the mass range over which they were calibrated on simulations; the respective approximate range, in terms of ν_c or b_1, is indicated. The third column indicates how halos were identified (Appendix D): using spherical overdensity (SO) or friends-of-friends (FOF) algorithms. In [312], halos are identified using a SO criterion $\Delta_{\rm SO}$ defined with respect to the background matter density, where $y = \log_{10} \Delta_{\rm SO}$, $\nu_c = \delta_{\rm cr}/\sigma(M)$, and $\delta_{\rm cr} = 1.686$ is fixed independently of the background cosmology; Ref. [314] use an SO finder with $\Delta_{\rm SO} = 200$. Note that the fitting formulas for $b_2(b_1)$, $b_3(b_1)$ provided in Refs. [227, 310] agree with those of [314] to well within the error bars. We choose the latter here as it was calibrated over a larger range of b_1.

bias parameter $b_{\rm td}$, the higher-derivative bias is very difficult to extract from the halo power spectrum.

Higher-derivative biases can be measured more easily in Lagrangian space, using either the halo-matter power spectrum [259, 174] or the projection method of [304, 306] discussed near the end of Sec. 4.2. In particular, [259, 174] measured the so-called peak bias b_{01}, which contributes to $-b_{\nabla^2\delta}^L$ along with the leading contribution from the filtering kernel [Eq. (2.69) in Sec. 2.6], and obtained a contribution $b_{01} \approx 2[R(M)]^2$ for halos with mass $M \geq 8 \times 10^{12}\,h^{-1}\,M_\odot$, with only a weak departure from the simple $[R(M)]^2$ scaling with mass. The negative sign of this contribution to $b_{\nabla^2\delta}^L$ is the opposite of what simple filtering yields, but is expected if halos collapse from Lagrangian patches near initial density peaks [221] (Sec. 6.6.1). However, due to the impact of halo velocity bias on the evolution of higher-derivative biases, it is not possible to relate Lagrangian higher-derivative biases to their Eulerian counterpart without using a model for the amplitude and time evolution of velocity bias [152, 174, 329] (see the discussion in Sec. 2.7). Further, Ref. [306] measured $b_{(\nabla\delta)^2}^L \sim -R(M)^{3.5}$ (which corresponds to χ_1 in their notation).

We summarize the most precise recently published fitting functions for the leading halo bias parameters in Tab. 7.

4.5.2 Galaxy bias

As we have discussed in Sec. 1.1, the linear bias parameter b_1 has been measured many times in galaxy surveys over the past 50 years. Here, we briefly review the measurements of nonlinear bias parameters from higher-order correlation functions. Ref. [330] and [331] have measured the linear and second-order LIMD bias parameters from galaxies in, respectively, IRAS (Infrared Astronomical Satellite) and 2dFGRS (2dF Galaxy Redshift Survey), using the bispectrum (they set $b_{K^2} = 0$). The results for b_2, a marginally detected value of $b_2 \approx -0.3$ for IRAS and a result consistent with zero for 2dFGRS, are broadly consistent with the expectation for dark matter halos of the same linear bias b_1 ($b_1 \approx 0.76$ and 1.0, respectively). Later, Ref. [332] showed that the relation between b_1 and b_2 predicted by the PBS relation Eq. (3.26) for halo mass functions can explain the amplitude of the galaxy bispectrum for equilateral configurations extracted from SDSS data. The three-point correlation function of the SDSS luminous red galaxy (LRG) sample was measured by [333]. Ref. [334] measured b_2 from the three-point correlation function of the WiggleZ survey. For the recent SDSS-III CMASS sample, Ref. [335] have constrained b_2 by combining the power spectrum and bispectrum, while Ref. [336] used the three-point function. Further, Ref. [337] employed the position-dependent correlation function [338], which corresponds to an integral over the three-point function in the squeezed limit, to constrain b_2 for the same galaxy sample. All three references emphasize that, unfortunately, these estimates are still limited by residual systematic uncertainties (in particular due to redshift-space distortions) in the model that is used to fit the bias parameters. A related method to estimate

the bispectrum in the squeezed configuration has been applied to the cross-correlation of the Lyman-α forest power spectrum and CMB lensing [339] to constrain the bias parameters b_1 and b_2 of the Lyman-α forest.

4.5.3 Stochasticity

The deterministic bias contributions discussed so far in this section are only a subset of the complete bias expansion, Eq. (4.1) up to third order, which also contains stochastic terms. We now discuss these stochastic contributions, restricting ourselves to the results obtained for halos in simulations. The qualitative picture also applies to galaxies; we will return to additional stochastic effects which appear for galaxies in Sec. 9.1. We begin with the most frequently studied contribution $P_\varepsilon(k) = \langle \varepsilon(\boldsymbol{k})\varepsilon(\boldsymbol{k}')\rangle'$. If halos are a perfect Poisson sample with a mean comoving density (per logarithmic mass interval) $\overline{n}_h$, then $P_\varepsilon(k) = P_\varepsilon^{\{0\}} = 1/\overline{n}_h$. Moreover, under this assumption the shot noise of halos of different mass is uncorrelated. One can succinctly summarize this prediction as [340]

$$P_\varepsilon(k; M, M') \equiv \langle \varepsilon(\boldsymbol{k}, M)\varepsilon(\boldsymbol{k}', M')\rangle' \overset{\text{Poisson}}{=} \frac{\delta_D(\ln M - \ln M')}{\overline{n}_h(M)} \,. \tag{4.49}$$

However, this is only approximately correct for actual halos [341, 127, 177, 342, 307]. On scales much smaller than the Lagrangian radii $R(M)$, $R(M')$ of the halos, one expects Eq. (4.49) to be accurate, since the likelihood of a halo center being located within a given cell is then a true Poisson process. Thus, Eq. (4.49) is expected to hold for $k \gg [R(M)+R(M')]^{-1}$ [340]. On the other hand, on large scales $k \ll [R(M)+R(M')]$ the halo model paradigm (see [343] for a comprehensive overview) provides a different constraint. If we make the assumption that all matter is contained within virialized halos, then mass and momentum conservation of matter imply that the matrix $P_\varepsilon^{\{0\}}(M, M') = \lim_{k\to 0} P_\varepsilon(k; M, M')$, which quantifies the halo stochasticity in the large-scale limit, has a zero eigenvalue, whose eigenvector corresponds to weighting each halo by mass [177, 340]. This is because any stochastic contribution to the nonlinear matter power spectrum has to scale as k^4 on large scales [344]. Thus, within the halo model framework, the mass-weighted power spectrum of all halos enjoys zero stochasticity [345]. In the presence of a diffuse matter component that is not associated with halos of any mass, this constraint does not need to hold exactly. No further robust constraints on $P_\varepsilon(k; M, M')$ are known beyond these two limiting cases, although, as mentioned above, the transition between the two limits is expected to happen at the Lagrangian radius $k \sim R^{-1}(M)$ of halos, i.e. the same scale that determines the amplitude of higher-derivative biases. Hence, one expects that

$$\left|P_\varepsilon^{\{2\}}\right| \sim [R(M)]^2 P_\varepsilon^{\{0\}} \,, \quad \text{where} \quad P_\varepsilon^{\{2\}} \equiv \left.\frac{\partial P_\varepsilon(k)}{\partial k^2}\right|_{k=0} \,, \tag{4.50}$$

keeping in mind the caveat about possible contributions that are controlled by k_{NL} instead of $R(M)$ (Sec. 2.8). Importantly, beyond the large-scale limit, measurements of stochasticity depend on the precise definition of the "deterministic" part of halo clustering, for example whether NLO corrections which involve higher-order biases are included following Sec. 4.1.4. The reason is that higher-order terms which are not accounted for in the deterministic model can contribute to the inferred stochastic contribution (e.g. [126, 127, 346]). Finally, an upper bound can be placed on the leading stochastic contribution to the halo-matter cross-power spectrum [Eq. (4.22)] by assuming that the fields ε (halos) and ε_m (matter) are perfectly correlated. This leads to

$$\left|P_{\varepsilon\varepsilon_m}^{\{2\}}\right| \le \sqrt{P_\varepsilon^{\{0\}} P_{\varepsilon_m}^{\{4\}}} \,, \tag{4.51}$$

where $P_{\varepsilon_m}^{\{4\}}$ is the leading stochastic contribution to the matter power spectrum in the EFT approach (Appendix B.3).

Ref. [177] presents a detailed simulation study of the halo stochasticity covariance, which they estimated as

$$\hat{P}_\varepsilon(k; M, M') \equiv \left\langle \left[\delta_h(\boldsymbol{k}; M) - b_1(M)\delta_m(\boldsymbol{k})\right]\left[\delta_h(\boldsymbol{k}'; M') - b_1(M')\delta_m(\boldsymbol{k}')\right] \right\rangle' \,, \tag{4.52}$$

by dividing the halo mass range ($\sim 10^{13} - 3 \cdot 10^{15}\,h^{-1}\,M_\odot$) into 10 bins of equal number density. The linear bias was determined using $P_{hm}(k)$ for $k < 0.024\,h\,\text{Mpc}^{-1}$. Interestingly, they do find an eigenvalue of

$\hat{P}_\varepsilon(k; M, M')$ that is significantly lower than $1/\overline{n}_h$, whose eigenvector is close to mass-weighting, as expected from the halo model argument made above. This is also reflected in the components of $\hat{P}_\varepsilon(k; M, M')$ at high masses, which are lower than the Poisson expectation. Moreover, there is one eigenvalue that is significantly larger than $1/\overline{n}_h$. The components of the associated eigenvector appear to be close to the quadratic bias $b_2(M)$. The other eigenvalues were found to be consistent with $1/\overline{n}_h$. All these results were found to be roughly scale independent for $k \lesssim 0.2\,h\,\mathrm{Mpc}^{-1}$. The structure of the halo stochasticity matrix is relevant for methods that use multiple, weighted tracers within the same volume to reduce the sample variance and shot noise in constraints of scale-dependent bias from primordial non-Gaussianity (see Sec. 7.6.2).

All these features can be qualitatively understood through a toy model of the clustering of proto-halos in Lagrangian space, combining exclusion at small-scale $r < R(M)$, and nonlinear bias at larger scales $r > R(M)$ [178]. In this ansatz, including quadratic LIMD bias, the stochastic contribution to the proto-halo power spectrum in Lagrangian space takes the form

$$P_\varepsilon(k, M) \overset{k\to 0}{=} \frac{1}{\overline{n}_h} + \frac{1}{2}(b_2^L)^2 \int_{R(M)}^\infty d^3r\,[\xi_{\mathrm{L}}(r)]^2 - (b_1^L)^2 \int_0^{R(M)} d^3r\,\xi_{\mathrm{L}}(r) - V_{\mathrm{excl}}\,, \tag{4.53}$$

where the second-term on the right-hand side is the $k \to 0$ limit that is subtracted in $\mathcal{I}^{[\delta^2, \delta^2]}$, and the exclusion volume is $V_{\mathrm{excl}} = 4\pi R^3(M)/3$. Note that while Eq. (4.53) is given in Lagrangian space, the constant stochastic contribution to $P_{hh}(k)$ in the large-scale limit trivially transforms to Eulerian space, since $\delta_h^E = \delta_h^L + \delta$ at linear order. Eq. (4.53) clearly shows that exclusion lowers the stochasticity (whence the low eigenvalue $< 1/\overline{n}_h$), whereas nonlinear bias enhances it (whence the large eigenvalue $> 1/\overline{n}_h$). Further, the scale dependence of the stochastic contribution is controlled by the scale $R(M)$. Quantitative agreement is difficult to obtain with such perturbative toy models because stochastic contributions to the low-k power spectrum are genuinely non-perturbative as discussed in Sec. 2.8 (see [347] for a non-perturbative, albeit one-dimensional approach).

We finally briefly discuss theoretical expectations for the other two stochastic amplitudes $P_{\varepsilon\varepsilon_\delta}^{\{0\}}$ and $B_\varepsilon^{\{0\}}$ that have appeared in our discussion of halo n-point functions. First, for Poisson shot noise, the skewness is given by $B_\varepsilon^{\{0\}} = (\overline{n}_h)^{-2}$; more specifically, if the stochasticity in halo counts of different mass is independent, we have [340]

$$B_\varepsilon(k; M, M', M'') \equiv \langle \varepsilon(\boldsymbol{k}, M)\varepsilon(\boldsymbol{k}', M')\varepsilon(\boldsymbol{k}'', M'')\rangle' \overset{\mathrm{Poisson}}{=} \frac{\delta_D(\ln M - \ln M')\delta_D(\ln M - \ln M'')}{[\overline{n}_h(M)]^2}\,. \tag{4.54}$$

Corrections to the Poisson bispectrum from exclusion and nonlinear bias are of similar magnitude as in the case of P_ε [348]. Under the same assumptions as made for Eq. (4.49) and Eq. (4.54), the cross-correlation of ε and ε_δ for halos is given by [323, 349, 340]

$$P_{\varepsilon\varepsilon_\delta}(k; M, M') \equiv \langle \varepsilon(\boldsymbol{k}, M)\varepsilon_\delta(\boldsymbol{k}', M')\rangle' \overset{\mathrm{Poisson}}{=} \frac{1}{2}b_1(M)\frac{\delta_D(\ln M - \ln M')}{\overline{n}_h(M)}\,. \tag{4.55}$$

The interpretation of this result is quite simple: a large-scale density perturbation δ modulates the local halo abundance, with $\delta_h = b_1\delta$ at linear order, and correspondingly modulates the shot noise amplitude.

At this point, it is worth making the connection to some previous references [116, 350, 176] which, instead of the relation Eq. (4.2), included the stochasticity by defining a scale-dependent bias $\check{b}_1(k)$ and correlation coefficient $\check{r}_{gm}(k)$,

$$P_{hh}(k) = [\check{b}_1(k)]^2 P_{mm}(k); \quad P_{hm} = \check{b}_1(k)\check{r}_{gm}(k)P_{mm}(k)\,, \tag{4.56}$$

where in some references the Poisson expectation is subtracted from the halo auto-power spectrum $P_{hh} \to P_{hh}^{\mathrm{p.sub.}}(k) \equiv P_{hh}(k) - \overline{n}_h^{-1}$. The real-space version of Eq. (4.56) is adopted in [351, 352]. These can be related to the quantities defined in Eq. (4.2) via [353]

$$\check{b}_1(k) = b_1\sqrt{1 + P_\varepsilon^{\{0\}}/[b_1^2 P_{mm}(k)]}; \quad \check{r}_{gm}(k) = \frac{b_1}{\check{b}_1(k)}\,. \tag{4.57}$$

In the Poisson-subtracted case, $P_\varepsilon^{\{0\}}$ should be replaced with $P_\varepsilon^{\{0\}} - \overline{n}_h^{-1}$ here. Ref. [354] includes one-loop nonlinear evolution in this parametrization. Ref. [350] estimated $\check{b}_1(k)$ and $\check{r}_{gm}(k)$ from SDSS galaxies on large scales, finding significant departures from the Poisson expectation. However, Ref. [355] pointed out that this could be caused by errors in the selection function. The author estimated the correlation coefficient to be larger than 0.9; a similarly high value > 0.95 was found by [356] for 2dF galaxies using counts-in cells.

The parametrization Eq. (4.56) does not cleanly distinguish between the shape of the matter power spectrum on large scales and the small-scale physics of halo formation, which are included in Eq. (4.2) through the *constants* b_1, $P_\varepsilon^{\{0\}}$. Equivalently, the deterministic and stochastic (noise) contributions are not cleanly separated. Hence, Eq. (4.2) is preferred in the modern literature.

4.6 Halo assembly bias

So far, we have only considered the dependence of the halo bias parameters on mass and redshift. However, it is now well established that halo bias depends on various other properties of halos at fixed mass and redshift. These dependencies have come to be known summarily as *assembly bias*. This section provides a brief overview of these results.

While the first numerical studies did not provide any conclusive evidence for a dependence of halo clustering on additional properties [357, 358], Refs. [359, 360, 361] showed that, at fixed halo mass and redshift, halos in dense regions collapse at a slightly higher redshift than in underdense environments, where the local environment density was defined as the matter density in spheres of radius $R_{\mathrm{env}} \sim \mathrm{few}\,h^{-1}\,\mathrm{Mpc}$ centered on the halo. The local environment density is expected to correlate with the halo formation time, and indeed Ref. [362] found a dependence of halo bias, at fixed mass, on the halo formation time, as shown in Fig. 19. As the halo profile shape or concentration also correlates with the halo assembly history [363, 364], assembly bias has also been found in terms of concentration. Assembly bias has since been found in several other halo properties, and most likely cannot be explained completely by the halo formation history [365]. We now provide an overview of the trends found in the literature to date, before discussing implications for models of halo biasing:

- *Concentration:* this parameter quantifies the shape of the spherically averaged halo profile. It is determined either by fitting a profile of the NFW form [366] to the measured halo profile, or by matching the measured maximum circular velocity of the halo to the NFW profile prediction. The latter estimate is sensitive to substructure, and less reliably related to the halo profile shape.

- *Formation redshift:* this is commonly defined as the epoch, phrased in terms of cosmological redshift, at which the main progenitor of the halo assembled 50% of the final halo mass. A large formation redshift thus corresponds to early-formed, or old, halos.

- *Lagrangian slope:* defined as $d\langle\delta^{(1)}\rangle_{\mathrm{proto\text{-}halo}}/dM$, this is measured in [367] by estimating the over-density of the Lagrangian proto-halo patch, and taking its derivative with respect to mass using the measured halo assembly history (i.e., removing particles first that have joined the halo last).

- *Spin parameter:* a measure of the angular momentum of halos, this is defined as $\lambda = |\boldsymbol{J}|/(\sqrt{2}M_\Delta V_\Delta R_\Delta)$, where R_Δ, M_Δ denote the radius and mass within which the mean density reaches Δ times the background matter density $\overline{\rho}_m$, and $V_\Delta = (GM_\Delta/R_\Delta)^{1/2}$ is the circular velocity at R_Δ.

- *Halo sphericity:* this parameter is defined as the square-root of the ratio of smallest to largest elements of the diagonalized inertia tensor within R_Δ.

- *Velocity anisotropy:* the parameter β is defined as $\beta = 1 - \sigma_{v,t}^2/(2\sigma_{v,r}^2)$, where $\sigma_{v,t}^2$, $\sigma_{v,r}^2$ are the mean tangential and radial velocity dispersions within the halo, respectively. A value of $\beta = 0$ corresponds to an isotropic velocity dispersion; positive values correspond to preferentially radial orbits, while negative values imply more circular orbits.

- *Subhalo mass fraction:* this is defined as the fraction of the total mass within R_Δ which is in the

Halo property	Low-mass halos	High-mass halos
Concentration[1] [363, 368, 231, 230, 302, 369, 370, 371, 365]	↗	↘
Formation redshift [362, 363, 368, 231, 230, 365]	↗	→
Lagrangian slope [367]		↗
Spin parameter [231, 369, 371, 365]	↗	↗
Sphericity[2] [369, 371]	↗	↗
Subhalo mass fraction[3] [231, 365]	↘	↗
Velocity anisotropy β [369]	↘	↘
Late-time accretion rate[4] [371, 365]		↘

[1] Note that different definitions of concentration are used in the literature. Refs. [363, 368, 230, 370, 365] use that obtained from a fit to the density profile, while Refs. [231, 302, 369, 371] use that obtained from the maximum circular velocity.
[2] Similar trends are found for the sphericity of the velocity field within halos [369].
[3] Ref. [365] use the number and average separation of subhalos for high-mass halos, and find the same trend.
[4] Ref. [371] finds trends that are not strictly monotonic.

Table 8: Overview of assembly bias trends that have been reported in the literature. An upward pointing arrow means that bias increases with increasing value of the given property, while a downward pointing arrow denotes a decreasing bias. A horizontal arrow indicates absent or weak trends, while no symbol indicates no known constraints. "Low-" and "high-mass" halos refer to halos with $M \ll M_\star$ and $M \gg M_\star$, respectively.

form of self-bound substructures. Similar trends hold for the mass fraction in the main subhalo [231]. Other subhalo-related properties are the number of subhalos and their mean separation from the center-of-mass of the main halos [365].

- *Late-time accretion rate:* this is defined as $d \ln M_\Delta / d \ln a = -d \ln M_\Delta / dz$ by following the halo's main progenitor.

Tab. 8 presents a summary of the trends with the properties listed above which have been reported in the literature. We split the trends into those for low-mass and high-mass halos, distinguished through the characteristic mass $M_\star(z)$ of halos collapsing at redshift z, defined through

$$\sigma(M_\star, z) = \delta_{\rm cr} . \tag{4.58}$$

We focus here on the trends in the linear bias b_1; only Refs. [302, 370, 371] have measured assembly bias in higher-order LIMD bias parameters. They found the same qualitative trends also in the higher-order biases. Most results presented in Tab. 8 are for spherical-overdensity identified halos with $\Delta = 200$; those of [369] are for FoF halos. Recently, Ref. [372] explored the dependence of assembly bias on the halo definition employed, and found that, while assembly bias can be mitigated by a judicious choice, no single halo definition can absorb all assembly bias effects. Clearly, the trends shown in Tab. 8 are only a subset of all possible correlations of bias with halo properties. For example, Ref. [373] very recently showed a significant trend with the distance to the nearest more massive halo ("neighbor bias").

Despite the fact that the reported results in the literature are in good agreement overall, an important caveat to the comparison of different results is that they do not strictly have to agree, since the bias has been estimated using different approaches and different scales. Refs. [362, 363, 231, 367, 369, 365] used the halo correlation function on scales in the range $10 - 60 \, h^{-1} \, {\rm Mpc}$, while [302] used moments. None of these references have converted their results to the bias parameters strictly in the large-scale limit. Ref. [230] employed the halo-matter cross-power spectrum on scales $k \lesssim 0.1 \, h \, {\rm Mpc}^{-1}$, which, following the discussion in Sec. 4.1, should fairly accurately measure the large-scale linear bias. Finally, Refs. [370, 371] used the response approach based on separate-universe simulations. These strictly recover the large-scale bias parameters. Note that Ref. [374] recently reported strongly scale-dependent features of assembly bias in the halo correlation function, although these are restricted to very small scales $r \lesssim 10 \, h^{-1} \, {\rm Mpc}$.
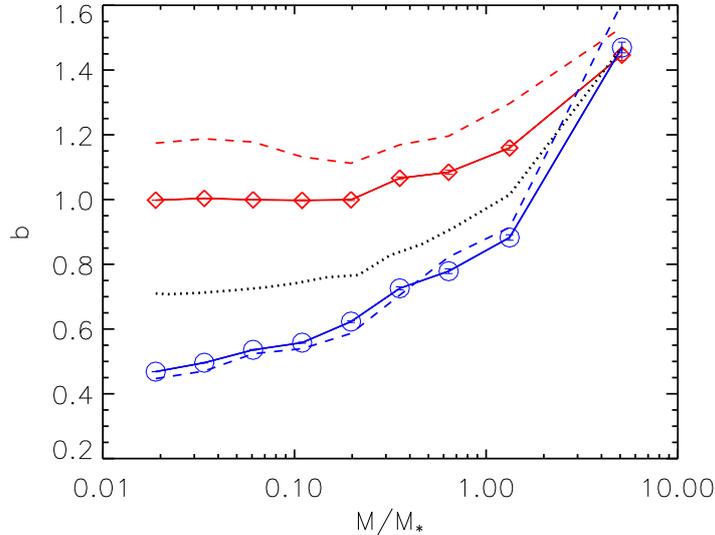
Figure 19: Bias of halos identified at $z = 0$, as a function of mass and formation redshift, i.e. the redshift when the most massive progenitor halo reaches $1/2$ of the final halo mass. The dotted black curve represents the mean bias in a given mass bin, whereas the solid red and blue curve show results for the 20 percent oldest and youngest halos, respectively. Similarly, the red and blue dashed curves refer to the upper and lower 10th percentile. Note that the halo mass is in units of $M_\star$ [Eq. (4.58)]. From [362].

For low-mass halos, late-forming (younger) halos have higher concentration, but lower bias than the early-forming (older), more concentrated halos. This trend for assembly bias with respect to concentration reverses for high-mass halos. This reversal has so far not been conclusively detected for the assembly bias with respect to formation time. The trend with concentration at high masses can be explained within the excursion set approach, as discussed in Sec. 5.10. Note that this requires going beyond the canonical formulation of the excursion set theory based on sharp k-space filters [375, 376], in which the fluctuations that collapse to form halos are statistically independent of the larger scales, i.e. their environment.

Another interesting trend is the dependence on the slope of the Lagrangian density profile of halos. Ref. [367] found that high-mass halos with steeper (more negative) slopes are less biased than those with shallower slopes. If shallower slopes are associated with lower final concentrations, then this trend matches that found for concentrations. On the other hand, if one relates the Lagrangian slope to the late-time accretion rate of halos, where shallower (less negative) slopes correspond to higher accretion rates [376], the halo assembly bias found with respect to Lagrangian slope has the *opposite* sign of that found with respect to direct measurements of the accretion rate [371].

At low mass ($M \lesssim M_\star$), the trend with concentration reverses sign, so that the more concentrated halos are clustered more strongly. In this regime, one also observes a strong trend with formation time, with early-forming halos being more clustered. This low-mass trend is illustrated in Fig. 19 where halos of a given mass are split according to their formation redshift. The effect becomes very significant for $M \lesssim M_\star$. Refs. [377, 378, 379] proposed that the assembly bias seen at low mass originates from tidal interactions with a larger neighbor. The mass accretion onto these "harassed" or "backsplash" low-mass halos is suppressed at late time by the tidal field of massive neighboring halos. This tidal stripping significantly affects the formation history of many field halos. The detailed zoomed simulation study of [380], as well as the correlation of halo bias with the tidal field (estimated on a scale of a few halo virial radii) found by [381], support this picture.

Observationally, constraining halo assembly bias is a difficult measurement, as one needs to identify at least two observable proxies, one for halo mass as well one that correlates with halo formation time or profile

shape at fixed mass. The mean halo mass of a given sample can be constrained by gravitational lensing. Various detections have been claimed in the literature [382, 383, 384, 385], although all of them have been disputed [386, 387], due to different projection and selection effects which mimic the assembly bias signature. The upper limits placed by [386, 387] are still completely consistent with the level of assembly bias expected in the standard ΛCDM cosmology.

Finally, it is worth noting that assembly bias is an important effect in phenomenological approaches to galaxy clustering that populate halos with galaxies; the most well-known approaches, which we will review in Sec. 9.1, are halo occupation distributions and sub-halo abundance matching. This is because certain types of galaxies might preferentially reside in, say, early-forming halos within a given mass range. Then, their clustering will not follow those of all halos of this mass, but only a special subset which, due to assembly bias, has different clustering properties [388, 389, 390]. We will return to this point in Sec. 9.1, but emphasize already here that this issue does not affect the general perturbative bias expansion: since the bias parameters directly describe the statistics of galaxies, without the intermediate step of the clustering of halos at fixed mass, this approach is entirely insensitive to assembly bias.

5 The excursion-set approach to the distribution of dark matter halos

In this and the next section, we review the predictions for the bias of dark matter halos in two well-developed physical models of halo formation: the excursion-set formalism (this section) and the peak approach (Sec. 6). To set the frame and illustrate the basic physical assumptions of these models, we begin with a discussion of recent simulation studies of the connection between virialized halos and their progenitors in the initial density field (Sec. 5.1).

Next, we introduce in detail the spherical collapse solution (Sec. 5.2.1), which provides an idealized model of halo formation. Moreover, it yields a natural, typical value of the mean overdensity of the Lagrangian regions (proto-halos) that form halos at late times. This fractional overdensity $\delta_{\rm cr}$ can then be used, for example, in the simple Lagrangian thresholding model introduced in Sec. 2.1. If naïvely applied, however, as is done by Press-Schechter theory (Sec. 5.2.2), the thresholding model leads to a mass function (number density of halos per unit mass) which is not properly normalized because of the so-called cloud-in-cloud problem (Sec. 5.2.3); that is, the matter density contained in halos, when integrating over all halo masses, is a factor of 2 smaller than the actually available matter density. The excursion-set formalism (Sec. 5.3) solves the cloud-in-cloud problem by adding the *first-crossing* condition. In the subsections that follow, we discuss the predictions for the halo mass function and bias from the excursion-set formalism with a constant threshold $\delta_{\rm cr}$ (Sec. 5.4–5.8) as well as a general threshold (moving, fuzzy barrier in Sec. 5.9). Finally, in Sec. 5.10, we revisit the phenomenon of assembly bias (Sec. 4.6) in the context of the excursion-set formalism.

In Sec. 5 and Sec. 6, the analysis will always be done in Lagrangian space, and in terms of the linear density contrast $\delta_R^{(1)}$ smoothed at a radius R. In these sections, we therefore simplify the notation with respect to that used in previous sections, to conform more closely with that followed in the literature. For convenience, we list frequently used symbols in Tab. 9 (p. 111).

Before delving into the details of the standard excursion-set picture which we will describe in Sec. 5.2–5.7, let us summarize the qualitative predictions for halo bias in this picture. Since the only quantity appearing in the calculation is the smoothed linear density field $\delta_R^{(1)}(\boldsymbol{q})$, the Lagrangian halo density field can be

written as a local function of $\delta_R^{(1)}$ and its derivatives with respect to R. Specifically,

$$
\begin{aligned}
\delta_h^L(\boldsymbol{q})\Big|_{\text{exc. set}} &= F\left[\delta_R^{(1)}(\boldsymbol{q}),\, \partial\delta_R^{(1)}(\boldsymbol{q})/\partial R,\, \cdots\right] \\
&= b_1^L \delta^{(1)}(\boldsymbol{q}) + \frac{1}{2}b_2^L[\delta^{(1)}]^2(\boldsymbol{q}) + \cdots \\
&\quad + b_{\nabla^2\delta}^L \nabla^2\delta^{(1)}(\boldsymbol{q}) + b_{\delta\nabla^2\delta}^L \delta^{(1)}(\boldsymbol{q})\nabla^2\delta^{(1)}(\boldsymbol{q}) + \cdots \\
&\quad + b_{\nabla^4\delta}^L \nabla^4\delta^{(1)}(\boldsymbol{q}) + \cdots,
\end{aligned}
\tag{5.1}
$$

where here and throughout this section, we drop the time arguments, as all quantities are linearly extrapolated to the time τ_0 of identification of halos. We have used that the higher-derivative terms introduced by smoothing $\delta^{(1)} \to \delta_R^{(1)}$ can all be written as $\nabla^{2n}\delta^{(1)}$ (see Sec. 2.6), which also holds for $\partial\delta_R^{(1)}/\partial R$, leading to a Lagrangian LIMD bias expansion with a subset of higher-derivative terms. Note that no stochastic contributions are predicted by the standard excursion set, though a shot-noise term ε is often added empirically. We will generalize this version of the excursion-set picture in Sec. 5.9, where we also take into account the tidal field in Lagrangian space, and discuss stochasticity in the threshold. These lead to additional contributions to the bias expansion beyond Eq. (5.1).

5.1 General considerations about the formation of dark matter halos

Since numerical simulations give us access to the entire formation history of dark matter halos, they enable us to ascertain the validity of the approximations made in Lagrangian approaches such as excursion sets or peaks. The correspondence between halos identified at low redshift and their progenitors in the initial conditions—the so-called *proto-halos*—can be studied on an object-by-object basis. Here, proto-halos are the Lagrangian regions obtained by tracing the dark matter particles belonging to virialized halos back to their initial positions. The typical comoving size of proto-halos of mass M thus is the Lagrangian radius $R(M)$ which encloses M at the mean comoving density $\overline{\rho}_m$.

In the excursion-set formalism, proto-halos are modeled as regions where the smoothed linear matter density on the scale $R(M)$ is above a threshold δ_{cr}. Moreover, any regions which are included in a region above threshold identified at a larger smoothing scale are excluded. On the other hand, the peak approach models proto-halos as local maxima of the linear density smoothed on the scale $R(M)$.

Ref. [391] were the first to investigate the robustness of the correspondence between peaks in the initial density field and proto-halos. On analyzing 32^3-particle N-body simulations of fairly small box size ($L \sim 14\,h^{-1}\,\text{Mpc}$), they concluded that the resolved dark matter halos form preferentially around high peaks of the initial density field. However, Ref. [392] used simulations evolving 144^3 particles in a much larger box ($L = 50\,h^{-1}\,\text{Mpc}$) and found a rather weak association between initial density peaks and collapsing halos. Their conclusions were confirmed partially by [393], who studied the properties of proto-halos in N-body simulations of 256^3 particles in a box of size $L \simeq 85\,h^{-1}\,\text{Mpc}$. The latter concluded that nearly half of the galactic-sized proto-halos do not contain a linear density peak within their Lagrangian volume.

Recently, Ref. [394] revisited this issue using two high-resolution 1024^3 N-body simulations in boxes of size 1200 and $150\,h^{-1}\,\text{Mpc}$, respectively. They found that essentially all the proto-halos trace Lagrangian patches with a linearly extrapolated density contrast $\delta_R^{(1)} \gtrsim \mathcal{O}(1)$. In addition, as many as $\sim 70\%$ of all halos identified with an FoF algorithm (with $N_{\text{FoF}} > 100$) in both of their simulations can be properly identified with a peak in the linear density field (i.e. the peak resides in the Lagrangian halo patch) when smoothed on mass scales $0.7M \leq M_{\text{pk}} \leq 1.3M$ (Fig. 20). Here, M is the mass of the actual halo, given by the number of bound dark matter particles, while M_{pk} is the *peak mass*, i.e. the mass enclosed in the filter. The fraction of peak associations increases to 85% upon extending the search to the mass range $0.5M \leq M_{\text{pk}} \leq 2M$. Furthermore, the fraction depends strongly on halo mass, with as many as $\sim 91\%$ of halos with $M > 5 \times 10^{14}\,h^{-1}\,M_\odot$ forming in the vicinity of peaks of the expected characteristic mass.

However, a non-negligible portion of halos show a considerable disparity between the predicted and measured masses. For example, $\sim 20\%$ of halos with > 100 particles have $M_{\text{pk}} < M$, 15% of which do not contain any peaks on any mass scale within a factor of four of the true halo mass scale. The authors refer to
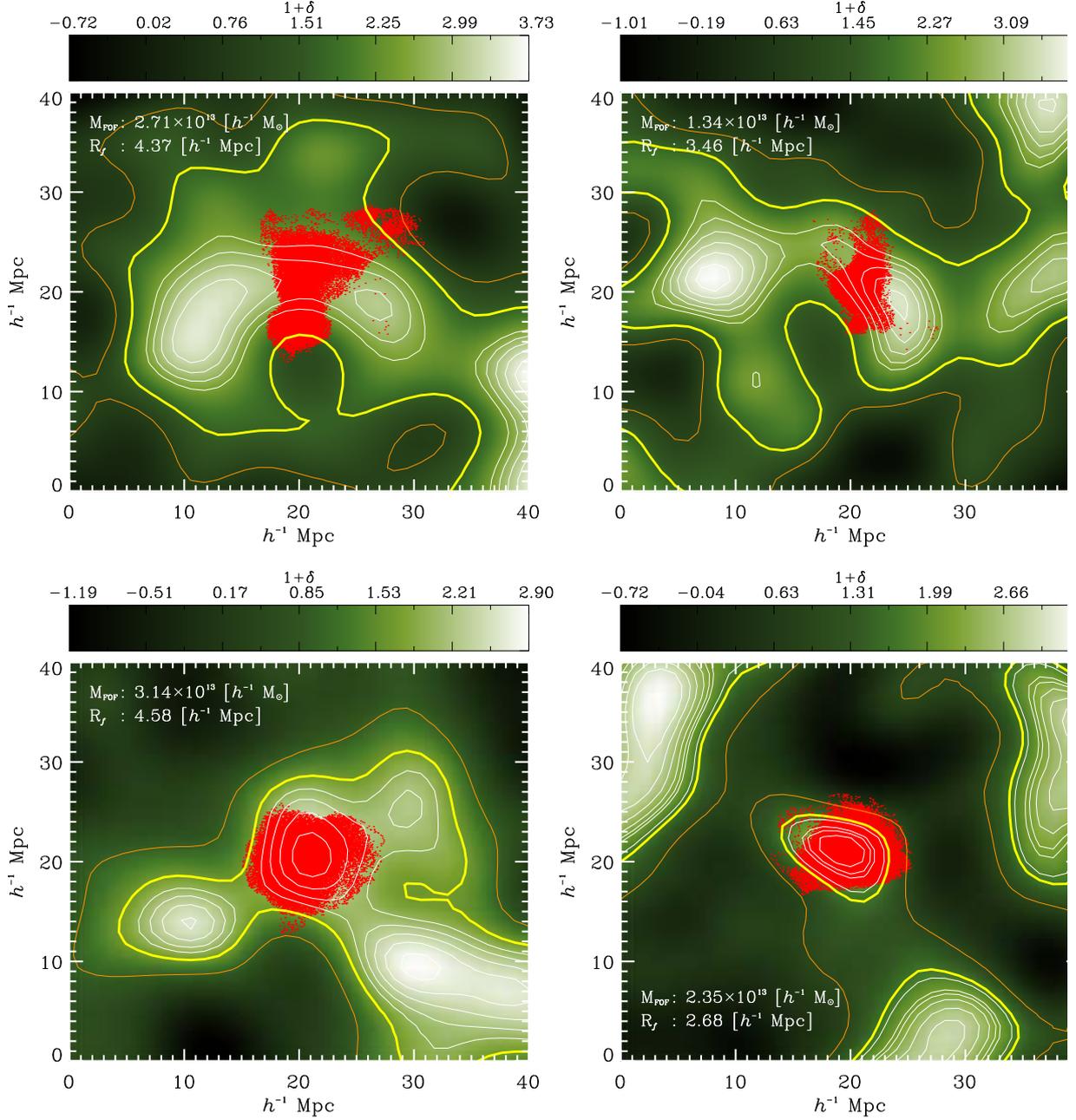
Figure 20: Examples of the overdensity field in the vicinity of four proto-halos. The top panels display results for halos with $M_{\rm pk} < M_{\rm halo}/4$, where $M_{\rm pk}$ is the mass enclosed within the filter radius R at which a peak was found. The bottom panels show results for two similar mass halos, yet with $M_{\rm pk} \approx M_{\rm halo}$. Linearly extrapolated density fields have been smoothed with a tophat filter with $R = R(M_{\rm halo})$. Particles that belong to the FoF halos at $z = 0$ are shown as red dots. In all panels, contours highlight the density gradients in the neighborhood of the halo. A density contrast $\delta_R = 1$ is shown as an orange curve, while the threshold for spherical collapse, $\delta_R = \delta_{\rm cr} = 1.686$, is shown as a thick yellow line. From [394].

107

these as "peakless" halos (see the upper panels in Fig. 20). The increasing fraction of "peakless" halos with decreasing mass reflects the fact that the approximation of an isolated spherical collapse of a perturbation centered around a density peak becomes increasingly inaccurate as the halo mass decreases.

These numerical findings help us in understanding the validity of the various approximations made in Lagrangian bias models. The standard Press-Schechter and excursion-set theory discussed in Sec. 5.2 – Sec. 5.8 rely on two key assumptions:

- The collapse of dark matter halos is approximately spherical.

- Proto-halo patches are characterized by a linearly extrapolated overdensity $\delta_R^{(1)}$ equal to the spherical collapse threshold $\delta_{\rm cr}$, Eq. (5.9).

In light of the numerical studies reviewed above, these assumptions are expected to be valid at high mass $M \gg M_\star$ only. Two important additional ingredients can be taken into account in order to improve the accuracy of Lagrangian bias models:

- The shear field, which is expected to be partly responsible for both the "peakless" halos and the mass dependence of the critical density threshold. In a first approximation, the effect of shear can be modeled as a moving barrier as discussed in Sec. 5.9.

- The peak constraint, i.e. the fact that halos collapse around local maxima in the initial density field. This extension is the focus of Sec. 6.

The peak constraint resolves a fundamental limitation of the excursion-set formalism, which stems from the assumption that *the total mass fraction enclosed in halos with mass greater than M is the same as the total volume fraction of above-threshold smoothed Lagrangian regions,* with filter size $R = R(M)$. For this assumption to hold, the *smoothed* density of all points inside halos must exceed the threshold. This is not guaranteed for actual halos. In particular, depending on the density profile, the outer part of halos can have values of the smoothed density that are significantly below the threshold (see Fig. 21 for an illustration). Indeed, the comparison between the halo mass function calculated from the excursion-set formalism with constant threshold and N-body simulations shows that the theory underpredicts the abundance of massive halos [395, 318, 396]. Moreover, this assumption is the very foundation of the excursion-set formalism. The two adjustable components in the excursion-set formalism, the scale dependence and fuzziness of the collapse threshold, are not directly related to the above-mentioned deficiency. This problem is addressed in the peak formalism (Sec. 6) where we define the *center* of halos as peaks of the density field. Hence, we do not need to rely on the volume fraction in order to calculate the halo abundance. Recently, the virtues of both approaches have been combined in the *excursion-set peaks* ansatz, which we will discuss in Sec. 6.8.

5.2   From Press-Schechter to excursion sets

5.2.1   The spherical collapse model of halo formation

Let us consider a spherical region with uniform density of initial radius R_i and initial average density contrast Δ_i in an expanding Universe [108, 397] with mean comoving matter density at the initial time $\overline{\rho}_{m,i} \equiv \overline{\rho}_m(t_i)$. That is, the total mass of the overdensity is $M = (4\pi/3)R_i^3 \overline{\rho}_{m,i}(1 + \Delta_i)$. For simplicity, we only present the case of an Einstein de-Sitter (EdS) Universe, although the effects of cosmological constant as well as curvature of the Universe can easily be incorporated [398, 376].

The interior of a spherically symmetric perturbation embedded in an FRW spacetime evolves independently of the surrounding spacetime and is only determined by the total mass enclosed [399, 400, 401]. The equation of motion of the radius R of the spherical region is given by

$$\frac{d^2 R}{dt^2} = -\frac{GM(<R)}{R^2} = -\frac{4\pi G}{3}\overline{\rho}_{m,i}(1 + \Delta_i)\frac{R_i^3}{R^2} = -\frac{H_i^2 R_i^3}{2R^2}(1 + \Delta_i), \tag{5.2}$$

where $H_i = 2/(3t_i)$ is the Hubble rate at the initial time t_i. It is worth noting that Eq. (5.2) holds for arbitrarily large-scale spherical perturbations; i.e. we do not need to assume $R \ll H^{-1}$. Integrating once
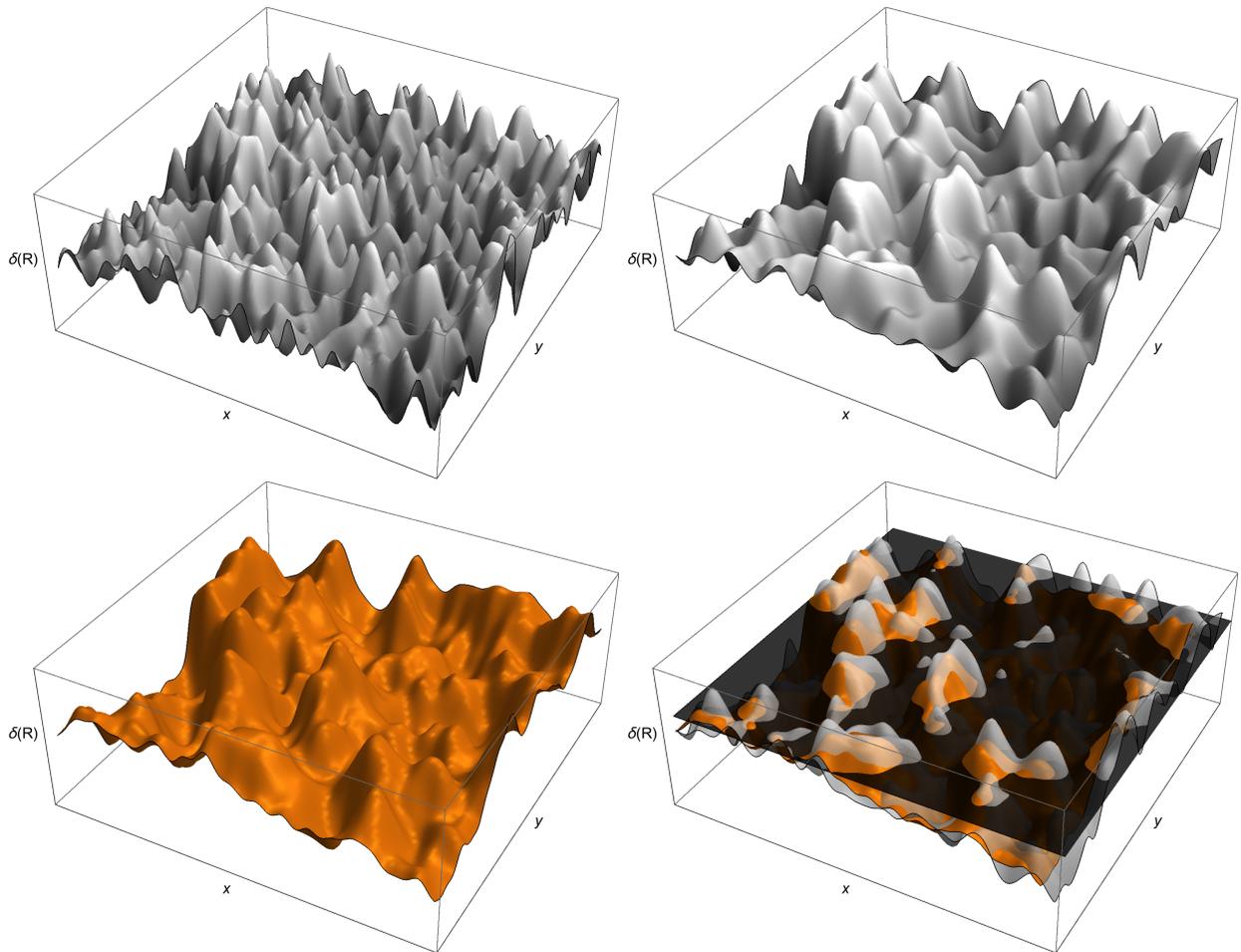
Figure 21: Illustration of a two-dimensional Gaussian density field $\delta(R)$ smoothed on different scales R: $0.4\,h^{-1}\,\mathrm{Mpc}$ (*top left*), $1\,h^{-1}\,\mathrm{Mpc}$ (*top right*), $2\,h^{-1}\,\mathrm{Mpc}$ (*bottom left*). The side length is $50\,h^{-1}\,\mathrm{Mpc}$. In the *bottom-right* panel, we superimpose the fields with $R = 1\,h^{-1}\,\mathrm{Mpc}$ and $R = 2\,h^{-1}\,\mathrm{Mpc}$, along with the spherical collapse threshold ($\delta_{\mathrm{cr}} = 1.686$, black plane). In the excursion-set approach, the mass fraction of halos is set equal to the fraction of area above the threshold at the smoothing scale $R(M)$. Note that there are regions which are above threshold at the larger smoothing scale $R = 2\,h^{-1}\,\mathrm{Mpc}$ but below the threshold at $R = 1\,h^{-1}\,\mathrm{Mpc}$. This illustrates the so-called *cloud-in-cloud* problem that is discussed in Sec. 5.2.3.

over time, we obtain the energy conservation equation:

$$\frac{1}{2}\dot{R}^2 - \frac{H_i^2 R_i^3}{2R}(1 + \Delta_i) = \mathrm{constant} \equiv E \, . \tag{5.3}$$

The spherical region is unbound if $E > 0$, i.e. if the initial expanding velocity exceeds $\dot{R}_i \geq H_i R_i \sqrt{1 + \Delta_i}$, but it will eventually turn around and re-collapse if $E < 0$. If $\Delta_i \ll 1$, then the initial velocity can be obtained by matching the linear adiabatic growing mode, $\dot{R}_i = H_i R_i (1 - \Delta_i/3)$. The total energy then becomes

$$E = -\frac{5}{3}\frac{(H_i R_i)^2}{2}\Delta_i, \tag{5.4}$$

and the turn around, or maximum, radius is obtained by setting $\dot{R} = 0$ and solving Eq. (5.3) for R:

$$R_{\mathrm{ta}} = \frac{3}{5}\left(\frac{1 + \Delta_i}{\Delta_i}\right)R_i \, . \tag{5.5}$$

109

The exact solution of Eq. (5.2) is given by a cycloid, which can be parameterized with a variable θ:

$$R = \frac{R_{\text{ta}}}{2}\left(1 - \cos\theta\right), \; t = \frac{t_{\text{ta}}}{\pi}(\theta - \sin\theta), \tag{5.6}$$

where t_{ta} is the time at turn-around ($\theta = \pi$). Noting the similarity between the case at hand and the Keplerian radial orbit (with semi-major axis $R = R_{\text{ta}}/2$ and period $T = 2t_{\text{ta}}$), we use Kepler's third law to relate R_{ta} and t_{ta} as $8GMt_{\text{ta}}^2 = \pi^2 R_{\text{ta}}^3$.

The parametric solution yields a relation between θ and the overdensity Δ,

$$1 + \Delta = \frac{\bar{\rho}_{m,i}(1 + \Delta_i)R_i^3}{\bar{\rho}_m R^3} = (1 + \Delta_i)\left(\frac{t}{t_i}\right)^2\left(\frac{R_i}{R}\right)^3 = \frac{9}{2}\frac{(\theta - \sin\theta)^2}{(1 - \cos\theta)^3}. \tag{5.7}$$

At early times ($\theta \simeq \theta_i$), we reproduce the linear growing modes $\Delta(t) \simeq 3\theta^2/20 \simeq \Delta_i(t/t_i)^{2/3} \propto a(t)$ in the EdS Universe. Moreover, the coefficients of the higher-order terms in the expansion of $\Delta(t)$ in powers of $a(t)$ are precisely the coefficients of the operators δ^n in the n-th order contribution to the density field in perturbation theory (e.g., $17/21$ at second order, $341/567$ at third order, and so on). Then, the spherical region evolves nonlinearly as time proceeds, reaching $1 + \Delta_{\text{ta}} \simeq 9\pi^2/16 \simeq 5.55$ at turn-around time ($\theta = \pi$). The spherical collapse model predicts the collapse to a singularity $R = 0$ at $\theta = 2\pi$. This is of course unphysical. Instead, gradient instabilities, which occur as soon as mass shells cross in physical space, will break spherical symmetry and lead to a complex bound structure. We can estimate the final radius and density of this structure (halo) assuming perfect virialization. The virial theorem then states that $E = T + W = W/2$, where T is the kinetic energy while W is the potential energy. Since $W = E$ holds at turn-around, we obtain for the radius of the virialized halo, $R_{\text{vir}} = R_{\text{ta}}/2$. Further assuming that virialization is completed precisely at $\theta = 2\pi$, the formal collapse epoch, we obtain for the density of the virialized halo

$$1 + \Delta_{\text{vir}} = (1 + \Delta_{\text{ta}})\left(\frac{\bar{\rho}_{m,\text{ta}}}{\bar{\rho}_{m,\text{vir}}}\right)\left(\frac{R_{\text{ta}}}{R_{\text{vir}}}\right)^3 = (1 + \Delta_{\text{ta}})\left(\frac{t_{\text{ta}}}{t_{\text{vir}}}\right)^2\left(\frac{R_{\text{ta}}}{R_{\text{vir}}}\right)^3 = 18\pi^2 \simeq 178. \tag{5.8}$$

The corresponding linear density contrast at the time of virialization (collapse), which defines the critical density δ_{cr}, is

$$\delta_{\text{cr}} \equiv \Delta^{(1)}(t_{\text{vir}}) = \Delta_i\left(\frac{t_{\text{vir}}}{t_i}\right)^{2/3} = \frac{3}{5}\left(\frac{3}{4}\right)^{2/3}(\theta_{\text{vir}} - \sin\theta_{\text{vir}})^{2/3} = \frac{3}{5}\left(\frac{3\pi}{2}\right)^{2/3} \simeq 1.686 . \tag{5.9}$$

There is some arbitrariness in the definition of the virialization epoch. Analytic arguments [402] and N-body simulations [403] show that the virialization process is not instantaneous. Furthermore, halos are often identified with threshold values Δ_{vir} different from 178 [313]. Therefore, the corresponding linear density contrast δ_{cr} also depends on the halo identification procedure. Nevertheless, in the high-mass limit where the spherical collapse assumption works best, Ref. [404] argues that a threshold $\delta_{\text{cr}} = \mathcal{F}^{-1}(\Delta_{\text{vir}})$, where $\mathcal{F}$ is the spherical collapse mapping from linear to nonlinear densities, should be used for halos identified with a nonlinear threshold Δ_{vir}. For $\Delta_{\text{vir}} = 200$, this rare-event limit yields $\delta_{\text{cr}} \simeq 1.59$. While this prescription is somewhat supported by N-body simulations [405], we will assume Eq. (5.9) as default value here.

As the linear growth factor is a monotonically increasing function of time, one can use it as a clock to measure the virialization time; thus, a spherical region of radius R will be virialized and form a halo when the linearly extrapolated density contrast exceeds δ_{cr}. This motivates the Press-Schechter theory that estimates the abundance of halos by interpreting the fraction of thresholded ($\delta_R^{(1)} > \delta_{\text{cr}}$) Lagrangian volume, following Sec. 2.1, as the fraction of mass contained in halos with mass greater than $M = (4\pi/3)\bar{\rho}_m R^3$.

Clearly, the spherical collapse is only a very rough toy model for how gravitational collapse to bound structures actually happens. In particular, tidal fields, angular momentum, as well as the effect of small-scale perturbations are entirely neglected. Models in which halos form through ellipsoidal collapse provide a better description of the formation times of dark matter halos [406, 403]. Therefore, δ_{cr} should be regarded

Quantity	Symbol
Smoothed linear density contrast	$\delta(R) \equiv \delta_R^{(1)}(\boldsymbol{q})$
Variance of the smoothed density field	$S \equiv \left\langle [\delta(R)]^2 \right\rangle$
Cumulative collapsed mass fraction	$F(>M)$
Differential volume fraction	$f(M) = -dF(>M)/dM$
Multiplicity function	$f(\nu_c) = f(M)\,|dM/d\nu_c|$
Differential survival probabilities of random walks:	
from $(\infty, 0)$ to (R, δ)	$\Pi(\delta; R)$
from (R_ℓ, δ_ℓ) to (R, δ)	$\Pi[(\delta; R), (\delta_\ell; R_\ell)]$
Conditional survival probability for a walk to arrive at (R, δ),	
given constraint that $\delta(R_\ell) = \delta_\ell$	$\Pi(\delta; R | \delta_\ell; R_\ell)$

Table 9: List of symbols used throughout Sec. 5.

as an idealized reference value, which we will later generalize to a scale-dependent stochastic barrier $B(\sigma)$ (see Sec. 5.9). In practice, since there is some freedom in the definition of a virialized dark matter halos (see Appendix D), the value of δ_{cr} effectively depends on the halo identification algorithm and, thus, can deviate from 1.686. This effective threshold can be measured by extrapolating the linear overdensity of the Lagrangian regions that collapse to form halos to the limit $M \to \infty$ (see Fig. 26 on p. 132).

Note that a similar calculation can be done for the formation of cosmic voids [407]. In particular, for very underdense spherical regions with a linearly extrapolated fractional underdensity $\delta_\ell^{(1)} \lesssim -1$, Ref. [408] has found that $\delta_\ell(\tau) = (1 - \delta_\ell^{(1)}(\tau)/1.5)^{1.5}$ is a good approximation to the interior density, which might be useful to model the bias relation in cosmic voids along with the arguments in Sec. 2.1 [409].

5.2.2 Press-Schechter formalism

We now recap the original derivation of the Press-Schechter mass function [217] (see also Sec. 2.1). The key assumptions are

- The linear matter density contrast $\delta^{(1)}$, whence the smoothed density contrast $\delta_R^{(1)}$, follows Gaussian statistics. One must relax this assumption in the presence of primordial non-Gaussianity (see Sec. 7.4.2).

- A Lagrangian region of volume $V = (4\pi/3)R^3$ collapses to form a halo of mass $M = \bar{\rho}_m V$ when its density contrast $\delta_R^{(1)}$ (linearly extrapolated to the epoch of interest) exceeds the spherical collapse threshold δ_{cr}.

Given the second assumption, the fraction of Lagrangian volume belonging to halos of mass *greater than M* is [cf. Eq. (2.5)]

$$p_G(\delta_R^{(1)} > \delta_{\mathrm{cr}}) = \frac{1}{\sqrt{2\pi\sigma^2(R)}} \int_{\delta_{\mathrm{cr}}}^{\infty} d\delta \, \exp\left[-\frac{1}{2}\frac{\delta^2}{\sigma^2(R)}\right] = \frac{1}{2}\mathrm{erfc}\left[\frac{\delta_{\mathrm{cr}}}{\sqrt{2}\sigma(R)}\right] . \tag{5.10}$$

Here, p_G stands for the Gaussian cumulative PDF, and erfc denotes the complementary error function defined in Tab. 2 on p. 10. R is the Lagrangian radius corresponding to the halo mass M,

$$M = 3.1389 \times 10^{11} h^{-1} M_\odot \left(\frac{\Omega_m}{0.27}\right)\left(\frac{R}{h^{-1}\mathrm{Mpc}}\right)^3 , \tag{5.11}$$

and $\sigma(R)$ is the root-mean-square of the linear smoothed density perturbations on the scale R,

$$S \equiv \sigma^2(R) \equiv \left\langle \left(\delta_R^{(1)}\right)^2 \right\rangle = \int_{\boldsymbol{k}} P_{\mathrm{L}}(k) W_R^2(k) , \tag{5.12}$$

111

where $W_R(k)$ denotes the filtering kernel in Fourier space. Note that, in Eq. (5.10), the smoothing scale R is held fixed although the volume fraction is an integral over contributions from halos with Lagrangian radius greater than R. This will be dealt with by taking a derivative with respect to mass.

Given the fundamental assumption of the excursion set stated in Sec. 5.1, the Lagrangian volume fraction $F(> M)$ of the total mass enclosed in halos of mass greater than M follows immediately from Eq. (5.10):

$$F(> M) = \frac{1}{\overline{\rho}_m} \int_M^\infty d\ln M' \, M' \overline{n}_h(M') = p_G(\delta_R^{(1)} > \delta_{\rm cr}) \,. \tag{5.13}$$

Therefore, the differential volume fraction in halos of mass M is

$$f(M) \equiv -\frac{dF(> M)}{dM} \,, \tag{5.14}$$

which formally leads to the halo mass function, i.e. the comoving number density of halos per logarithmic mass interval,

$$\overline{n}_h(M) \equiv \frac{d^2 \overline{N}_h}{dV \, d\ln M} = \overline{\rho}_m f(M) = -\overline{\rho}_m \frac{dF(> M)}{dM} \,. \tag{5.15}$$

Unfortunately, this mass function does not appear to be properly normalized, as an integration over all the mass included in halos only recovers half of the total mass:

$$\int_0^\infty d\ln M \, M \, \overline{n}_h(M) = -\overline{\rho}_m \int_0^\infty dM \frac{dF}{dM} = -\overline{\rho}_m \Big[p_G(R = \infty) - p_G(R = 0) \Big] = \frac{1}{2}\overline{\rho}_m \,. \tag{5.16}$$

On the other hand, the basic premise of this argument states that, when including arbitrarily small halo masses, the mass in halos should add to the total matter density, since

$$\lim_{R \to 0} \sigma(R) = \infty \quad \text{and} \quad \lim_{R \to \infty} \sigma(R) = 0 \,, \tag{5.17}$$

so that there are no regions that do not collapse to halos of some (arbitrarily small) mass. Note that we have assumed a hierarchical density field without any cutoff here. To resolve this issue, Ref. [217] introduced an *ad hoc*, "fudge" factor of two which leads to the *Press-Schechter* (PS) mass function:

$$\overline{n}_{h,\rm PS}(M) = \overline{\rho}_m \left(-2\frac{dF}{dM} \right) = \frac{\overline{\rho}_m}{M} \sqrt{\frac{2}{\pi}} \nu_c e^{-\nu_c^2/2} \left| \frac{d\ln \sigma(R)}{d\ln M} \right| \,, \tag{5.18}$$

where $\nu_c \equiv \delta_{\rm cr}/\sigma(R)$ is the significance of the critical density in terms of the standard deviation of matter fluctuations on the scale R. Massive halos, for which $\nu_c \gg 1$, are rare, while low-mass halos such that $\nu_c \ll 1$ are common. It is convenient to parametrize the halo mass function with a *multiplicity function* $\nu_c f(\nu_c)$ defined as

$$\nu_c f(\nu_c) \frac{d\nu_c}{\nu_c} = \frac{M}{\overline{\rho}_m} \overline{n}_h(M) \frac{dM}{M} \,. \tag{5.19}$$

The notation $d\nu_c$ implicitly signifies that σ is varied while $\delta_{\rm cr}$ is kept fixed. As we shall see in Sec. 5.3, this multiplicity function corresponds to the first-crossing distribution of random walks in the excursion-set theory. Hence, the halo mass function can be recast into the generic form

$$\overline{n}_h(M) = \frac{\overline{\rho}_m}{M} \nu_c f(\nu_c) \left| \frac{d\ln \sigma(R)}{d\ln M} \right| \,. \tag{5.20}$$

Note that the Jacobian often appears as $d\ln \nu_c / d\ln M$ in the literature. Still, one should keep in mind that σ is varied while $\delta_{\rm cr}$ is held fixed. In Press-Schechter theory, the multiplicity function is

$$\nu_c f_{\rm PS}(\nu_c) = \sqrt{\frac{2}{\pi}} \nu_c e^{-\nu_c^2/2} \,. \tag{5.21}$$

Eq. (5.20) is frequently referred to as *universal mass function* (we have already encountered this form in Sec. 3.3), since it only depends on redshift, cosmological parameters, and the power spectrum of initial fluctuations through $\nu_c = \delta_{\rm cr}/\sigma(M)$ [16, 395]. As discussed in Sec. 3.3, in an EdS universe with exact power-law initial conditions, the mass function can always be written in this form. In a more general cosmology, for example a realistic ΛCDM model, the multiplicity function will also depend on variables other than the peak significance ν_c. For example, the mass function of density peaks (see Sec. 6) cannot be written in the form Eq. (5.20). Still, the cosmology dependence of the peak multiplicity function is captured by a small set of moments $\sigma_n(R)$.

5.2.3 Cloud-in-cloud problem and its resolutions*

Despite an unsatisfactory treatment of the normalization, the Press-Schechter formalism captures the qualitative features of the halo mass function measured in N-body simulations (a power-law function with an exponential cutoff; see Fig. 9). As such, it provides a baseline for more detailed models of the halo mass function. In this section, we review the origin of the factor of two difference in the normalization of the Press-Schechter mass function, and the attempts to resolve the issue.

Refs. [410, 375] explained the underlying reason for the "fudge factor" in the Press-Schechter formalism. Namely, the Press-Schechter approach does not take into account the possibility that an underdense region identified on a scale R may be embedded in a halo on a bigger scale $R' > R$ (see the derivation of Eq. (5.15) above). This is the so-called *cloud-in-cloud* problem first identified in [411] and named in [410].

To properly address this problem, we have to take into account the underdensities $\delta_R^{(1)} < \delta_{\rm cr}$ that belong to a collapsed region with larger radius $R' > R$. That is, the fraction of Lagrangian volume belonging to collapsed objects with mass greater than M, $F(>M)$, should be written as

$$F(>M) = p_G(\delta_R^{(1)} > \delta_{\rm cr}) + \int_{-\infty}^{\delta_{\rm cr}} d\delta \frac{dp_G}{d\delta} p_{up}\left(\delta(r_{>R}) > \delta_{\rm cr}; \delta\right), \tag{5.22}$$

instead of Eq. (5.10), which includes only the first term. Here, $\delta(r_{>R})$ denotes the linear density contrast smoothed over some radius r greater than R, whereas the first-crossing probability p_{up} (subscript up because the first crossing always happens upward) in the second term of Eq. (5.22) is the probability that regions with $\delta_R^{(1)} < \delta_{\rm cr}$ belong to a collapsed object with larger radius. In other words, it is the probability that there exists $R' > R$ such that $\delta_{R'}^{(1)} > \delta_{\rm cr}$. Importantly, Eq. (5.22) is, by construction, properly normalized as $p_{up} \to 1$ in the small-scale limit $R \to 0$ (where $\sigma(R) \to \infty$). This means that, for a sufficiently small length scale, virtually all underdense regions are parts of collapsed objects of bigger size. Note, however, that it is still not obvious why the second term on the right-hand side of Eq. (5.22) (the probability that an underdense region is part of a bigger collapsed object) is exactly the same as the first term (the probability that the density contrast of a given region exceeds critical value $\delta_{\rm cr}$).

Alternatively, Ref. [412] formulates the cloud-in-cloud problem as a lack of distinction between sub-halos, i.e. collapsed regions (with size R) enclosed in a bigger collapsed region (with size R'), and *isolated* collapsed regions. In their formulation, the cloud-in-cloud problem can be resolved upon introducing a conditional probability function $p(M, M')$, which quantifies the probability that an overdense region of mass M belongs to a larger isolated region of mass $M'(\geq M)$. This can also be interpreted as a probability distribution of substructure of mass M inside a larger, isolated virialized object of mass M'. Thus, Eq. (5.13) becomes

$$F(>M) = \frac{1}{\bar{\rho}_m} \int_M^{\infty} dM' \bar{n}_h(M') p(M, M'). \tag{5.23}$$

Clearly, the function $p(M, M')$ satisfies $p(M, M') = 0$ when $M > M'$, and $\lim_{M' \to M^+} p(M, M') = 1$ by

---

* This section is of a more technical nature and a slight digress. Readers interested in the excursion set itself may go directly to Sec. 5.3.

construction. In addition, the $M \to 0$ limit of Eq. (5.23) yields [see Eq. (5.13) and Eq. (5.16)]:

$$F(> 0) = \frac{1}{\overline{\rho}_m} \int_0^\infty dM' \overline{n}_h(M') p(M \to 0, M') = \frac{1}{2} \ . \tag{5.24}$$

This implies $p(M \to 0, M') \to 1/2$ in order to ensure that the halo mass function be correctly normalized. That is, only half of the small scale ($M \ll M'$) regions in the collapsed halos of mass M' are dense enough to form a sub-halo. This follows from the fact that, on extremely small scales ($M \to 0$), the threshold density contrast $\delta_{\rm cr}$ can be safely ignored compared to the standard deviation of matter perturbations: $\delta_{\rm cr} \ll \sqrt{\sigma_M^2 - \sigma_{M'}^2}$. Comparing Eq. (5.23) to Eq. (5.13), however, we find that the Press-Schechter case corresponds to setting $p(M, M') = 1$ for all $M' > M$, which means that the Press-Schechter formalism implicitly assumes that every collapsed region is part of a bigger isolated region. This fallacious assumption also leads to an error in the normalization, as $p(M, M')$ has the wrong asymptote in the limit $M \ll M'$. Finally, in analogy with the step following Eq. (5.13), we take the mass derivative of Eq. (5.23) which leads to an integro-differential equation,

$$\frac{dF(> M)}{dM} = -\frac{1}{\overline{\rho}_m} \overline{n}_h(M) + \frac{1}{\overline{\rho}_m} \int_M^\infty dM' \overline{n}_h(M') \frac{dp(M, M')}{dM}, \tag{5.25}$$

which allows us to solve for $\overline{n}_h(M)$ for a given function $dp(M, M')/dM$.

In both formulations of the cloud-in-cloud problem, new functions $p_{up}(\delta(r_{>R}) > \delta_{\rm cr}; \delta)$ and $p(M, M')$ are introduced to describe the statistics of small-scale regions (either over-dense or under-dense) in conjunction with the large-scale collapsed regions. The excursion-set formalism provides a systematic way of calculating those functions given the statistical properties of the underlying matter density field. In fact, as we shall see shortly, the correct normalization of the Press-Schechter mass function can be derived within an excursion-set approach based on the sharp-k filter, in which fluctuations on different smoothing scales are completely uncorrelated.

5.3 Excursion-set formalism: setting up the scene

The basic building block of the excursion-set formalism is the smoothed, linearly evolved initial density field at a point $\boldsymbol{q}$:

$$\delta(R) \equiv \delta_R^{(1)}(\boldsymbol{q}) = \int d^3x \, W_R(|\boldsymbol{x}|) \delta^{(1)}(\boldsymbol{q} + \boldsymbol{x}) \ . \tag{5.26}$$

Since we always center the filter at the same Lagrangian position $\boldsymbol{q}$, regardless of the value of R, and we always deal with the linear density field, we will omit the explicit dependence on $\boldsymbol{q}$ and the superscript (1) to simply write $\delta(R) \equiv \delta_R^{(1)}(\boldsymbol{q})$ in this section (see Tab. 9). Frequently used filtering kernels $W_R(x)$ and their Fourier transform can be found in Appendix A.2. Note that the shape of the filter in real space reflects the Lagrangian density profile of halos. In this regard, the sharp-k filter is not realistic, as the corresponding real-space filtering kernel $W_R(x)$ is not always positive.

The statistical properties of the smoothed density field $\delta(R)$ are entirely specified by its correlation functions. In particular, when the initial density contrast follows Gaussian statistics, the linear matter power spectrum extrapolated to the present epoch $P_{\rm L}(k)$ encodes the entire statistics (Appendix A.3).

We are now ready to formulate the excursion-set formalism for halo statistics. The key assumption is highlighted in the following box:

At a given redshift z, a Lagrangian point $\boldsymbol{q}$ belongs to a halo of size R if R is the *maximum smoothing scale* at which the smoothed linear density contrast $\delta(R)$ exceeds the critical density contrast $\delta_{\rm cr}(z)$.

The spherical collapse model (Sec. 5.2.1) provides the connection between the extrapolated density field $\delta(R)$ and dark matter halos. Appealing to the assumption of statistical homogeneity of the Universe and the ergodicity of the cosmic density field, the probability that a random Lagrangian point $\boldsymbol{q}$ belongs to a
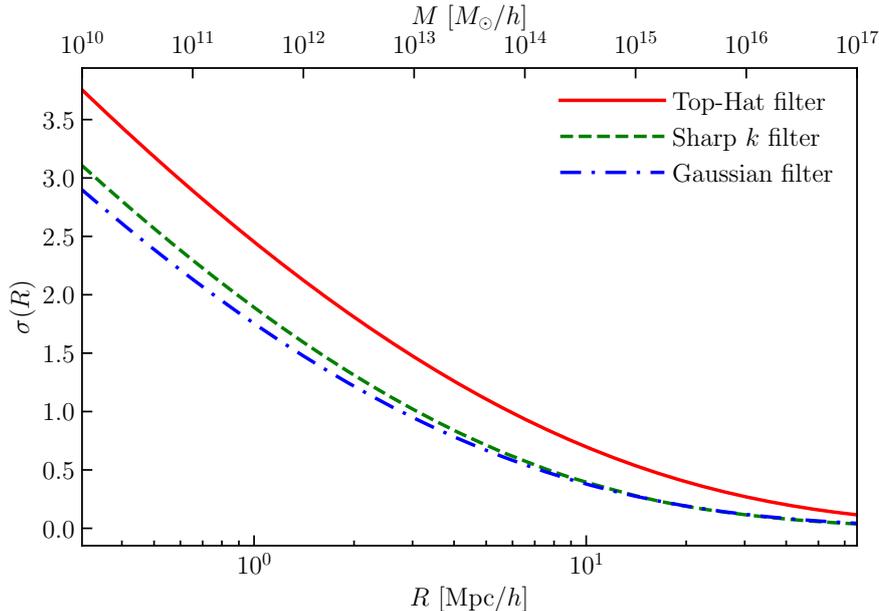
Figure 22: Relation between the halo mass (M), Lagrangian radius (R) and the r.m.s. density fluctuation smoothed on the scale R, $\sigma(R)$, at $z = 0$ for three different filters: tophat filter (red solid line), Sharp-k filter (green dashed line), and Gaussian filter (blue dot-dashed line). Note that $\sigma(R)$ for the tophat filter is always larger than the others because the ringing of the spherical Bessel function captures power from higher-k modes than the other two filters. Throughout this section, we use R, M, and S interchangeably.

halo of mass between M and $M + dM$ is proportional to the fractional number density of halos in the same mass range. This allows us to calculate the halo mass function $\overline{n}_h(M)$.

Note, however, that the excursion set with a constant critical density threshold generically underestimates the collapsed fraction [395, 318, 396]. As we have mentioned earlier, this is because any region below the threshold that is associated with a collapsed object is not counted as a part of the collapsed region in the excursion set. N-body simulations show that the underestimation is most severe for massive halos. A possible remedy to this problem is to reduce the threshold in a mass-dependent way, although this does not address the underlying physical deficiency of the approach. An alternative is to explicitly (albeit approximately) incorporate the associated regions below the threshold as is done in the peak model (Sec. 6), where all the mass associated with a peak in the smoothed Lagrangian density field is counted towards the final halo, and one does not have to calculate the collapsed volume fraction explicitly.

When implementing the excursion-set formalism, we start from an infinitely large smoothing radius R around any given point in the Universe. Since the density averaged over this region is the cosmic mean density, the density contrast must vanish as $R \to \infty$. We then gradually decrease the smoothing length R, and trace the evolution of the density contrast $\delta(R)$ smoothed at scale R. We proceed further until the smoothed density contrast *first* exceeds the critical value $\delta_{\mathrm{cr}}(z)$, which signifies that the point belongs to a halo with mass given by the smoothing radius. The density contrast at any given smoothing scale fluctuates around zero with corresponding variance $\sigma^2(R)$. Unlike what is done in the Press-Schechter formalism however, excursion-set theory records the "trajectory" $\delta(R)$ starting from $R = \infty$ in order to find the *first-crossing* fraction. In this way, the cloud-in-cloud problem is resolved by counting only *isolated* virialized regions and keeping track of underdensities within larger virialized objects.

Trajectories can be thought of as random walk with R as time variable (Fig. 23). Alternatively, they can also be parametrized with the variance $S(R) \equiv \sigma^2(R)$ of density fluctuations at smoothing scale R [Eq. (5.12)]. For standard cosmological models, $S(R)$ is a monotonically decreasing function of the filtering scale R and, therefore, is in one-to-one correspondence with R and M. Hence, we shall hereafter use M, R

115

and S interchangeably when we describe the halo scale. The relation between them is shown in Fig. 22 for our fiducial cosmology.

In order to find the first-crossing probability as a function of either R, S or M, given the matter power spectrum $P_L(k)$, one can use Monte-Carlo techniques [375, 318]. Analytical solutions of the excursion set are available in some specific cases. In particular, two *exact* solutions for the excursion set formalism are known in two extreme cases, where the statistics of density contrast at different smoothing scales are (i) completely independent (for example, when employing sharp-k filtering [375]), or (ii) completely correlated. Expansions around these exact solutions are often employed in the literature [413, 153] in order to solve excursion-set problems for more realistic filter shapes, such as a real-space tophat.

5.4 Survival probability, halo mass function, and bias

One of the key quantities computed in the excursion-set formalism is the *differential survival probability* $\Pi(\delta; R_0)$ that the density field $\delta(R)$ does not exceed the critical value $\delta_{\rm cr}$ for all smoothing radii R greater than R_0, and reaches some value δ at smoothing radius R_0: $\delta(R_0) = \delta$.

The knowledge of $\Pi(\delta; R_0)$ leads to the halo mass function as follows. Firstly, we calculate the probability that the smoothed density contrast never exceeded $\delta_{\rm cr}$ for all radii greater than R_0 by integrating the differential survival probability $\Pi(\delta; R_0)$ from $-\infty$ to $\delta_{\rm cr}$. Its complement is the probability that $\delta(R)$ exceeded $\delta_{\rm cr}$ *at least once* at some radius $R > R_0$, i.e. it is the fraction $F(> M)$ of random field points belonging to halos of mass greater than $M(R_0)$:

$$F(> M) = 1 - \int_{-\infty}^{\delta_{\rm cr}} d\delta\, \Pi(\delta; R_0) . \tag{5.27}$$

The differential volume fraction

$$f(M)dM = -\frac{dF(> M)}{dM}dM = \frac{dR}{dM}\frac{\partial}{\partial R}\left[\int_{-\infty}^{\delta_{\rm cr}} d\delta\, \Pi(\delta; R)\right]dM , \tag{5.28}$$

must now be interpreted as a *first-crossing distribution*, i.e. the probability that the smoothed density contrast exceeds the threshold $\delta_{\rm cr}$ within a mass range $[M, M + dM]$. The halo mass function follows from Eq. (5.15),

$$\overline{n}_h(M) = \overline{\rho}_m f(M) = \overline{\rho}_m\left[\frac{dR}{dM}\frac{\partial}{\partial R}\int_{-\infty}^{\delta_{\rm cr}} d\delta\, \Pi(\delta; R)\right] . \tag{5.29}$$

To calculate the halo bias in the excursion-set formalism, we consider a *conditional* survival probability in analogy with the peak-background split discussed in Sec. 3. Namely, the LIMD bias is given by the response of the number density of collapsed objects to a long-wavelength density perturbation $\delta_\ell \equiv \delta(R_\ell)$, where $R_\ell \gg R$ is some arbitrary large scale, in the large-scale limit $R_\ell \to \infty$. Therefore, we must compute the fraction $F(> M|\delta_\ell)$ of overdense regions conditioned on the presence of a long-wavelength density perturbation.

In line with the aforementioned argument, the number density of collapsed objects with size R in the large-scale region with density contrast δ_ℓ is given by the *conditional survival probability* $\Pi(\delta; R|\delta_\ell; R_\ell)$:

$$\overline{n}_h(M|\delta_\ell; R_\ell) = \overline{\rho}_m\left[\frac{dR}{dM}\frac{\partial}{\partial R}\int_{-\infty}^{\delta_{\rm cr}} d\delta\, \Pi(\delta; R|\delta_\ell; R_\ell)\right] . \tag{5.30}$$

The density contrast in Lagrangian space then reads (cf. the discussion in Sec. 2.1)

$$1 + \delta_h^L \equiv \frac{\overline{n}_h(M|\delta_\ell, R_\ell)}{\overline{n}_h(M)}$$

$$= \frac{f(M|\delta_\ell; R_\ell)}{f(M)} = \left[\frac{\partial}{\partial R}\int_{-\infty}^{\delta_{\rm cr}} d\delta\, \Pi(\delta; R)\right]^{-1}\left[\frac{\partial}{\partial R}\int_{-\infty}^{\delta_{\rm cr}} d\delta\, \Pi(\delta; R|\delta_\ell; R_\ell)\right] . \tag{5.31}$$

Here, the conditional probability can be evaluated as

$$\Pi(\delta; R|\delta_\ell; R_\ell) = \frac{\Pi[(\delta; R), (\delta_\ell; R_\ell)]}{\Pi(\delta_\ell; R_\ell)} \ , \tag{5.32}$$

where $\Pi[(\delta; R), (\delta_\ell; R_\ell)]$ is the survival probability that the smoothed density contrast is δ_ℓ at the larger smoothing scale R_ℓ (*background* scale), and δ at the smaller smoothing R (*peak* scale). This can be pictorially represented as random walks starting from $(\delta, R) = (\delta_\ell, R_\ell)$ instead of the origin $(\delta, R_\ell) = (0, \infty)$ (see Fig. 24).

Similarly, with the variance S as time variable, the fraction of random walks $f(S)$ that first cross the barrier in the range $[S, S + dS]$ is related to the halo mass function through

$$f(S)dS = \frac{1}{\overline{\rho}_m} \overline{n}_h(M)dM \ . \tag{5.33}$$

Therefore, we have the relation

$$f(M)dM = f(S)dS = f(\nu_c)d\nu_c \ . \tag{5.34}$$

Since $\nu_c = \delta_{\rm cr}/\sqrt{S}$, this yields also

$$\nu_c f(\nu_c) = 2Sf(S) \ . \tag{5.35}$$

Furthermore, the bias coefficients are computed by considering

$$1 + \delta_h^L = \frac{f(S|\delta_\ell; S_\ell)}{f(S)} \ . \tag{5.36}$$

Note again that the LIMD bias parameters, which describe the response to an infinite-wavelength density perturbation, are obtained in the limit $S_\ell \to 0$. The leading correction for a finite S_ℓ can be mapped onto a higher-derivative bias $b_{\nabla^2\delta}$ [152, 305].

5.5 Numerical solution of excursion sets: Langevin equation

One approach to derive the first-crossing distribution is to simulate a random walk of the density at a fixed location, with a smoothing radius decreasing from $R = \infty$ or, equivalently, a variance increasing from $S = 0$. In this case, the excursion-set formalism reduces to a diffusion problem that is described by the Langevin equation where the rate of change of the smoothed linear density field $\delta(R)$ as a function of smoothing length is given by a *stochastic force* $\mathcal{Q}(R)$ [375] (this force also depends on the position $\boldsymbol{q}$, which we suppress throughout since it is arbitrary, but fixed):

$$\frac{d\delta(R)}{dR} = \mathcal{Q}(R) \ . \tag{5.37}$$

The stochastic force has a vanishing expectation value $\langle\mathcal{Q}(R)\rangle = 0$, but nonzero variance. For any isotropic filtering kernel, the covariance, or two-point function, of the stochastic force is given by

$$\langle\mathcal{Q}(R_1)\mathcal{Q}(R_2)\rangle = \int_{\boldsymbol{k}} P_{\rm L}(k) \frac{dW_{R_1}(k)}{dR_1} \frac{dW_{R_2}^*(k)}{dR_2} \ . \tag{5.38}$$

5.5.1 Monte-Carlo solution with sharp-k filter

Solving the Langevin equation is particularly simple for the case of the sharp-k filter, because stochastic forces at each step are independent owing to the Gaussian nature of the linear density field, and the statistical homogeneity of the Universe. For the sharp-k filter, the derivative is $dW_R(k)/dR = d\Theta_H(1 - kR)/dR = -k/R\,\delta_D\,(k - 1/R)$, where Θ_H is the Heaviside step function. Consequently, Eq. (5.38) becomes

$$\langle\mathcal{Q}(R_1)\mathcal{Q}(R_2)\rangle = -\left.\frac{k^2 P_{\rm L}(k)}{2\pi^2}\right|_{k=\frac{1}{R_1}} \frac{1}{R_1^2}\delta_D\,(R_1 - R_2) \ . \tag{5.39}$$

117

This suggests rewriting the Langevin equation in terms of $\ln k \equiv -\ln R$:

$$\frac{d\delta(R = 1/k)}{d\ln k} = \mathcal{Q}(\ln k) \,, \tag{5.40}$$

where

$$\langle \mathcal{Q}(\ln k_1)\mathcal{Q}(\ln k_2)\rangle = \frac{k_1^3 P_{\mathrm{L}}(k_1)}{2\pi^2}\delta_D(\ln k_1 - \ln k_2) \,. \tag{5.41}$$

Eq. (5.41) allows us to generate realizations of the stochastic force $\mathcal{Q}(\ln k)$ from a given linear matter power spectrum $P_{\mathrm{L}}(k)$. We can calculate the density contrast $\delta(R = 1/k)$ from the formal solution of Eq. (5.40), which may be written as

$$\delta(R) = \int_{-\infty}^{-\ln R} d(\ln k')\mathcal{Q}(\ln k') \,, \tag{5.42}$$

and from there calculate the survival probability $\Pi(\delta; S)$ as the fraction of random walks that first cross the barrier $\delta_{\mathrm{cr}}(z)$ between S and $S + dS$.

Note that both the Langevin equation for sharp-k filtering, Eq. (5.40), as well as the stochastic force, Eq. (5.41), do not explicitly depend on the step R or the density contrast δ_L. Therefore, changing the *initial time* R_{init} and the density contrast $\delta(R_{\mathrm{init}})$ simply re-defines the integration boundaries of Eq. (5.42). The usual choice for evaluating the halo mass function is to set the initial value as $(R_{\mathrm{init}}, \delta(R_{\mathrm{init}})) = (\infty, 0)$, because the density contrast averaged over the entire Universe vanishes.

In practice, the numerical implementation can be done as follows. Upon discretizing the steps in Fourier space, $\ln k_i = (\Delta \ln k)i$, the stochastic force is drawn from a Gaussian distribution centered around zero and with variance

$$\sigma_{\mathcal{Q}}^2 = \frac{k_i^3 P_{\mathrm{L}}(k_i)}{2\pi^2 \Delta \ln k} \,. \tag{5.43}$$

Then, a random walk of the density contrast can be calculated by discretizing the integration as

$$\delta(R) = \sum_{k_i < 1/R} r_i \sqrt{\frac{k_i^3 P(k_i)}{2\pi^2}\Delta \ln k} \qquad \text{(for the sharp-k filter)} \,, \tag{5.44}$$

where r_i is a random number drawn from the normal distribution $\mathcal{N}(0,1)$. For illustration, we show in Fig. 23 two excursion-set random walks with the sharp-k filter as solid lines.

5.5.2 Monte-Carlo solution with general filters

Dor general filter functions, such as the real-space tophat or Gaussian filters, the derivative $dW_R(k)/dR$ is not a Dirac delta in Fourier space, and the stochastic forces at different smoothing lengths $\mathcal{Q}(R_i)$, $\mathcal{Q}(R_j)$ are no longer independent from each other. That is, the stochastic force at each step $\mathcal{Q}(R_i)$ depends on the entire set of previous steps $\{\mathcal{Q}(R_0), \mathcal{Q}(R_1), \cdots \mathcal{Q}(R_{i-1})\}$, so that that the random walk is *non-Markovian*. In principle, for a given filter function and linear matter power spectrum, one can draw a random Gaussian realization of $\mathcal{Q}(R_i)$ from the full covariance matrix $\langle \mathcal{Q}(R_i)\mathcal{Q}(R_j)\rangle$ given in Eq. (5.38).

Alternatively, we can exploit the fact that the linear density contrast $\delta(\boldsymbol{k})$ is Gaussian, which in particular implies that Fourier amplitudes of different wavenumber are independent. To do so, we first re-write the smoothed linear density field in Eq. (5.26) in Fourier space as

$$\delta(R) = \int_{\boldsymbol{k}} \delta(\boldsymbol{k})W_R(k), \tag{5.45}$$

where we set the Lagrangian coordinate of interest $\boldsymbol{q} = 0$ without loss of generality. Comparing this to the case for the sharp-k filter in Eq. (5.42), it is clear that the stochastic force $\mathcal{Q}^{\mathrm{sk}}(\boldsymbol{k})$ in the sharp-k case precisely corresponds to the linear density contrast integrated over a spherical shell in Fourier space with fixed logarithmic width $d\ln k$. Using this, we can write the solution of the Langevin equation in terms of
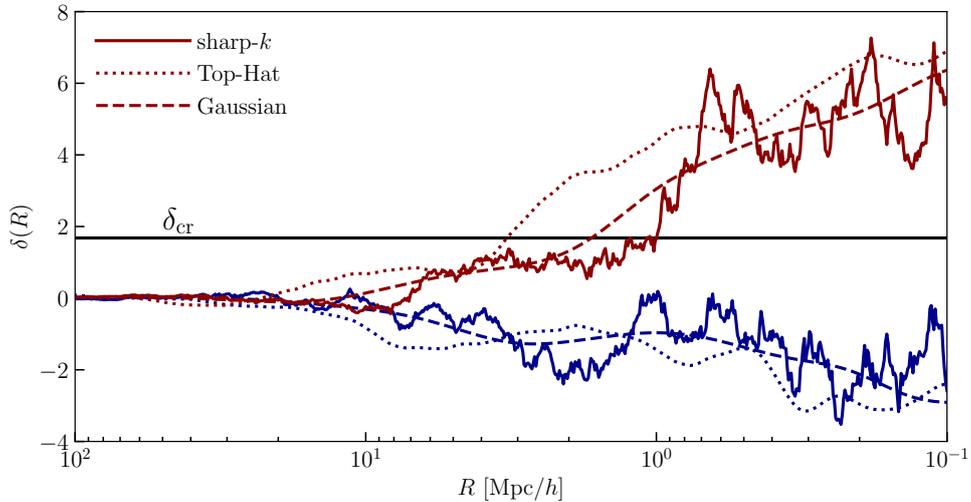
Figure 23: Two sets of three example excursion-set random walks of the Lagrangian density contrast $\delta(R)$ as a function of the filter radius R. Lines with the same color show the same realization of excursion-set random walk but with three different filter functions: sharp-k filter (*solid lines*), Gaussian filter (*dashed lines*), tophat filter (*dotted lines*). While the blue (bottom) realization has not formed a halo for scales $R > 0.1$ [Mpc/h], the red (upper) realization forms a halo at $R = 1 \sim 3$ [Mpc/h] depending on the filter function. For this calculation, we use 4000 equal steps in logarithmic space between $k = 10^{-4} h/\mathrm{Mpc}$ and $10^2 h/\mathrm{Mpc}$.

the filter function $W_R(k)$ and sharp-k stochastic force $\mathcal{Q}^{\mathrm{sk}}(\boldsymbol{k})$ as

$$\delta(R) = \int_0^\infty d(\ln k') \mathcal{Q}^{\mathrm{sk}}(\ln k') W_R(\ln k). \tag{5.46}$$

Note that the integration runs from 0 to ∞. By taking a derivative of Eq. (5.46) with respect to R, we find the stochastic force $\mathcal{Q}(R)$ for a general filter function which satisfies the required covariance matrix given in Eq. (5.38).

For a given set of stochastic forces $\mathcal{Q}^{\mathrm{sk}}(k_i)$, or a given set of random variables r_i in Eq. (5.44), the corresponding excursion-set random walk with filter $W_R(k)$ can be calculated from the discretized version of Eq. (5.46),

$$\delta(R) = \sum_i r_i W_R(k_i) \sqrt{\frac{k_i^3 P(k_i)}{2\pi^2} \Delta \ln k} \qquad \text{(for a generic filter)} . \tag{5.47}$$

Here, i labels the steps in log wavenumber that are sampled. For comparison, we overlay in Fig. 23 two random walks obtained with a Gaussian (dashed lines) and tophat (dotted lines) filter. They were generated with the same sequence of random numbers r_i used for the sharp-k random walk. As the random walk at each radius depends on all Fourier modes (the dependence is determined by the filter function), the random walk with real-space tophat and Gaussian filters are correlated and, consequently appear much smoother than that constructed with the sharp-k filter.

5.6 Analytical approaches I: excursion set with uncorrelated steps

Although the survival probability, mass function and halo bias can be calculated numerically as explained in Sec. 5.4, there are a number of analytical studies of the excursion-set formalism in the literature. The analytical approaches help us better understand the problem, in addition to providing closed form expressions for the halo bias. We shall now review some of the recent developments along this direction. Throughout, $S \equiv \sigma^2(R)$, while s is a dummy time variable.

119

5.6.1 Completely independent (Markovian) steps with sharp-k filter

Consider first the excursion-set random walks with sharp-k filter. The real-space kernel corresponding to this choice is oscillatory and thus very different from physical Lagrangian density profiles of halos. Nevertheless, this case is the most popular in the literature because an exact solution to the first-crossing problem can easily be derived.

As we have shown in Sec. 5.5.1, the shark-k filtering guarantees (for Gaussian initial conditions) that the steps of the random walk are fully independent or Markovian. That is, the density contrast $\delta(s+\Delta s)$ is only determined by the density contrast at the previous step $\delta(s)$, and the variance Δs: the steps which the random walk has gone through to reach $\delta(s)$ are irrelevant. When performing the step from s to $s+\Delta s$ (i.e. the filter decreases from R to $R-\Delta R$), the difference $\delta(s+\Delta s)-\delta(s)$ in the density contrast is a Gaussian random variate with zero mean and variance Δs. Only Fourier modes in the range $k \in [1/R, 1/(R-\Delta R)]$ contribute to this random variate, proving that the step is independent from all the others.

Since the sum of independent Gaussian random variables is also a Gaussian with a variance given by the sum of individual variances, the probability distribution of trajectories reaching $\delta(S) = \delta$ at time S is

$$p_G(\delta)d\delta = \frac{1}{\sqrt{2\pi S}}e^{-\delta^2/2S}d\delta \; . \tag{5.48}$$

Then, the differential survival probability is given by [414]

$$\Pi(\delta;S) = \begin{cases} \frac{1}{\sqrt{2\pi S}}\left[e^{-\delta^2/2S} - e^{-(2\delta_{\rm cr}-\delta)^2/2S}\right] & (\delta < \delta_{\rm cr}) \\ 0 & (\delta \geq \delta_{\rm cr}) \end{cases} \; . \tag{5.49}$$

This can be understood as follows. The total fraction of trajectories attaining (S,δ) with $\delta < \delta_{\rm cr}$ is given in Eq. (5.48), which is the first term in Eq. (5.49). However, not all trajectories passing through (S,δ) always remained below the threshold $\delta_{\rm cr}$ at earlier time. Therefore, we must subtract all paths which crossed the threshold $\delta_{\rm cr}$ for some $S' < S$. To estimate this fraction, consider a path reaching (δ, S) after crossing $\delta_{\rm cr}$ at an earlier time $S' < S$. We can construct another path which arrives at $(S, \delta_{\rm cr} + (\delta_{\rm cr} - \delta))$ and which, at time $S' < s < S$, is the exact reflection of the original path off the "mirror" $\delta = \delta_{\rm cr}$. Clearly, this reflection establishes a one-to-one correspondence between trajectories ending up at $(S, 2\delta_{\rm cr} - \delta)$ and the paths that reach (S,δ) after crossing the threshold at time $S' < S$. Therefore, we must subtract $p_G(2\delta_{\rm cr} - \delta)$ from $p_G(\delta)$, which is nothing but Eq. (5.49).

The survival probability is thus given by

$$\int_{-\infty}^{\delta_{\rm cr}} d\delta\,\Pi(\delta;S) = \int_{-\infty}^{\delta_{\rm cr}} \frac{1}{\sqrt{2\pi S}}\left[e^{-\delta^2/2S} - e^{-(2\delta_{\rm cr}-\delta)^2/2S}\right]d\delta = {\rm erf}\left[\frac{\delta_{\rm cr}}{\sqrt{2S}}\right] \; , \tag{5.50}$$

from which we calculate the collapsed fraction as

$$F(> M) = 1 - \int_{-\infty}^{\delta_{\rm cr}} d\delta\,\Pi(\delta;S) = {\rm erfc}\left[\frac{\delta_{\rm cr}}{\sqrt{2S}}\right] \; . \tag{5.51}$$

This yields the halo mass function, via Eq. (5.29),

$$\overline{n}_{h,\rm sk}(M) = \frac{\overline{\rho}_m}{M}\nu_c f_{\rm sk}(\nu_c)\left|\frac{d\ln\sigma(R)}{d\ln M}\right| \; , \tag{5.52}$$

where

$$\nu_c f_{\rm sk}(\nu_c) = \sqrt{\frac{2}{\pi}}\nu_c e^{-\nu_c^2/2} \; . \tag{5.53}$$

This shows that the Markovian (sharp-k) excursion-set formalism recovers the Press-Schechter mass function [Eq. (5.18)] with the correct normalization factor.

5.6.2 Halo bias with sharp-k filter

In order to calculate the Lagrangian bias coefficients, we need to evaluate the conditional survival probability

$$\Pi(\delta; S|\delta_\ell; S_\ell) = \frac{\Pi\left[(\delta; S), (\delta_\ell; S_\ell)\right]}{\Pi(\delta_\ell; S_\ell)} . \tag{5.54}$$

For the case of a sharp-k filter, each step of the random walk is independent so that the trajectory is invariant under translation—that is, it does not matter where the walk has started. Consequently, the differential survival probability is given by

$$\Pi\left[(\delta; S), (\delta_\ell; S_\ell)\right] = \Pi(\delta - \delta_\ell; S - S_\ell)\Pi(\delta_\ell; S_\ell). \tag{5.55}$$

We can then calculate the conditional collapsed fraction from a simple rescaling of Eq. (5.51),

$$F(> M|\delta_\ell; S_\ell) = \mathrm{erfc}\left[\frac{\delta_{\mathrm{cr}} - \delta_\ell}{\sqrt{2(S - S_\ell)}}\right] . \tag{5.56}$$

The conditional mass function immediately follows,

$$\overline{n}_{h,\mathrm{sk}}(M|\delta_\ell) = \frac{\overline{\rho}_m}{M} 2S f_{\mathrm{sk}}(S|\delta_\ell; S_\ell) \left|\frac{d\ln\sigma(R)}{d\ln M}\right| , \tag{5.57}$$

with a conditional first-crossing distribution $f_{\mathrm{sk}}(S|\delta_\ell; S_\ell)$ given by

$$f_{\mathrm{sk}}(S|\delta_\ell; S_\ell) = \frac{1}{\sqrt{2\pi}} \frac{\delta_{\mathrm{cr}} - \delta_\ell}{(S - S_\ell)^{3/2}} \exp\left[-\frac{(\delta_{\mathrm{cr}} - \delta_\ell)^2}{2(S - S_\ell)}\right] . \tag{5.58}$$

Therefore, the halo density contrast in a Lagrangian region of size R_ℓ and overdensity δ_ℓ is

$$1 + \delta_h^L(\delta_\ell) \equiv \frac{f_{\mathrm{sk}}(S|\delta_\ell, S_\ell)}{f_{\mathrm{sk}}(S)} = \left(1 - \frac{\delta_\ell}{\delta_{\mathrm{cr}}}\right) \left(\frac{S}{S - S_\ell}\right)^{3/2} \exp\left[-\frac{(\delta_{\mathrm{cr}} - \delta_\ell)^2}{2(S - S_\ell)} + \frac{\delta_{\mathrm{cr}}^2}{2S}\right], \tag{5.59}$$

which, in the limit where $R_\ell \to \infty$ and $S_\ell \to 0$, reduces to the Lagrangian LIMD bias expansion:

$$1 + \delta_h^L(\delta_\ell) = \left(1 - \frac{\delta_\ell}{\delta_{\mathrm{cr}}}\right) \exp\left[\frac{(2\delta_{\mathrm{cr}} - \delta_\ell)\delta_\ell}{2S}\right] = 1 + \frac{\delta_\ell}{\sigma(R)} \left(\nu_c - \frac{1}{\nu_c}\right) + \frac{1}{2} \frac{\delta_\ell^2}{\sigma^2(R)} \left(\nu_c^2 - 3\right) + \mathcal{O}(\delta_\ell^3), \tag{5.60}$$

The linear- and second-order coefficients of δ_ℓ in Eq. (5.60) are, respectively, the linear- and the second-order Lagrangian LIMD bias parameters.

Let us compare this with the bias parameters predicted by the peak-background split following Sec. 3.3, which are given by

$$b_n^L = (-1)^n \frac{1}{\overline{n}_{h,\mathrm{sk}}(M)} \frac{\partial^n \overline{n}_{h,\mathrm{sk}}(M)}{\partial \delta_{\mathrm{cr}}^n} . \tag{5.61}$$

It is easy to see, via Eqs. (5.52)–(5.53), that these are exactly the same as Eq. (5.60), the biases derived in the excursion set from the conditional first crossing distribution in the large-scale limit $S_\ell \to 0$. Explicitly, the first two bias parameters are the same as Eq. (3.28),

$$(b_1^L)_{\mathrm{sk}} = (b_1^L)_{\mathrm{PS}} = -\frac{1}{n_{h,\mathrm{sk}}(M)} \frac{\partial \overline{n}_{h,\mathrm{sk}}(M)}{\partial \delta_{\mathrm{cr}}} = \frac{1}{\sigma(R)} \left(\nu_c - \frac{1}{\nu_c}\right) \tag{5.62}$$

$$(b_2^L)_{\mathrm{sk}} = (b_2^L)_{\mathrm{PS}} = \frac{1}{n_{h,\mathrm{sk}}(M)} \frac{\partial^2 \overline{n}_{h,\mathrm{sk}}(M)}{\partial^2 \delta_{\mathrm{cr}}} = \frac{1}{\sigma^2(R)} \left(\nu_c^2 - 3\right) . \tag{5.63}$$

Note that, in the limit $\sigma(R) \to \infty$, the linear bias b_1^L tends towards $-1/\delta_{\mathrm{cr}}$. That is, unlike the simple thresholding procedure described in Sec. 2.1, low mass objects are predicted to be anti-biased. This can be traced to the first-crossing constraint, which results in low-mass *isolated* virialized objects being preferentially found in underdense regions.

121
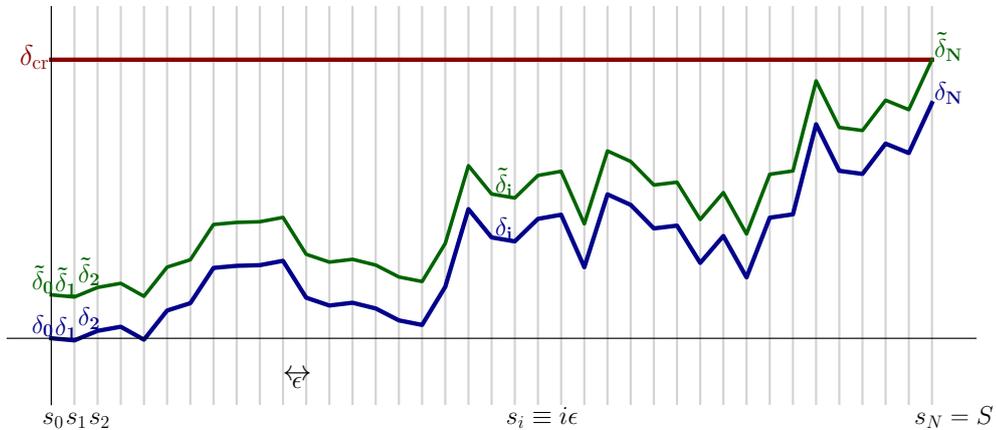
Figure 24: Illustration of the variables used in Sec. 5.6.3. We show two excursion-set paths with exactly the same realization of the local density field, but with different large-scale environment. The blue path starts from $\delta(S_0 = 0) = 0$, corresponding to an environment at the cosmic mean density, while the green path starts from $\delta(S_0 = 0) = \tilde{\delta}_0 > 0$, corresponding to an overdense environment.

5.6.3 Expanding around the Markovian (sharp-k) solution*

Above we have obtained an analytical solution for the excursion set in the Markovian case, i.e. with a sharp-k filter. For other filters, the covariance matrix of the stochastic force $\mathcal{Q}$ [Eq. (5.37)] is non-diagonal, which means that all steps at different filter radii are correlated. As a result, the random walk is generally non-Markovian. We now describe generalizations of the excursion-set procedure that are able to deal with these non-Markovian walks, as long as the cross-correlation between steps of the random walk remains small.

An approximation due to [413, 415, 416] describes the correlated random walks in terms of a path-integral of discretized trajectories (with step size $\epsilon \equiv s_{i+1} - s_i$, that is, $s_i = i\epsilon$) starting from $\delta_0 \equiv \delta(s = 0)$ and reaching $s_N \equiv S$. They first write down a joint probability distribution function $\mathcal{W}(\delta_0; \delta_1, \cdots \delta_N)$ in the N-dimensional space of random walk trajectories $\delta_i \equiv \delta(s_i)$. This is given by (see Fig. 24 for an illustration and definitions of symbols)

$$
\mathcal{W}(\delta_0; \delta_1, \cdots, \delta_N) = \left\langle \delta_D\big(\delta(s_1) - \delta_1\big) \cdots \delta_D\big(\delta(s_N) - \delta_N\big) \right\rangle
$$
$$
= \int_{-\infty}^{\infty} \frac{d\lambda_1}{2\pi} \cdots \int_{-\infty}^{\infty} \frac{d\lambda_N}{2\pi} e^{i \sum_i \lambda_i \delta_i} \left\langle e^{-i \sum_i \lambda_i \delta(s_i)} \right\rangle , \tag{5.64}
$$

where we have used the Fourier representation of the Dirac delta distribution $\delta_D(x)$. The differential survival probability $\Pi[(\delta_0; s_0 = 0), (\delta_N; s_N)]$ immediately follows from $\mathcal{W}(\delta_0; \delta_1, \cdots, \delta_N)$ since it is the probability that the path $(\delta_1, \cdots, \delta_{N-1})$ never exceeded the threshold density contrast δ_{cr} along the path:

$$
\Pi_\epsilon(\delta_0; \delta_N; s_N) = \int_{-\infty}^{\delta_{cr}} d\delta_1 \cdots \int_{-\infty}^{\delta_{cr}} d\delta_{N-1} \mathcal{W}(\delta_0; \delta_1, \cdots, \delta_N) . \tag{5.65}
$$

Here, we define the shorthand notation of $\Pi_\epsilon(\delta_0; \delta_N; s_N) \equiv \Pi_\epsilon[(\delta_0; s_0), (\delta_N; s_N)]$ where the subscript ϵ in Eq. (5.65) reminds us that we are dealing with a discrete random walk, and we have to eventually take the limit $\epsilon \to 0$ at the end of the calculation.

---

  * This section is of a more technical nature and is not essential for the remainder of the review. The main result of this section are Eqs. (5.75)–(5.77).

Because $\delta(s_i)$ follow Gaussian statistics, the random variate $X \equiv i \sum_i \lambda_i \delta(s_i)$ is also Gaussian distributed. Therefore, we can simplify further the expectation value of the exponential[17] in Eq. (5.64) to

$$\left\langle \exp\left[-i\sum_i \lambda_i \delta(s_i)\right]\right\rangle = \exp\left[\frac{1}{2}\left\langle \left(i\sum_i \lambda_i \delta(s_i)\right)^2\right\rangle\right] = \exp\left[-\frac{1}{2}\sum_{ij}\lambda_i \lambda_j \langle \delta(s_i)\delta(s_j)\rangle\right]. \qquad (5.66)$$

The differential survival probability then reads, by combining Eq. (5.64) and Eq. (5.66),

$$\Pi_\epsilon(\delta_0; \delta_N; s_N) = \int_{-\infty}^{\delta_{\rm cr}} d\delta_1 \cdots \int_{-\infty}^{\delta_{\rm cr}} d\delta_{N-1} \int_{-\infty}^{\infty}\frac{d\lambda_1}{2\pi}\cdots\int_{-\infty}^{\infty}\frac{d\lambda_N}{2\pi}\exp\left[i\sum_{i=1}^{N}\lambda_i \delta_i - \frac{1}{2}\sum_{i,j=1}^{N}\lambda_i \lambda_j \langle \delta(s_i)\delta(s_j)\rangle\right].$$
$$(5.67)$$

Note that, once the differential survival probability $\Pi_\epsilon(\delta_0; \delta_N; s_N)$ is known, the survival probability, collapsed fraction $F(> M)$ and mass function $f(M)$ follow from the steps described in Sec. 5.4. Also note that, for the sharp-k filter, each increment in $\delta(s_i)$ is independent. Therefore, the two-point correlator is given by

$$\langle \delta(s_i)\delta(s_j)\rangle = \min\left(s_i, s_j\right) = \epsilon\min(i, j), \qquad (5.68)$$

and the PDF can be calculated analytically:

$$\Pi_\epsilon^{\rm sk}(\delta_0; \delta_N; s_N) = \int_{-\infty}^{\delta_{\rm cr}} d\delta_{N-1}\frac{1}{\sqrt{2\pi\epsilon}}\exp\left[-\frac{(\delta_N - \delta_{N-1})^2}{2\epsilon}\right]\Pi_\epsilon^{\rm sk}(\delta_0; \delta_{N-1}; s_{N-1}). \qquad (5.69)$$

This equation is indeed consistent with the fact that the sharp-k filtering yields a Markovian process and, therefore, yields the differential survival probability Eq. (5.49).

One intermediate result for the sharp-k filter case shown in [413] is the finite-ϵ correction. The $\epsilon \to 0$ limit solution in Eq. (5.49) vanishes when the random walk ends at the critical density $\delta = \delta_{\rm cr}$, $\Pi_\epsilon^{\rm sk}(\delta_0; \delta_{\rm cr}; S) = 0$, or begins from the critical density $\Pi_\epsilon^{\rm sk}(\delta_{\rm cr}; \delta; S) = 0$. With the finite step size ϵ, Ref. [413] has shown that the leading corrections for these cases are of order $\sqrt{\epsilon}$. Namely,

$$\Pi_\epsilon^{\rm sk}(\delta_0; \delta_{\rm cr}; S) = \sqrt{\frac{\epsilon}{\pi S^3}}(\delta_{\rm cr} - \delta_0)\exp\left[-\frac{(\delta_{\rm cr} - \delta_0)^2}{2S}\right] + \mathcal{O}(\epsilon) \qquad (5.70)$$

$$\Pi_\epsilon^{\rm sk}(\delta_{\rm cr}; \delta; S) = \sqrt{\frac{\epsilon}{\pi S^3}}(\delta_{\rm cr} - \delta)\exp\left[-\frac{(\delta_{\rm cr} - \delta)^2}{2S}\right] + \mathcal{O}(\epsilon) \qquad (5.71)$$

if δ_0 and δ are both smaller than $\delta_{\rm cr}$. The finite-ϵ corrections are necessary to calculate the non-Markovian correction because the same parameter ϵ appears when converting a finite sum, for example, in Eq. (5.73), to the continuous limit as $\sum \to 1/\epsilon \int ds$.

For the general filters other than the sharp-k filter, different time steps are correlated, and the two-point correlators may be written as (for $0 \le s_i \le s_j$)

$$\langle \delta(s_i)\delta(s_j)\rangle = s_i + \Delta(s_i, s_j) \simeq s_i + \kappa\frac{s_i(s_j - s_i)}{s_j}, \qquad (5.72)$$

where $\Delta(s_i, s_j) \equiv \Delta_{ij}$ is the cross-correlation between different smoothing scales ($\Delta_{ij} = 0$ for the sharp-k case). Therefore, κ parametrizes the deviation from the sharp-k case for which we have an analytical solution. For the spherical tophat filter (which [413] refers to as *sharp-x* filter), $\kappa \simeq 0.4592 - 0.0031\,(R/[1{\rm Mpc}/h])$ whereas, for the Gaussian filter, $\kappa \simeq 0.35$ for a ΛCDM linear power spectrum. Next, Ref. [413] integrated

---

[17]For a Gaussian random variable X, $\langle e^X \rangle = \exp\left[\frac{1}{2}\langle X^2\rangle\right]$. One way to prove this relation is to Taylor-expand the left-hand side and use Wick's theorem: $\langle X^{2n}\rangle = (2n-1)!!\left[\langle X^2\rangle\right]^n$.

Eq. (5.67) with Eq. (5.72) by treating Δ_{ij} (and therefore κ) as a small perturbation. To linear order in κ, the differential survival probability is

$$\Pi_\epsilon(\delta_0; \delta_N; s_N) = \Pi_\epsilon^{(0)}(\delta_0; \delta_N; s_N) + \frac{1}{2} \sum_{i,j=1}^N \Delta_{ij} \int_{-\infty}^{\delta_{\rm cr}} d\delta_1 \cdots \int_{-\infty}^{\delta_{\rm cr}} d\delta_{N-1} \frac{\partial}{\partial \delta_i} \frac{\partial}{\partial \delta_j} \mathcal{W}^{\rm sk}(\delta_0; \delta_1, \cdots, \delta_N; s_N),$$
(5.73)

where the leading-order term $\Pi_\epsilon^{(0)}$ is the same expression as the sharp-k filtering case, which reduces to Eq. (5.49) in $\epsilon \to 0$ limit. $\mathcal{W}^{\rm sk}$ is the probability density function of the sharp-k case, that is, a multivariate Gaussian with diagonal covariance matrix, where the variance $s(R_i)$ is calculated with the general filter of interest. Note that $s(R_i) = \sigma^2(R_i)$ depends on the shape of filter chosen (Fig. 22).

The perturbative expansion is justified as the maximum of $\Delta_{ij}/s_i \simeq \kappa(1 - s_i/s_j) < \kappa$ is less than unity; at the same time, however, truncation at the leading order may not lead to an accurate approximation as $\kappa \simeq 0.3 - 0.4$. The linear-order result is, in the $\epsilon \to 0$ limit, given by

$$\Delta\Pi(\delta; S) = \kappa \frac{\partial}{\partial \delta} \left[\frac{\delta_{\rm cr}(\delta_{\rm cr} - \delta)}{S} \operatorname{erfc}\left(\frac{2\delta_{\rm cr} - \delta}{\sqrt{2S}} \right) \right]$$
$$+ \frac{\kappa \delta_{\rm cr}}{\sqrt{2\pi S}} \frac{\partial}{\partial \delta} \left[e^{-(\delta_{\rm cr} - \delta)^2/(2S)} \int_0^S \frac{ds}{s} e^{-\delta_{\rm cr}/(2s)} \operatorname{erfc}\left((\delta_{\rm cr} - \delta)\sqrt{\frac{s}{2S(S-s)}} \right) \right], \quad (5.74)$$

which, on using Eq. (5.29), yields the halo mass function

$$\overline{n}_h(M) = \sqrt{\frac{2}{\pi}} \frac{\overline{\rho}_m}{M} \left[(1 - \kappa)\nu_c e^{-\nu_c^2/2} + \frac{\kappa \nu_c}{2} \Gamma\left(0, \frac{\nu_c^2}{2} \right) \right] \left| \frac{d\ln\sigma}{d\ln M} \right|, \quad (5.75)$$

where the incomplete Gamma function is $\Gamma(0, x) = \int_x^\infty e^{-t} d\ln t$. This mass function has been found to be in good agreement with simulations [417, 418]. The linear and quadratic Lagrangian bias parameters can be calculated from the peak-background split approach in the same way as in Eq. (5.62), leading to[18]

$$b_1^L(\nu_c) = \frac{1}{\delta_{\rm cr} \left[1 - \kappa + \frac{\kappa}{2} e^{\nu_c^2/2} \Gamma(0, \nu_c^2/2) \right]} \left\{ (1 - \kappa)(\nu_c^2 - 1) + \frac{\kappa}{2} \left[2 - e^{\nu_c^2/2} \Gamma\left(0, \frac{\nu_c^2}{2} \right) \right] \right\}, \quad (5.76)$$

$$b_2^L(\nu_c) = \frac{1}{\delta_{\rm cr}^2 \left[1 - \kappa + \frac{\kappa}{2} e^{\nu_c^2/2} \Gamma(0, \nu_c^2/2) \right]} \left\{ (1 - \kappa)\nu_c^4 + (4\kappa - 3)\nu_c^2 - \kappa \right\}, \quad (5.77)$$

where the appropriate value for κ should be inserted, depending on the filter used.

5.7 Analytical approaches II: excursion set with correlated steps

Another method to computing the first-crossing probability as a function of the filtering scale R for general (non sharp-k) filters is to expand around the completely-correlated solution (the original Press-Schechter result) [396, 419, 153, 305, 420]. This method originated from the observation of [410] that the trajectories $\delta(R)$ can, to a first approximation, be decomposed into a series of independent steps (in $\ln R$) of finite size Δ. Here, the critical step size Δ is a correlation length, i.e. $\delta(R)$ at two different smoothing scales R_1 and R_2 are strongly correlated if $|\ln R_1 - \ln R_2| \lesssim \Delta$, and weakly correlated if $|\ln R_1 - \ln R_2| \gtrsim \Delta$. In the limit $\Delta \to 0$, we recover the Markovian walks (sharp-k filtering case, Sec. 5.6.1) whereas, in the opposite limit $\Delta \to \infty$, the walks are referred to as completely correlated.

---

[18]This expression corrects a typo in [416].

5.7.1 First-crossing with completely correlated steps

As noted in [419], an example of completely correlated walks is the set of straight lines passing through the origin and a point (δ, S). These curves are fully characterized by the slope $\nu = \delta/\sqrt{S}$. The first-crossing distribution associated with this ensemble of walks will depend on the assumed distribution of ν. If this distribution is Gaussian, $p(\nu) = \exp(-\nu^2/2)/\sqrt{2\pi}$, then the corresponding survival probability for the constant barrier of height $\delta_{\rm cr}$ is given by

$$\int_{-\infty}^{\delta_{\rm cr}} d\delta \,\Pi(\delta; S) = \int_{-\infty}^{\delta_{\rm cr}/\sqrt{S}} d\nu \, p(\nu) = \frac{1}{2}\left(1 + {\rm erf}\left(\delta_{\rm cr}/\sqrt{2S}\right)\right) . \tag{5.78}$$

Consequently, the fraction of walks first crossing the barrier in the range $(S, S + dS)$ is one half of the solution for completely uncorrelated steps,

$$S f(S) = \frac{1}{2} S f_{\rm sk}(S) = \frac{1}{2}\frac{\delta_{\rm cr}}{\sqrt{2\pi S}}e^{-\delta_{\rm cr}^2/2S} , \tag{5.79}$$

i.e. the original Press-Schechter solution [217] (see Sec. 5.2.2) with a halo mass function of

$$\overline{n}_h(M) = \frac{\overline{\rho}_m}{M}\frac{\delta_{\rm cr}}{\sqrt{2\pi S}}e^{-\delta_{\rm cr}^2/2S}\left|\frac{d\ln\sigma}{d\ln M}\right| . \tag{5.80}$$

This expression characterizes very well the first-crossing distribution at small S (that is, for high-mass halos), even when the steps are not fully correlated [421]. Comparison with Monte-Carlo realizations of random walks show that, at small S, the first-crossing distribution also asymptotes to $1/2$ times the corresponding distribution for sharp-k walks when the shape of the underlying power spectrum is varied, or primordial non-Gaussianity (PNG) is added [421].

PNG generates a skewness in the initial density field which affects the first-crossing distribution and, thus, the halo mass function. This effect is largest at high mass because the tails of the density PDF $p(\delta_R)$ are very sensitive to any non-zero skewness in the initial conditions [422, 423, 424]. In practice, non-Gaussian corrections to the initial density PDF can be computed using e.g. saddle-point techniques and Edgeworth expansions [425, 426, 427]. We will not discuss this issue any further here, and refer the reader to [428] for additional details on the halo mass function in the presence of PNG. The effect of PNG on the clustering of biased tracers in the context of the excursion set is the topic of Sec. 7.4.2.

5.7.2 Up-crossing probability distribution function*

In order to calculate the survival probability, one must in principle enforce the constraint $\delta(s) < \delta_{\rm cr}$ for all scales $s < S$. When the steps are strongly (yet not perfectly) correlated, however, Refs. [153, 396] (see also [429]) have suggested to replace this condition with the milder requirement that *the first crossing must happen from below*: $\delta(S - \Delta S) < \delta_{\rm cr}$ and $\delta(S) > \delta_{\rm cr}$. In the limit of small increment ΔS, the condition of first crossing happening from below is equivalent to

$$\delta_{\rm cr} < \delta(S) < \delta_{\rm cr} + \delta'\Delta S , \tag{5.81}$$

where a prime denotes a derivative with respect to S. For this condition to make sense, we must have $\delta' \geq 0$.

Therefore, when identifying the fraction of walks which *up-cross* $\delta_{\rm cr}$ in the range $(S, S + \Delta S)$ as the fraction of Lagrangian volume enclosed in halos of the corresponding mass M, the multiplicity function is given by

$$f(S)\Delta S = \int_0^\infty d\delta' \int_{\delta_{\rm cr}}^{\delta_{\rm cr}+\delta'\Delta S} d\delta \, P(\delta, \delta') \approx \Delta S \int_0^\infty d\delta' \, \delta' \, P(\delta = \delta_{\rm cr}, \delta') . \tag{5.82}$$

---

*This section is of a more technical nature and is not essential for the remainder of the review. Readers can go directly to Sec. 5.7.3 where the halo bias obtained from this method is discussed.

Here, the joint probability distribution of δ and δ' is given by

$$P(\delta, \delta') = \frac{\gamma}{\pi\sqrt{1-\gamma^2}} \exp\left[-\frac{2\gamma^2}{1-\gamma^2}\left\{\frac{\delta^2}{4S\gamma^2} - \delta\delta' + S\delta'^2\right\}\right], \tag{5.83}$$

where

$$\gamma^2 \equiv \frac{[\langle\delta(S)\delta'(S)\rangle]^2}{\langle[\delta(S)]^2\rangle\langle[\delta'(S)]^2\rangle}, \tag{5.84}$$

and we have $\langle\delta(S)\delta'(S)\rangle = 1/2$ since $\delta'(S)$ is the derivative of $\delta(S)$ with respect to $S = \langle\delta^2(R[S])\rangle$. Then, the conditional probability $P(\delta'|\delta_{\mathrm{cr}})$ has mean $\langle\delta'|\delta_{\mathrm{cr}}\rangle = \delta_{\mathrm{cr}}/2S$ and variance $\mathrm{Var}(\delta'|\delta_{\mathrm{cr}}) = 1/(4S\Gamma^2)$, where $\Gamma^2 \equiv \gamma^2/(1-\gamma^2)$. The integral in Eq. (5.82) can, then, be easily performed so that the first-crossing distribution reads

$$Sf(S) = \frac{e^{-\delta_{\mathrm{cr}}^2/2S}}{2\sqrt{2\pi S}}\delta_{\mathrm{cr}}\left[\frac{1+\mathrm{erf}\left(\Gamma\delta_{\mathrm{cr}}/\sqrt{2S}\right)}{2} + \frac{e^{-\Gamma^2\delta_{\mathrm{cr}}^2/2S}}{\sqrt{2\pi/S}\,\Gamma\delta_{\mathrm{cr}}}\right]. \tag{5.85}$$

The bias parameters can be correspondingly computed by considering the first crossing distribution with the additional constraint that the excursion-set random walks went through $\delta_\ell \equiv \delta(S_\ell)$ on the large scale $S_\ell \ll S$, analogously to the Markovian case (Sec. 5.4). The conditional first crossing can be calculated similarly to Eq. (5.82) except that we now need to integrate the conditional probability function $P(\delta, \delta'|\delta_\ell)$

$$f(S|\delta_\ell; S_\ell)\Delta S = \int_0^\infty d\delta' \int_{\delta_{\mathrm{cr}}}^{\delta_{\mathrm{cr}}+\delta'+\Delta S} d\delta P(\delta, \delta'|\delta_\ell) \approx \Delta S\, P(\delta_{\mathrm{cr}}|\delta_\ell)\int_0^\infty d\delta'\, \delta' P(\delta'|\delta_{\mathrm{cr}}, \delta_\ell). \tag{5.86}$$

Upon taking the limit $\Delta S \to 0$, we obtain

$$f(S|\delta_\ell; S_\ell) = P(\delta_{\mathrm{cr}}|\delta_\ell)\int_0^\infty d\delta'\, \delta'\, P(\delta'|\delta_{\mathrm{cr}}, \delta_\ell) \tag{5.87}$$

for the constant deterministic barrier that we are considering here. In order to obtain the conditional probability, we use that the joint probability of $(\delta, \delta', \delta_\ell)$ is a trivariate Gaussian [305] specified by the parameters that we have defined earlier and the variance of the long-wavelength fluctuations $S_\ell \equiv \langle\delta_\ell^2\rangle = \langle\delta^2(R_\ell)\rangle$, the covariance of long- and short-wavelength fluctuations, $S_\times \equiv \langle\delta(S)\delta(S_\ell)\rangle$, and the cross-correlation coefficient between $\delta'(S)$ and $\delta(S_\ell)$, $\epsilon_\times = 2S\langle\delta'(S)\delta(S_\ell)\rangle/S_\times$. Then, the mean and variance of the conditional Gaussian distribution function $P(\delta'|\delta_{\mathrm{cr}}, \delta_\ell) = P(\delta', \delta_{\mathrm{cr}}|\delta_\ell)/P(\delta_{\mathrm{cr}}|\delta_\ell)$ are

$$\bar{\delta}'(R) \equiv \langle\delta'(R)|\delta_{\mathrm{cr}}, \delta_\ell\rangle = \frac{1}{2SQ}\left[\delta_{c\times} + \epsilon_\times\frac{S_\times}{S_\ell}\left(\delta_\ell - \delta_{\mathrm{cr}}\frac{S_\times}{S_\ell}\frac{S_\ell}{S}\right)\right], \tag{5.88}$$

and

$$\bar{\sigma}^2(R) \equiv \mathrm{Var}\left(\delta'(R)|\delta_{\mathrm{cr}}, \delta_\ell\right) = \frac{1}{4S\Gamma^2}\left[1 - \frac{\Gamma^2 S_\ell}{QS}\frac{S_\times^2(1-\epsilon_\times)^2}{S_\ell^2}\right], \tag{5.89}$$

with

$$\delta_{c\times} \equiv \delta_{\mathrm{cr}} - \frac{S_\times}{S_\ell}\delta_\ell, \qquad Q \equiv 1 - \frac{S_\times^2}{SS_\ell} = 1 - \left(\frac{S_\times}{S_\ell}\right)^2\frac{S_\ell}{S}. \tag{5.90}$$

Finally, integrating over Gaussian conditional distribution function $P(\delta'|\delta_{\mathrm{cr}}, \delta_\ell)$ yields the conditional first-crossing probability distribution function as [305]

$$\begin{aligned}
Sf(S|\delta_\ell; S_\ell) &= SP(\delta_{\mathrm{cr}}|\delta_\ell)\int_0^\infty d\delta'\delta' P(\delta'|\delta_{\mathrm{cr}}, \delta_\ell) \\
&= S\frac{\bar{\delta}' e^{-\delta_{c\times}^2/2SQ}}{\sqrt{2\pi SQ}}\left[\frac{1+\mathrm{erf}\left(\bar{\delta}'(S)/\sqrt{2}\bar{\sigma}(S)\right)}{2} + \frac{e^{-\bar{\delta}'^2(S)/2\bar{\sigma}^2(S)}}{\sqrt{2\pi}(\bar{\delta}'(S)/\bar{\sigma}(S))}\right].
\end{aligned} \tag{5.91}$$

126

Note that explicit expressions for $S_\times$ and $\epsilon_\times$ can be derived for e.g. power-law spectra with Gaussian filter. In the case of the sharp k-space filter, $S_\times/S_\ell = 1$ and $\epsilon_\times = 0$. Although $f(S|\delta_\ell; S_\ell)$ in Eq. (5.91) remains positive even when $\delta_\ell > \delta_{\mathrm{cr}}$, we need to impose the condition $\delta_\ell \le \delta_{\mathrm{cr}}$ in order to satisfy the "first crossing" criterion.

5.7.3 Halo mass function and bias

The first-crossing PDF in Eq. (5.85) can be directly translated to the halo mass function through

$$\bar{n}_h(M) = \frac{\bar{\rho}_m}{M} 2Sf(S) \left| \frac{d\ln\sigma}{d\ln M} \right| = \frac{\bar{\rho}_m}{M} \frac{e^{-\delta_{\mathrm{cr}}^2/2S}}{\sqrt{2\pi S}} \delta_{\mathrm{cr}} \left[\frac{1 + \mathrm{erf}\left(\Gamma\delta_{\mathrm{cr}}/\sqrt{2S}\right)}{2} + \frac{e^{-\Gamma^2\delta_{\mathrm{cr}}^2/2S}}{\sqrt{2\pi/S}\,\Gamma\delta_{\mathrm{cr}}} \right] \left| \frac{d\ln\sigma}{d\ln M} \right| . \qquad (5.92)$$

Here, $\Gamma^2 = \gamma^2/(1-\gamma^2)$ and γ is the cross-correlation coefficient between $\delta(S)$ and $\delta'(S)$ defined in Eq. (5.84). This mass function is in general not normalized to unity owing to the imperfect correlation between δ and δ'. Comparison with Eq. (5.80) shows that the square bracket in Eq. (5.85) is the leading-order correction to the fully correlated solution, which is attained in the limit $\Gamma \to \infty$ i.e. when δ and δ' are completely correlated ($\gamma = 1$). Note that this limit is hardly achieved in reality; for a Gaussian filter and a power-law power spectrum $P_{\mathrm{L}}(k) \propto k^n$, $\Gamma^2 = (3+n)/2$ for example, and the limit can only be achieved when $n \to \infty$. Therefore, the first-crossing distribution of partially correlated walks is fundamentally different than that of either entirely correlated or fully Markovian walks. Surprisingly, Monte-Carlo realizations of the first-crossing distributions indicate that the approximation Eq. (5.92) works very well over a large range of scales for a range of choices of power spectrum and filters (including tophat-filtered ΛCDM spectra) [153]. Note that, for "moving barriers" $\delta_{\mathrm{cr}} \to B(S)$, the halo mass function is not simply given by Eq. (5.92) with δ_{cr} replaced by $B(S)$ because the derivative with respect to S in Eq. (5.81) also brings a factor of $B'(S)$ which changes the lower limit of the integral in Eq. (5.82). We will discuss this in Sec. 5.9.

We then calculate the halo bias parameters from the conditional first-crossing probability distribution function in Eq. (5.91):

$$\bar{n}_h(M|\delta_\ell; S_\ell) = \frac{\bar{\rho}_m}{M} \sqrt{\frac{2S}{\pi}} \frac{\bar{\delta}' e^{-\delta_{c\times}^2/2SQ}}{\sqrt{Q}} \left[\frac{1 + \mathrm{erf}\left(\bar{\delta}'(S)/\sqrt{2}\bar{\sigma}(S)\right)}{2} + \frac{e^{-\bar{\delta}'^2(S)/2\bar{\sigma}^2(S)}}{\sqrt{2\pi}(\bar{\delta}'(S)/\bar{\sigma}(S))} \right] \left| \frac{d\ln\sigma}{d\ln M} \right| , \qquad (5.93)$$

where we refer to Sec. 5.7.2 for the definition of the symbols. Because in this case the first-crossing probability depends on two distinct variables δ and $\delta' \equiv d\delta/dS$, the halo bias is bivariate and we shall denote the corresponding (Lagrangian) bias parameters as b_{ij}^L. Note that δ' is a higher-derivative operator, and proportional to $R^2\nabla^2\delta$ at leading order in derivatives (see Sec. 2.6). Thus, the Lagrangian LIMD biases are given by b_{n0}^L, while the b_{ni}^L with $i > 0$ are higher-derivative biases. We see that the Lagrangian halo density δ_h^L predicted by this ansatz can be expanded as written in Eq. (5.1).

When the peak-background split is implemented with conditional mass functions, one considers the limit $R_\ell \gg R$ or, equivalently, $S_\ell \to 0$ at fixed value of δ_ℓ [15, 16]. Particular attention must be paid to the ratio $S_\times/S_\ell$, which tends towards $\mathcal{O}(1)$ in this limit (the exact value depends on the shape of the filter). Hence, on taking the limit $R_\ell \to \infty$, all the corrections to $Q = 1$, $\bar{\sigma}^2 = (4S\Gamma^2)^{-1}$ vanish, except for $\delta_{c\times} = \delta_{\mathrm{cr}} - \delta_\ell(S_\times/S_\ell) \approx \delta_{\mathrm{cr}} - \delta_\ell$, which implies $\bar{\delta}' \to (\delta_{\mathrm{cr}} - \delta_\ell S_\times/S_\ell)/(2SQ)$. In other words, the result is the same as differentiating w.r.t. δ_{cr}, and one recovers the linear Lagrangian LIMD bias parameter from the peak-background split, Eq. (3.26). Explicitly, for the constant barrier considered here, this yields

$$b_{10}^L \equiv -\left[\nu_c f(\nu_c)\right]^{-1} \frac{\partial\left[\nu_c f(\nu_c)\right]}{\partial\delta_{\mathrm{cr}}} = \frac{\nu_c^2 - 1}{\delta_{\mathrm{cr}}} + \frac{1}{\delta_{\mathrm{cr}}} \left[1 + \frac{1 + \mathrm{erf}(\Gamma\nu_c/\sqrt{2})}{2} \frac{\sqrt{2\pi}\Gamma\nu_c}{e^{-\Gamma^2\nu_c^2/2}} \right]^{-1} \qquad (5.94)$$

$$b_{20}^L \equiv \left[\nu_c f(\nu_c)\right]^{-1} \frac{\partial^2\left[\nu_c f(\nu_c)\right]}{\partial\delta_{\mathrm{cr}}^2} = \frac{\nu_c^4 - 3\nu_c^2}{\delta_{\mathrm{cr}}^2} + \frac{2 + \gamma^2}{s} \left[1 + \frac{1 + \mathrm{erf}(\Gamma\nu_c/\sqrt{2})}{2} \frac{\sqrt{2\pi}\Gamma\nu_c}{e^{-\Gamma^2\nu_c^2/2}} \right]^{-1} , \qquad (5.95)$$

where we have inserted Eq. (5.85) for $\nu_c f(\nu_c) = 2Sf(S)$. The Lagrangian bias for completely correlated walks (which equals the Markovian result Eq. (5.62) since the corresponding multiplicity functions differ only by a constant factor of 2) is recovered in the high-peak limit $\nu_c \gg 1$.

Ref. [152] showed that, on simultaneously taking the limit $S_\times/\sqrt{SS_\ell} \to 0$ but retaining the dependence on $\epsilon_\times$, one can obtain higher-derivative biases such as $b_{\nabla^2\delta}$ from Eq. (5.93), generalizing the PBS approach to higher-derivative biases. This corresponds to the dependence of δ_h^L on other variables apart from δ, such as δ' or the curvature $\nabla^2\delta$. In this particular limit, we still find $Q \to 1$, $\bar{\sigma}^2 \to (4S\Gamma^2)^{-1}$ and $\delta_{c\times} \to \delta_{\rm cr} - \delta_\ell(S_\times/S_\ell)$. However, we now have

$$\bar{\delta}' \to \frac{1}{2s}\left[\delta_{\rm cr} - \frac{S_\times}{S_\ell}\big(1 - \epsilon_\times\big)\delta_\ell\right].$$ (5.96)

Taking the ratio of the conditional to unconditional first-crossing distributions and expanding in powers of δ_ℓ, we eventually obtain

$$1 + \delta_h^L(\delta_\ell) = \frac{f(S|\delta_\ell; S_\ell)}{f(S)} = 1 + \frac{S_\times}{S_\ell}\big(b_{10}^L + \epsilon_\times b_{01}^L\big)\delta_\ell + \frac{1}{2}\left(\frac{S_\times}{S_\ell}\right)^2 \big(b_{20}^L + 2\epsilon_\times b_{11}^L + \epsilon_\times^2 b_{02}^L\big)\delta_\ell^2 + \dots$$ (5.97)

which differs from the Lagrangian LIMD expansion owing to the presence of higher-derivative bias contributions induced by δ'. These arise with non-zero powers of $\epsilon_\times$. The appearance of $\epsilon_\times$ in Eq. (5.97) indicates that the bias in Fourier space is $\propto k^2$. Namely, we have at linear order $\delta_h^L(\boldsymbol{k}) = c_1^L(k)\delta^{(1)}(\boldsymbol{k})$ with [305]

$$c_1^L(\boldsymbol{k}) = \left(b_{10}^L + \frac{k^2 S}{\langle[\delta'(S)]^2\rangle}b_{01}^L\right)W_R(k)$$ (5.98)

This shows that excursion-set theory generically predicts a k-dependence in Fourier space. The peak theory discussed in Sec. 6 and, more generally, any Lagrangian bias scheme more sophisticated than LIMD (e.g. [113]) leads to k-dependent Lagrangian bias functions $c_n^L(\boldsymbol{k}_1, \dots, \boldsymbol{k}_n)$. Crucially however, the k-dependence can be mapped exactly onto the general set of higher-derivative terms described in Sec. 2.6 in all cases.

We can read off both the LIMD bias parameters b_{i0}^L and the higher-derivative bias parameters b_{ij}^L ($j > 0$) from Eq. (5.97). In particular, we have

$$b_{01}^L = \frac{1}{\delta_{\rm cr}}\left\{1 - \left[1 + \frac{1 + {\rm erf}(\Gamma\nu_c/\sqrt{2})}{2}\frac{\sqrt{2\pi}\Gamma\nu_c}{e^{-\Gamma^2\nu_c^2/2}}\right]^{-1}\right\}.$$ (5.99)

Note that the b_{0j}^L, which represent the jth-order response of the first-crossing distribution to a long-wavelength perturbation δ_ℓ', are dimensionless because the derivative is taken with respect to the variance S. Adding Eq. (5.94) and Eq. (5.99), the first-order Lagrangian biases are linearly related through

$$b_{10}^L + b_{01}^L = \frac{\delta_{\rm cr}}{S}.$$ (5.100)

Similar relations hold at any order, as well as for more sophisticated Lagrangian prescriptions such as the peak constraint [305, 221]. They can be used to infer the value of the higher-derivative, or "scale-dependent", bias parameters once the LIMD, or "scale-independent" biases are known. These relations follow from the possibility of expanding the joint distribution of δ and δ' in Hermite polynomials. In fact, the b_{ij}^L can be written as an ensemble average over Hermite polynomials [99, 305, 430]. We will discuss polynomial expansions in more detail in Sec. 6.6.1, and show that this can be generalized to variables which do not follow Gaussian distributions.

5.8 Summary: bias of halos in the standard excursion-set formalism

So far in this section, we have reviewed the theoretical predictions for the linear and second-order LIMD bias parameters from the excursion-set formalism with a constant density threshold $\delta_{\rm cr}$. This approximation is motivated by the spherical collapse model (Sec. 5.2.1). Therefore, these results are built on a strongly simplified scenario of halo formation.

When employing this assumption, however, exact analytical solutions are known for two special cases of the excursion set: when each step of the random walk is (i) completely independent from the previous
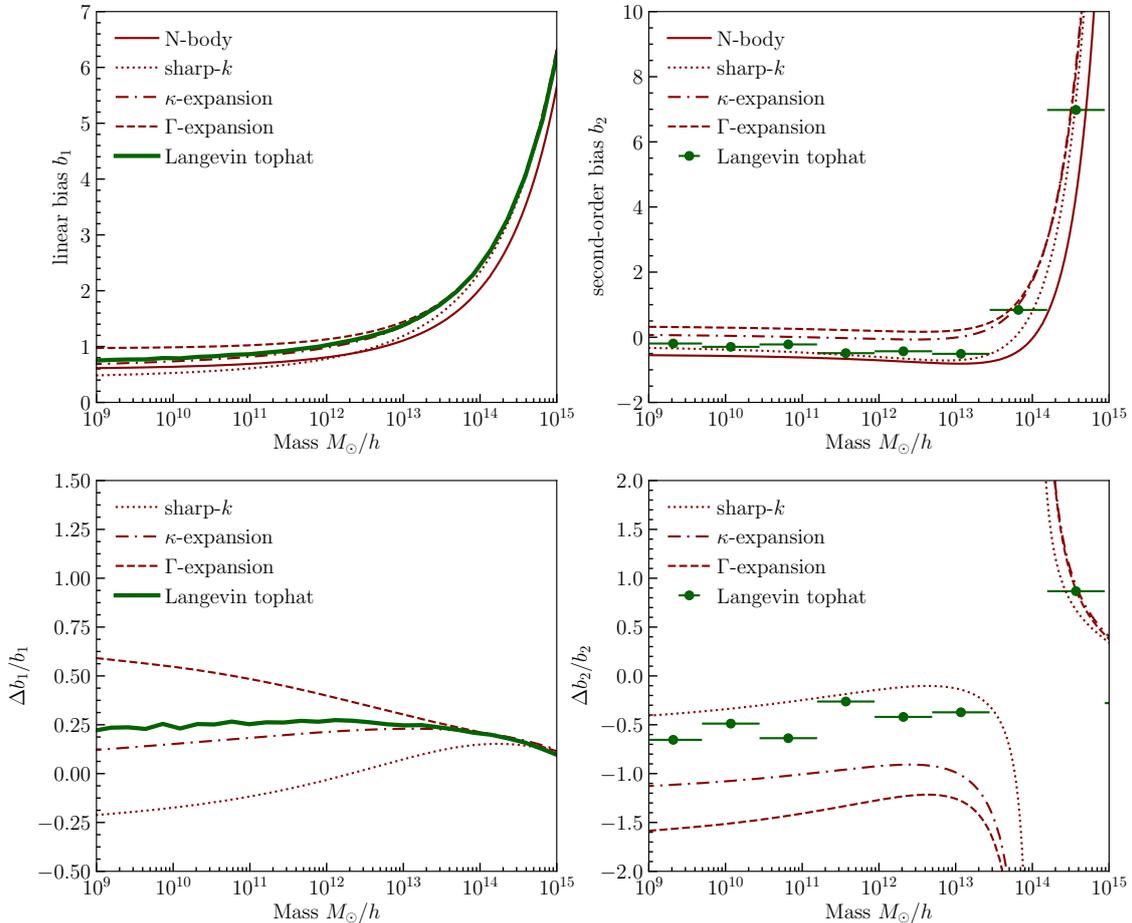
Figure 25: Excursion-set predictions for the Eulerian bias parameters b_1 (upper left panel) and b_2 (upper right panel) for halos at redshift $z = 0$ (red lines). Also shown is the result from N-body simulations, parametrized using the fitting functions given in Tab. 7 in Sec. 4 (using $\Delta_{SO} = 200$). The lower panels show the corresponding fractional deviation of the excursion-set predictions from the N-body results. We show four excursion-set predictions (all with $\delta_{cr} = 1.686$ and using $b_1 = 1 + b_1^L$ and $b_2 = b_2^L + 8/21 b_1^L$): (1) Markovian, or Press-Schechter case [Eqs. (5.62)–(5.63), dotted lines], (2) first order in the κ-expansion around the Markovian case [Eqs. (5.76)–(5.77), dot-dashed lines], (3) first order in the Γ expansion around the fully correlated case [Eqs. (5.94)–(5.95), dashed lines], and (4) the numerical integration result of the exact Langevin equation with a tophat filter [Eq. (5.47); thick solid line]. We have integrated the Langevin equation to obtain 5 billion independent random-walk chains, each with 2000 steps from $R = 10^{3.5}\,h^{-1}\,\mathrm{Mpc}$ to $10^{-1.5}\,h^{-1}\,\mathrm{Mpc}$. An accurate determination of b_2 demands even more random-walk chains, however. Note that, even with the numerical solution of the Langevin equation, the excursion-set predictions do not agree well with the bias parameters measured from N-body simulations.

steps (Markovian case obtained by employing the sharp-k filter, Sec. 5.6.1), or is (ii) completely correlated with the previous steps (Sec. 5.7.1). Both cases are far from being realistic descriptions of Lagrangian halo density profiles. For more general filters, approximate solutions can be obtained by expanding around these analytic solutions, as shown in Sec. 5.6.3 and Sec. 5.7.3.

Specifically, the predictions for the linear and second-order bias parameters are given in:

- Eqs. (5.62)–(5.63) for the Markovian treatment (sharp-k filter, or Press-Schechter)

- Eqs. (5.76)–(5.77) for the expansion around the Markovian case (i.e. $0 < \kappa < 1$)

- Eqs. (5.94)–(5.95) for the expansion around the fully correlated case ($\Gamma \gg 1$).

For both expansions around analytical solutions, we recover a multiplicity function that is proportional to that of the Markovian case in certain limits, namely $\kappa \to 0$ and $\Gamma \to \infty$, respectively. Note however that they describe expansions around two completely different solutions. We compare these three predictions in Fig. 25, assuming a spherical tophat filter for $\sigma(M)$, and a constant density threshold $\delta_{\rm cr} = 1.686$. Further, we choose $\kappa = 0.44$ in Eqs. (5.76)–(5.77). For Eqs. (5.94)–(5.95), we calculate Γ from our reference cosmology (Tab. 1) which gives $\Gamma = [0.35...0.57]$ in the mass range plotted in Fig. 25. We also show the result of a numerical integration of the Langevin equation Eq. (5.47) for the tophat filter. For both b_1 and b_2, the κ-expansion follows the exact numerical result more closely than the Γ-expansion. For both b_1 and b_2, all four predictions agree for high-mass halos, but diverge for lower-mass halos. Note, however, that the excursion-set predictions do not agree well with the bias parameters measured from N-body simulations, which are represented in Fig. 25 using the empirical fitting formulas given in Tab. 7 (Sec. 4.5). This also holds for the numerical integration of the Langevin equation, which corresponds to the exact solution of the excursion set. This result illustrates the need for a more realistic description of the collapse, to which the spherical model is a crude approximation. We turn to extensions such as the ellipsoidal collapse model and stochastic barriers next.

5.9 Beyond the spherical collapse model

For an initially spherical perturbation, the spherical collapse solution is a good approximation until the first orbit crossing. However, actual overdense regions in a Gaussian density field are not spherical, but triaxial [431, 13, 432]. Moreover, the traceless part of $\partial_i \partial_j \Phi$, i.e. the tidal shear K_{ij}, has been shown to play a crucial role in the formation of nonlinear structures [433, 434, 435, 436]. N-body simulations have established that the principal frame of the proto-halo strongly correlates with the local tidal field [437, 393, 438, 394, 439, 440]. The collapse of an isolated, homogeneous ellipsoid has been studied extensively [441, 442, 443, 444, 445], although these are still only approximations to actual Lagrangian proto-halo patches. In the formulation of [435], the initial conditions and external tides are chosen to recover the Zel'dovich approximation in the linear regime. For a given fixed Lagrangian overdensity, the final density can attain a range of values depending on the local tidal shear, as estimated through ellipsoidal collapse calculations in [446, 447].

In a first approximation, the dynamics of ellipsoidal collapse can be incorporated through a mass-dependent, or "moving" barrier $B(S)$ in the excursion-set approach. The tidal shear effectively slows down the collapse of low-mass objects and, therefore, yields a barrier $B(S)$ which grows monotonically with S. As a result, the relative abundance of high-mass objects increases [448], such that the multiplicity function $f(\nu_c)$ furnishes a better fit to the halo mass function measured in N-body simulations [224].

In practice, whereas the spherical evolution leads to a condition on the trace δ of the deformation tensor only, the critical density for non-spherical collapse will also depend on the other two invariants of $\partial_i \partial_j \Phi$ (equivalently, $\mathrm{tr}[(K_{ij})^2]$ and $\mathrm{tr}[(K_{ij})^3]$). Following [136, 121, 196], the three invariants of the second-rank tensor $\partial_i \partial_j \Phi$ can be conveniently parametrized as (see Sec. 2.2)

$$\delta \,, \qquad K_2 \equiv \frac{3}{2}\,\mathrm{tr}\left(K_{ij}^2\right)\,, \qquad K_3 \equiv \frac{9}{2}\,\mathrm{tr}\left(K_{ij}^3\right)\,. \tag{5.101}$$

In K_2, a multiplicative factor of $3/2$ is added such that $\langle K_2 \rangle = S$, which implies that, for Gaussian initial conditions, $5K_2$ is χ^2-distributed with 5 d.o.f.. Similarly, the factor of $9/2$ in the definition of K_3 ensures that $|K_3| \leq K_2^{3/2}$. Note that $K_2 = 3e^2 + p^2$, where e and p are the shear ellipticity and prolateness [223]. Ellipsoidal collapse will then give rise to a dependence of δ_h^L on K_2, K_3 of the form (see also Sec. 2.4)

$$\delta_h^L(\boldsymbol{q}) = b_1^L \delta + \frac{1}{2}b_2^L \delta^2 + b_{K_2}^L K_2 + \frac{1}{6}b_3^L \delta^3 + \frac{5}{\sqrt{7}}b_{K_3}^L K_3 + b_{\delta K_2}^L \delta K_2 + \ldots \,. \tag{5.102}$$

This generalizes the Lagrangian LIMD bias obtained in the standard excursion set, as summarized in Eq. (5.1), to include the local tidal terms induced by non-spherical collapse. Specifically, the bias parameters introduced in Eq. (5.102) are related to those defined in Sec. 2.4 through

$$b_{K_2}^L = \frac{2}{3}b_{K^2}^L \, ; \quad b_{\delta K_2}^L = \frac{2}{3}b_{\delta K^2}^L \, ; \quad b_{K_3}^L = \frac{2\sqrt{7}}{45}b_{K^3}^L \, . \tag{5.103}$$

In order to calculate the Lagrangian tidal bias parameters $b_{K_2}^L$, $b_{K_3}^L$, and so on, we need a model for the barrier B, which now characterizes the first-crossing of multidimensional random walks $(\delta, K_2, \dots)$ [449, 224, 450, 451, 326]. That is, a complete theory of Lagrangian bias must take into account the dependence of the galaxy number density on $K_{2\ell}$ and $K_{3\ell}$ separately, but this has not been done yet. As a rule of thumb, any additional variable adds a dimension to the first-crossing problem. For illustration, Ref. [121] considered the simple model

$$B(S, K_2) = \delta_{\mathrm{cr}} \left(1 + \sqrt{K_2/K_c} \right) \, , \tag{5.104}$$

which is motivated by setting $p = 0$ in the moving barrier of [223]. Here, K_c is a characteristic scale for the effect of tidal shear. The barrier depends on S, but only through $K_2 \propto S$. Furthermore, we always have $B \geq \delta_{\mathrm{cr}}$ since $K_2 \geq 0$. In analogy with our previous calculation of the Lagrangian LIMD bias b_n^L, the tidal bias $b_{K_2}^L$ can be obtained by considering the response of the unconditional first-crossing distribution to a long-wavelength perturbation $K_{2\ell}$. The conditional distribution $f(S|\delta_\ell, K_{2\ell}; S_\ell)$ now describes the first-crossing distribution of 2-dimensional walks that start from some non-zero δ_ℓ and $K_{2\ell}$. Of course, if the shape of the overdensity also affects the collapse, then the barrier would also depend on the misalignment between the tidal shear and shape tensor [326].

In practice, a simpler approach, in which K_2 is replaced by its mean value $\langle K_2 \rangle = S$, provides a good approximation to the exact result [224]. In this approximation, the barrier becomes $B(S, K_2) \to B(S) \approx \delta_{\mathrm{cr}}(1 + \sqrt{S/K_c})$, such that $f(S|\delta_\ell, K_{2\ell}; S_\ell)$ is well represented by one-dimensional walks crossing the barrier

$$\delta_{\mathrm{cr}} \left[1 + \sqrt{(S + K_{2\ell})/K_c} \right] - \delta_\ell \, . \tag{5.105}$$

Expanding this barrier in a series of powers of δ_ℓ and $K_{2\ell}$ and dividing by the unconditional first-crossing distribution, we arrive at

$$\delta_{\mathrm{cr}}^2 b_{K_2}^L \approx -\nu_c^3 \frac{\delta_{\mathrm{cr}}}{K_c} \left(1 + \frac{\delta_{\mathrm{cr}}}{\nu_c K_c} \right) \, , \tag{5.106}$$

in the limit $\nu_c \gg 1$. Unsurprisingly, $b_{K_2}^L < 0$ since a large shear increases the barrier height and, therefore, impedes the collapse of halos. Furthermore, $b_{K_2}^L \propto -\frac{\sigma_0}{K_c}b_2^L$ in the high-peak limit, so that the spherical collapse approximation—within which only the LIMD bias parameters are non-zero in Lagrangian space—is recovered. Higher-order local bias parameters involving K_2 can be computed analogously. Evidence for a non-zero $b_{K^2}^L$ has recently been presented in the literature (see the discussion in Sec. 4.5.1).

Numerical implementations of the ellipsoidal collapse [e.g., 223, 452, 453] and numerical studies of Lagrangian halos [e.g., 318, 394] have shown that the collapse barrier is "fuzzy", i.e. there is a range of values of δ_{cr} at fixed S. This is apparent in Fig. 26, which shows the distribution of smoothed linear overdensities associated with Lagrangian regions collapsing into halos [318]. Namely, for each halo identified in N-body simulations, one can trace its constituent dark matter particles back to their initial position, and compute the initial overdensity in a tophat sphere centered at the initial center-of-mass. These overdensities, once linearly extrapolated to $z = 0$, furnish a snapshot of the collapse barrier. As seen in Fig. 26, the mean as well as median barriers increase with decreasing halo mass, in broad agreement with the ellipsoidal collapse prediction. The scatter Σ around the mean barrier is generated by variables other than the density such as the local shear, coupling with the large-scale environment. Note that Σ is proportional to $\sigma(M)$, which is consistent with a lognormal distribution of the barrier.

Thus, an alternative to explicitly including the tidal shear in the barrier is to use a "fuzzy" moving barrier, which includes scatter around the mean, such as the square-root barrier [419]

$$B(S) = \delta_{\mathrm{cr}} \left(1 + \beta \sqrt{S}/\delta_{\mathrm{cr}} \right) , \tag{5.107}$$
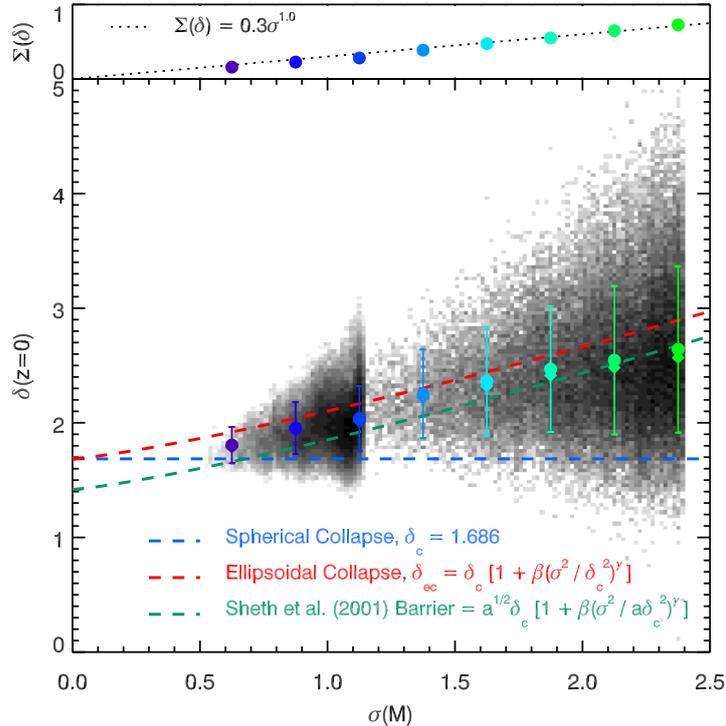
Figure 26: Smoothed linear overdensity, extrapolated to $z = 0$ for the Lagrangian regions that collapse and form halos by $z = 0$, as a function of halo mass parametrized through $\sigma(M)$. Circles correspond to the mean overdensities, whereas diamonds show the median overdensities. The errorbars indicate the halo-to-halo scatter. Shown for comparison are the spherical collapse barrier (blue dashed line), the ellipsoidal collapse barrier of [223] (red dashed line), and the collapse barrier associated with the Sheth-Tormen mass function [16] (green dashed line). The upper panel shows the scatter Σ in the barrier height as a function of $\sigma(M)$. From [318].

where β can be calibrated with N-body simulations using the method of [318] discussed above. For the purpose of predicting the halo mass function $\overline{n}_h(M)$ and LIMD bias parameters b_N, this approximation is more convenient than introducing an explicit dependence on the shear K_{ij} and other fields. The results in Sec. 5.6–5.7 can indeed be extended to this barrier. Moreover, this ansatz is adopted in the excursion-set peak approach described in Sec. 6.8. Alternatively, the scatter has also been modeled as a "diffusive barrier" in the context of the path integral approach described in Sec. 5.6.3 [454]. The resulting mass functions have been found to be in good agreement with simulations [417, 418].

One should bear in mind that the scatter in the barrier reflects the existence of hidden variables and, therefore, is a consequence of projecting the actual, multi-dimensional collapse barrier onto a simpler subspace in which S is the sole variable. A moving barrier $B(S)$ cannot fully capture the dependence of halo collapse on the tidal shear (or deformation) tensor K_{ij} unless the latter is explicitly accounted for. As we shall see in Sec. 7.4, a microscopic description of the barrier B in terms of K_{ij} is, in fact, essential to obtain a physically consistent prediction for the non-Gaussian bias in Lagrangian bias models.

5.10   Halo assembly bias in the excursion-set formalism

As discussed in Sec. 4.6, studies of halo clustering in numerical simulations have established that, at fixed halo mass, the halo bias depends on other properties such as formation time, concentration, or environment density, a phenomenon known as assembly bias. In the high-mass limit and for halo properties that can be related to Lagrangian quantities, the excursion set provides a physical picture for the origin of assembly bias, as we will now discuss. This holds in particular for the environment density and formation time.
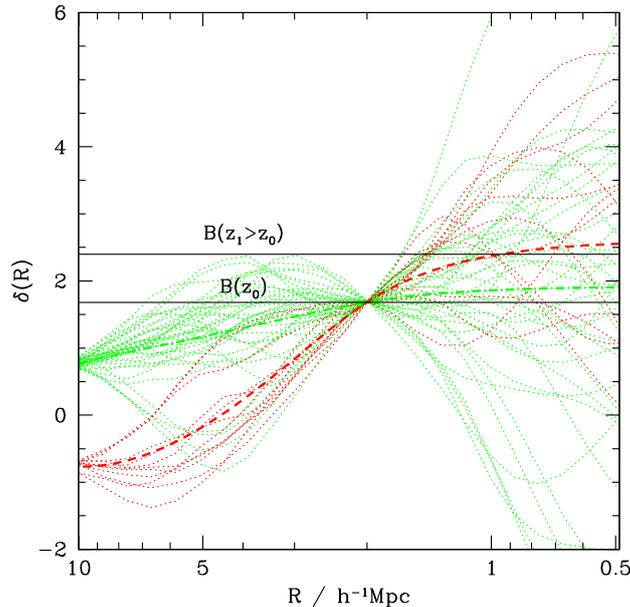
Figure 27:   Trajectories $\delta(R)$ as a function of the tophat smoothing radius R, extracted from random realizations of the linear density field. All the trajectories obey the constraint $\delta(R) = \delta_{cr}$ at the halo scale. The green and red curves show the realizations which have $\delta(R_{env}) = +2\sigma_{env}$ and $-2\sigma_{env}$, respectively. The thick dashed and dot-dashed curve represent the average trajectory $\bar{\delta}(R|\delta_{env})$. Here, we have used $R_{env} = 5\,h^{-1}$ Mpc.

Assembly bias cannot be explained in the standard Markovian formulation of excursion-set theory (see Sec. 5.6.1) since there is no correlation between the largest scales which characterize the clustering, and the small scales which characterize the accretion history [455]. Of course, correlations will arise as soon as the Markovian assumption is relaxed.

As shown in [456] and [367], the trend at high halo mass can be explained with the statistics of correlated random walks $\delta(s)$. To see this, we follow common practice and adopt the redshift $z_{1/2}$ at which one half of the mass has been accreted onto the halo as a proxy for the halo formation redshift. Simple considerations suggest a correlation between the large-scale overdensity δ_{env} and the slope δ' [456, 452]. A peak constraint as invoked in [367] leads to similar conclusions since the peak curvature J_1 (see Sec. 6.3)) strongly correlates with δ' (J_1 and δ' are, in fact, completely correlated for a Gaussian filter, see Sec. 6.8). Namely, when the value of δ is held fixed to $\delta = \delta_{cr}$ at the halo mass scale, we find:

$$\begin{aligned} \delta_{env} > 0 &\quad\Longleftrightarrow\quad \text{smaller } \delta' \quad\Longleftrightarrow\quad \text{low } z_{1/2} \\ \delta_{env} < 0 &\quad\Longleftrightarrow\quad \text{larger } \delta' \quad\Longleftrightarrow\quad \text{high } z_{1/2}\,. \end{aligned} \tag{5.108}$$

To illustrate this point, we have extracted trajectories $\delta(R)$ smoothed with a tophat filter, as shown in Fig. 27. All the trajectories are subject to the two constraints $\delta(R) = \delta_{cr}$ on the halo scale, and $\delta(R_{env}) = \pm 2\sigma_{env}$ [$R = 1.5\,h^{-1}$ Mpc and $R_{env} = 5\,h^{-1}$ Mpc for illustration]. The solid and dashed thick curves show the expected "mean" field $\bar{\delta}(R|\delta_{env})$ subject to the large-scale constraint. The horizontal lines represent the collapse barrier $B(z) = \delta_{cr}D(z_0)/D(z)$ at redshifts z_0 and $z_1 > z_0$, where z_0 is the redshift at which the linear density field $\delta(R)$ is extrapolated. Clearly, the steeper the slope $|\delta'(R)|$, the higher $z_{1/2}$ is, since, at constant final mass M (i.e. R), the collapse barrier $B(z)$ is reached relatively earlier. Furthermore, these early-forming halos reside in more isolated environment (underdense regions with $\delta_\ell < 0$) with less late-time accretion, which leads to steeper halo profiles or higher concentration. The converse holds for the late-forming halos which preferentially collapse in regions with $\delta_\ell > 0$.

Now, $\delta_{env} > 0$ corresponds to halos with a larger bias. An easy way to see this is to re-express $b_1^L \equiv b_{10}^L$

133

defined in Eq. (5.94) as

$$b_1^L = \frac{\nu^2}{\delta_{\rm cr}(1 - \gamma^2)} - 2\Gamma^2 \bar{\delta}' \, , \tag{5.109}$$

where $\bar{\delta}'$ is the average of the slope *given the threshold and first-crossing constraints* $\delta \equiv \delta_{\rm cr}$, $\delta' \geq 0$. When the large scale overdensity $\delta_{\rm env}$ is fixed, then $\bar{\delta}'$ is replaced by some $\delta' < \bar{\delta}'$ if $\delta_{\rm env} > 0$, and $\delta' > \bar{\delta}'$ if $\delta_{\rm env} < 0$. Consequently, b_1^L is larger than average if $\delta_{\rm env} > 0$, and smaller if $\delta_{\rm env} < 0$. We thus reach the conclusion that "young" halos (i.e., those with low $z_{1/2}$) have lower concentration and are more strongly clustered than "old" halos (i.e., those with high $z_{1/2}$). This prediction is consistent with the N-body results for high-mass halos. At low mass ($M \lesssim M_\star$), however, the assembly bias seen in N-body simulations is reversed, such that low-concentration halos are more biased (see Sec. 4.6).

The excursion-set formalism can also be adapted to investigate the impact of the anisotropic cosmic web on assembly bias. For instance, Ref. [457] investigated the effect of tides generated by filaments—modeled as one-dimensional saddle points—on the properties of neighboring halos. In the excursion-set language, the presence of a saddle point adds a contribution of the form $\hat{r}_i \hat{r}_j K_{ij}$, where $\hat{\boldsymbol{r}}$ is the unit separation vector between the halo and the saddle, to the mean and covariance of δ and δ'. This quadrupolar term, which is negative along the filament axis and positive perpendicular to it, can be used to quantify the response of halo counts, accretion rate and formation times to the presence of a nearby filament.

6 The statistics and evolution of Lagrangian density peaks

The peaks approach to the clustering of dark matter halos is built upon the assumption that halos form from peaks in the initial Lagrangian density field [431, 458, 12, 459, 13]. This is clearly an oversimplification of the complicated process of halo formation. Nevertheless, peaks provide an interesting non-perturbative toy model of discrete tracers which exhibits several interesting features of the bias expansion of general tracers (Sec. 2), including nonlinear and higher-derivative bias as well as stochasticity. The fundamental quantity in peak theory is the set of local maxima of the smoothed density field; therefore, peaks define a point process. Since the evolved density field is highly nonlinear, the peak constraint is generally applied to the initial (Lagrangian) Gaussian density field, with the assumption that the most prominent peaks should be in one-to-one correspondence with luminous galaxies or massive halos in the Universe.

The formalism became popular after [12] demonstrated that the bias inherent to overdense regions of the Universe could explain the higher clustering amplitude of Abell clusters (see the discussion in Sec. 2.1). The first numerical and analytic investigations of density peaks of the initial, Gaussian density field have been performed in [459, 13]. Since then, their abundance, clustering and internal properties have been studied in detail. In particular, their profiles have been investigated in an attempt to predict internal properties of virialized halos such as their profiles and angular momentum [460, 433, 461, 462, 463, 464, 465, 466, 467, 468]. As this review is focused on large-scale bias, however, we shall not discuss this aspect here. Some of these results have been used to constrain the power spectrum of matter density perturbations [469, 470] (using the fact that the number density of peaks in a Gaussian random field is sensitive to the shape of the power spectrum of density perturbations); to model the column density distribution of the Lyman-α forest [471], the cosmic web [472] or the alignment of galactic spins [473]; to study velocity bias and redshift-space distortions [474, 475, 257] as well as assembly bias [367]; and to generate constrained realizations of the large-scale structure [476, 477]. Furthermore, the peak formalism has been applied to study the importance of higher-derivative bias on the BAO scale [156, 152], and to understand the interplay between bias and gravitational evolution [152, 174]. More recently, [396] have proposed an approach which combines peak theory with the excursion-set approach described in the previous section, while [196] have shown how the peak clustering statistics can be derived from a perturbative bias expansion.

The empirical evidence for the association of massive halos with Lagrangian density peaks is reviewed in Sec. 5.1. Briefly, while massive halos with $M \gg M_\star$ are, to a good approximation, in one-to-one correspondence with prominent ($\nu_c \gg 1$) initial density peaks—simply because they are very rare and their proto-halo patch is approximately spherical [13]—the association weakens as one goes to lower halo mass. In fact, overdense regions of a Gaussian density fields are inherently triaxial [431, 13, 432]. Furthermore,

as discussed in Sec. 5.9, the initial shear field, i.e. the tidal field K_R^{ij} smoothed on the peak scale, certainly plays an important role in the formation of nonlinear structures [433, 434, 478, 479]. Taking into account these complications would very likely extend the predictability of peak theory to lower halo mass. The peak-patch picture of [435, 465, 466] has taken steps in this direction by incorporating the ellipsoidal collapse, solving the cloud-in-cloud problem and relying on an exclusion algorithm to avoid overlap of peak-patches. However, this approach does not allow for analytical expressions of the large-scale clustering of peak patches, although it is guaranteed to fit into the general bias expansion of Sec. 2 on scales where perturbation theory applies. Moreover, we expect the same qualitative features to appear within the simple thresholded peaks approach. Therefore, we shall hereafter restrict ourselves to the spherical approximation, and ignore the finite extent of the peak-patches (except for the filtering kernel). We will discuss each of these issues in detail in the context of the analytical peak model. As we will see, many insights can be gained even with the simplifications made here.

The outline of this section is as follows:

- Sec. 6.1 to Sec. 6.4 review the basic formalism, introduce the peak constraint and present a simple derivation of the peak number density (1-point function).

- Sec. 6.5 reviews calculations of peak correlation functions, while Sec. 6.6 delineates the peak bias expansion and its connection with the peak-background split and renormalization. Sec. 6.7 illustrates how the LIMD as well as higher-derivative peak bias parameters can be measured from N-body simulations using cross-correlations at two scales (see Sec. 4.2).

- Sec. 6.8 presents the excursion-set peaks (ESP) approach, which merges ideas from excursion-set theory (see Sec. 5) with the peak formalism.

- Sec. 6.9 focuses on the gravitational evolution of Lagrangian peaks, with a particular emphasis on the peak velocity bias.

Throughout this section, the peak number densities are defined as comoving Lagrangian densities. The definition of symbols used in this section can be found in Tab. 10.

6.1 Spectral moments and characteristic scales

The fundamental ingredient in all calculations within the peak approach is the smoothed initial density field $\delta_R^{(1)}$, extrapolated to the collapse epoch τ_0 using linear theory, where τ_0 does not necessarily correspond to today's epoch. In this section, we will denote the smoothed linear density field as $\delta_R(\boldsymbol{q}) \equiv \delta_R^{(1)}(\boldsymbol{q})$, where $\boldsymbol{q}$ is the Lagrangian position. This is the same field as used in the thresholding example considered in Sec. 2.1. However, owing to the peak constraint, the clustering properties of density peaks also depend on the statistics of the first and second derivatives $\partial_i \delta_R(\boldsymbol{q})$ and $\partial_i \partial_j \delta_R(\boldsymbol{q})$ of the smoothed density field. Throughout, we assume δ_R to be a homogeneous Gaussian random field of zero mean.

It is convenient to introduce the normalized dimensionless variables $\nu \equiv \delta_R/\sigma_0(R)$ [to be distinguished from the peak significance $\nu_c \equiv \delta_{\rm cr}/\sigma_0(R)$], $\eta_i \equiv \partial_i \delta_R/\sigma_1(R)$ and $\zeta_{ij} \equiv \partial_i \partial_j \delta_R/\sigma_2(R)$ (Tab. 10), where $\sigma_n(R)$ are the spectral moments of the matter power spectrum smoothed on the scale R,

$$\sigma_n^2(R) \equiv \int_{\boldsymbol{k}} k^{2n} W_R^2(k) P_{\rm L}(k) \,, \tag{6.1}$$

where $W_R(k)$ is a spherically symmetric filtering kernel. Note that this includes $\sigma(R) \equiv \sigma_0(R)$ as a special case. Here and in what follows, we omit the dependence on conformal time, which should always be taken to be the time at which halos are identified. Further, we shall often drop the argument R when no confusion is possible. The window function $W_R(k)$ reflects the fact that the formation of a given, isolated dark matter halo is insensitive to small-scale (UV) matter density fluctuations with $k \gg 1/R$. The filter shape $W_R(k)$ and its corresponding filtering scale $R(M) \propto M^{1/3}$ can be measured upon tracing halos back to the initial conditions [e.g. 174, 480]. Numerical simulations indicate that $W_R(k)$ is more extended than a tophat, but

Quantity	Symbol
Zero-lag spectral moments	$\sigma_n(R)$ [Eq. (6.1)]
Correlation functions (non-zero lag spectral moments)	$\xi_\ell^{(n)}(R,r)$ [Eq. (6.4)]
Spectral shape parameters	$\gamma_n(R)$ [Eq. (6.3)]
Normalized smoothed density	$\nu(\boldsymbol{q}) \equiv \frac{1}{\sigma_0}\delta_R(\boldsymbol{q})$
Normalized smoothed density gradient	$\eta_i(\boldsymbol{q}) \equiv \frac{1}{\sigma_1}\partial_i\delta_R(\boldsymbol{q})$
Normalized smoothed density Hessian	$\zeta_{ij}(\boldsymbol{q}) \equiv \frac{1}{\sigma_2}\partial_i\partial_j\delta_R(\boldsymbol{q})$
Eigenvalues of $-\zeta_{ij}$	$\lambda_1, \lambda_2, \lambda_3$
Rotational invariants of ζ_{ij}	J_1, J_2, J_3
Peak curvature	$J_1 \equiv -\operatorname{tr}[\zeta_{ij}]$
Scaled, uniformly distributed invariant	$x_3 \equiv J_3/(J_2)^{3/2}$
Lagrangian number density of BBKS peaks	$n_{\mathrm{pk}}(\boldsymbol{q})$
Mean comoving BBKS peak number density	$\overline{n}_{\mathrm{pk}}$
Derivative of δ_R with respect to smoothing scale	$\mu_R = -\frac{d\delta}{dR}$
Lagrangian number density of ESP peaks	$n_{\mathrm{ESP}}(\boldsymbol{q})$
Mean comoving ESP peak number density	$\overline{n}_{\mathrm{ESP}}$
Lagrangian 2-point correlation function of peaks	$\xi_{\mathrm{pk}}^L(r)$
Orthogonal polynomials and their duals	$O_{\mathbf{n}}(\boldsymbol{w}), O_{\mathbf{n}}^\star(\boldsymbol{w})$
Polynomials corresponding to J_2 and J_3	$F_{lm}(5J_2, J_3)$
n-th order Lagrangian bias	$c_n^L(\boldsymbol{k}_1,\dots,\boldsymbol{k}_n)$
n-th order Eulerian bias	$c_n^E(\boldsymbol{k}_1,\dots,\boldsymbol{k}_n,\tau)$
Linear peak displacement	$\boldsymbol{s}_{\mathrm{pk}}(\boldsymbol{k},\tau)$
Linear peak velocity bias	$c_{v,\mathrm{pk}}(k)$
Linear peak velocity dispersion	$\sigma_{v,\mathrm{pk}}(\tau)$

Table 10: List of symbols used throughout Sec. 6. Note that J_1 is often denoted as u in the literature.

more compact than a Gaussian [367, 480]. In order for the peak constraint to be meaningful, the actual filtering kernel $W_R(k)$ used must be such that the convergence of the spectral moments up to $\sigma_2(R)$ is ensured. While this is not the case for the tophat filter, this is certainly true for the Gaussian filter and, apparently, also for the actual filter $W_R(k)$, which can be measured upon tracing back halos to the initial conditions [367, 174, 480]. In what follows, we will assume a Gaussian filter for simplicity, until we reach the excursion-set peaks, Sec. 6.8, where a tophat filter is used to smooth the density and a Gaussian is used to smooth its derivatives.

Following Ref. [13] (BBKS), characteristic scales can be defined by taking ratios of spectral moments. In particular,

$$R_n(R) \equiv \sqrt{3}\frac{\sigma_n(R)}{\sigma_{n+1}(R)} \tag{6.2}$$

defines an ordered sequence of characteristic lengths $R_0 \geq R_1 \geq R_2 \geq \dots$ for typical values of interest of R, due to the shape of the linear matter power spectrum $P_{\mathrm{L}}(k)$. The first two scales are the typical separation R_0 between zero-crossings of the density field, and the mean distance R_1 between stationary point [13, 472]. These are the only scales involved in the calculation of the peak correlation functions. For subsequent use,

we also introduce the dimensionless parameters

$$\gamma_n(R) = \frac{\sigma_n^2(R)}{\sigma_{n-1}(R)\sigma_{n+1}(R)} \tag{6.3}$$

which quantifies the range over which $R^{-3}(kR)^{2(n-1)}P_{\mathrm{L}}(k)W_R^2(k)$ is significant, i.e. whether it is sharply peaked or broad.

The analogous quantities to σ_n^2 at non-zero separation are defined as follows:

$$\xi_\ell^{(n)}(R, r) = \int_{\boldsymbol{k}} k^{2(n+1)} W_R^2(k)\, P_{\mathrm{L}}(k)\, j_\ell(kr)\ , \tag{6.4}$$

where $j_\ell(x)$ are spherical Bessel functions. As ℓ is increased at fixed r, these harmonic transforms become increasingly sensitive to the small scale power.

The auto- and cross-correlations of the fields $\eta_i(\boldsymbol{q})$, $\nu(\boldsymbol{q})$ and $\zeta_{ij}(\boldsymbol{q})$ can generally be decomposed into components with definite transformation properties under three-dimensional rotations. Explicit expressions can be found in [481, 156, 152].

6.2 The Kac-Rice formula

Let $\{\boldsymbol{q}_1, \boldsymbol{q}_2, \cdots, \boldsymbol{q}_p, \cdots\}$ be the Lagrangian positions of point particles such as centers of halos in some (finite or infinite) volume. The comoving Lagrangian density $n_g(\boldsymbol{q})$ of these point particles (for a brief introduction to point processes in a cosmological context, see [482]) is formally written as a sum of Dirac distributions

$$n_g(\boldsymbol{q}) = \sum_p \delta_D(\boldsymbol{q} - \boldsymbol{q}_p)\ . \tag{6.5}$$

As shown in [483, 484, 13], in the case of stationary points (maxima, minima and saddle points) of a random field, the density $n_{\mathrm{sp}}(\boldsymbol{q})$ can be entirely expressed in terms of ν and its derivatives. To see why this is the case, note that the density of stationary points, defined as points $\boldsymbol{q}_p$ where the gradient of the density field satisfies $\boldsymbol{\eta}_i(\boldsymbol{q}_p) = 0$, can be written as

$$n_{\mathrm{sp}}(\boldsymbol{q}) = \delta_D[\boldsymbol{\eta}(\boldsymbol{q})] \left| \frac{\partial \boldsymbol{\eta}(\boldsymbol{q}_p)}{\partial \boldsymbol{q}_p} \right|\ , \tag{6.6}$$

where the Jacobian ensures that one obtains a proper density following Eq. (6.5). Thus, we need to evaluate the derivative of the gradient $\boldsymbol{\eta}$ with respect to $\boldsymbol{q}$. Now, in the neighborhood of a stationary point located at $\boldsymbol{q}_p$, the gradient is given by

$$\eta_i(\boldsymbol{q}) = \frac{\sqrt{3}}{R_1}\, \zeta_{ij}(\boldsymbol{q}_p)\, (q - q_p)^j + \mathcal{O}\!\left((\boldsymbol{q} - \boldsymbol{q}_p)^2\right)\ . \tag{6.7}$$

Then, Eq. (6.6) can be rephrased as

$$n_{\mathrm{sp}}(\boldsymbol{q}) = \sum_p \delta_D(\boldsymbol{q} - \boldsymbol{q}_p) = \frac{3^{3/2}}{R_1^3}\, |\det \zeta_{ij}(\boldsymbol{q})| \delta_D[\boldsymbol{\eta}(\boldsymbol{q})]\ , \tag{6.8}$$

provided that the Hessian ζ_{ij} is invertible. The Dirac delta $\delta_D[\boldsymbol{\eta}(\boldsymbol{q})]$ ensures that all the extrema are included. This expression, known as the Kac-Rice formula [483, 484] holds for arbitrary smooth random fields.

Since we are only interested in counting the density maxima, we further require that $\zeta_{ij}(\boldsymbol{q}_p)$ be negative definite at the position $\boldsymbol{q}_p$ of the stationary point. Finally, one usually restricts the set to those maxima with a certain threshold height ν_c following [12, 13], as in the case of thresholding (Sec. 2.1). The *localized* number density of "BBKS peaks" of height ν_c then reads

$$n_{\mathrm{pk}}(\boldsymbol{q}) = \frac{3^{3/2}}{R_1^3}\, |\det \zeta_{ij}(\boldsymbol{q})| \delta_D[\boldsymbol{\eta}(\boldsymbol{q})]\, \Theta_H(\lambda_3) \delta_D[\nu(\boldsymbol{q}) - \nu_c]\ , \tag{6.9}$$

138

where Θ_H is the Heaviside step function and λ_3 is the smallest eigenvalue of $-\zeta_{ij}$. The determinant somewhat complicates calculations. To circumvent this difficulty, one could weight the peaks by the inverse of $|\det\zeta_{ij}|$ [485]. However, it is unclear how such a weighted field exactly relates to dark matter halos.

6.3 Rotational invariants and their distribution

Since clustering statistics of peaks are invariant under three-dimensional rotations, they depend only on the scalar functions that can be constructed from the independent variables ν, η_i, ζ_{ij} (each of which contains, respectively, one, three and six degrees of freedom). Furthermore, since the peak constraint does not induce a dependence on $\eta_i\zeta_{ij}\eta_j$, only five scalar functions are relevant: $\nu(\boldsymbol{q})$, $J_1(\boldsymbol{q}) = -\operatorname{tr}\left(\zeta_{ij}(\boldsymbol{q})\right)$, the essentially χ^2-distributed quantity $\eta^2(\boldsymbol{q}) = \sum_i \eta_i^2(\boldsymbol{q})$, and the two rotational invariants

$$J_2(\boldsymbol{q}) = \frac{3}{2}\operatorname{tr}\left(\bar{\zeta}_{ij}^2(\boldsymbol{q})\right) \qquad J_3(\boldsymbol{q}) = \frac{9}{2}\operatorname{tr}\left(\bar{\zeta}_{ij}^3(\boldsymbol{q})\right) , \tag{6.10}$$

where $\bar{\zeta}_{ij} = \zeta_{ij} + (1/3)J_1\delta_{ij}$ are the components of the traceless part of the Hessian (ζ_{ij}). We will group all these rotational invariants into the vector of variables

$$\boldsymbol{w} \equiv \left\{\nu(\boldsymbol{q}), J_1(\boldsymbol{q}), 3\eta^2(\boldsymbol{q}), 5J_2(\boldsymbol{q}), J_3(\boldsymbol{q})\right\} , \tag{6.11}$$

and denote their PDF as $p(\boldsymbol{w})$. The reason for using $3\eta^2$ and $5J_2$ will become apparent shortly. The remaining $10-5=5$ angular degrees of freedom, which we group into the vector $\boldsymbol{\Omega}$, characterize the direction of the gradient $\boldsymbol{\nabla}\delta_R$ (two angles) and the orientation of the principal axis frame of ζ_{ij} (three "Euler" angles). The corresponding PDF is $p(\boldsymbol{\Omega})$. These angles will factorize out of the following calculations, so we can simply ignore them.

For Gaussian initial conditions, the probability density $p(\boldsymbol{w})$ simplifies to

$$p(\boldsymbol{w}) = \frac{5\sqrt{5}}{8\pi^2\sqrt{3}}\frac{\sqrt{\eta^2}}{\sqrt{1-\gamma_1^2}}\exp\left[-\frac{\nu^2 + J_1^2 - 2\gamma_1\nu J_1}{2(1-\gamma_1^2)} - \frac{3}{2}\eta^2 - \frac{5}{2}J_2\right] . \tag{6.12}$$

Importantly, we always have $J_2 > 0$ and $J_3^2 \leq (J_2)^3$ since ζ_{ij} is symmetric. Therefore, $p(\boldsymbol{w})$ is normalized such that

$$1 = \int_{-\infty}^{+\infty}d\nu \int_{-\infty}^{+\infty}dJ_1 \int_0^\infty d(3\eta^2) \int_0^\infty d(5J_2) \int_{-J_2^{3/2}}^{+J_2^{3/2}}dJ_3\, p(\nu, J_1, 3\eta^2, 5J_2, J_3) , \tag{6.13}$$

even though it does not explicitly depend on J_3. Clearly, the integral over J_3 is trivial and results in $5J_2$ being χ^2-distributed with 5 degrees of freedom. These 5 degrees of freedom correspond to the 5 independent components of $\bar{\zeta}_{ij}$. This is the reason for writing $5J_2$. Similarly, one can easily show that $3\eta^2$ is χ^2-distributed with 3 degrees of freedom. Note also that $\int d\boldsymbol{\Omega}\, p(\boldsymbol{\Omega}) = 1$.

The variable J_3 does not explicitly appear in Eq. (6.12) because it is, in fact, uniformly distributed. In order to emphasize the point, we introduce $x_3 \equiv J_3/(J_2)^{3/2}$. Then, $p(\boldsymbol{w})d\boldsymbol{w}$ can be written as

$$p(\boldsymbol{w})d\boldsymbol{w} = \mathcal{N}(\nu, J_1)d\nu dJ_1 \times \chi_3^2(3\eta^2)d(3\eta^2) \times \chi_5^2(5J_2)d(5J_2) \times \frac{1}{2}\Theta_H(1-x_3^2)dx_3 . \tag{6.14}$$

In the above expression, $\mathcal{N}(x, y)$ is a bivariate Normal distribution with unit variances,

$$\mathcal{N}(\nu, J_1) = \frac{1}{2\pi\sqrt{1-\gamma_1^2}}\exp\left[-\frac{\nu^2 + J_1^2 - 2\gamma_1\nu J_1}{2(1-\gamma_1^2)}\right] , \tag{6.15}$$

whereas $\chi_k^2(x)$ is the χ^2 distribution with k degrees of freedom,

$$\chi_k^2(x) = \frac{1}{2^{k/2}\Gamma(k/2)}x^{k/2-1}e^{-x/2} . \tag{6.16}$$

Clearly, x_3 has uniform probability density in the range $-1 \leq x_3 \leq +1$; thus, we can think of it as the cosine of an angle. Note, however, that the "angle" x_3 has nothing to do with spatial rotations. In fact, as the cubic root of the characteristic polynomial $\det(\zeta - \lambda I)$, where I is the 3×3 identity matrix, x_3 itself is a rotational invariant. Below, we will see that, unlike $\boldsymbol{\Omega}$ which does not contribute to the clustering of peaks, $J_3(\boldsymbol{q})$ actually does contribute, and, therefore, x_3 does not describe a rotation in space.

6.4 Average peak number density*

Owing to the peak constraint, the calculation of the N-point correlation functions of peaks of a 3-dimensional random field requires the evaluation of high-dimensional integrals over a joint probability distribution in $5N$ variables, after the $5N$ angular variables have been integrated out. Therefore, even the evaluation of the 1-point correlation function or average density of peaks $\overline{n}_{\rm pk}$ is not completely trivial.

The average number density of peaks of a 3-dimensional Gaussian random field,

$$\overline{n}_{\rm pk}(\nu_c) \equiv \langle n_{\rm pk}(\boldsymbol{q}) \rangle = \int d^5\boldsymbol{w}\, n_{\rm pk}(\boldsymbol{w})\, p(\boldsymbol{w}) \,, \tag{6.17}$$

was first calculated in BBKS [13]. Here however, we shall not proceed along the lines of [13], who explicitly wrote the volume element $d^6\zeta_{ij}$ in terms of the three eigenvalues λ_i of $-\zeta_{ij}$. Rather, we will follow the calculation of [196], which exploit the invariance under rotations and, hence, is far simpler.

Namely, the peak constraint implies that all three eigenvalues of the Hessian ζ_{ij} be negative. In terms of the rotational invariants, the restriction to local maxima of the density field translates into the conditions $J_1 > 0$, $J_2 < J_1^2$ and $J_3 < (J_1/2)(J_1^2 - 3J_2)$. Taking into account the symmetry of ζ_{ij}, the last condition implies that x_3 must satisfy

$$-1 < x_3 < \min\bigl[1, (y/2)(y^2 - 3)\bigr] \,, \tag{6.18}$$

where $y \equiv J_1/\sqrt{J_2}$. This splits the parameter space into two different regions depending on whether the inequality $(y/2)(y^2 - 3) < 1$ holds. For $0 < J_2 < J_1^2/4$, one finds $-1 < x_3 < +1$ whereas, for $J_1^2/4 < J_2 < J_1^2$, the more stringent constraint $-1 < x_3 < (y/2)(y^2 - 3)$ applies. Therefore, the multiplicative factor of $\Theta_H(\lambda_3)$ in the localized peak number density Eq. (6.9) can also be written as

$$\Theta_H(\lambda_3) = \Theta_H(J_1)\biggl\{ \Theta_H\bigl(J_1^2/4 - J_2\bigr) + \Theta_H\bigl(J_2 - J_1^2/4\bigr)\, \Theta_H\bigl(J_1^2 - J_2\bigr)$$
$$\times\, \Theta_H\bigl(y^3/2 - 3y/2 - x_3\bigr)\biggr\} \,. \tag{6.19}$$

We are now in the position to derive the well-known BBKS formula for the average peak number density $\overline{n}_{\rm pk}$ in a very simple way. To compute $\overline{n}_{\rm pk}$, one usually expresses the measure $d(5J_2)$ in terms of the ellipticity v, the prolateness w and three Euler angles so that, in these new variables, $J_2 = 3v^2 + w^2$. The calculation would then proceed along the lines of BBKS. However, this change of variable is, in fact, unnecessary as the calculation can be explicitly carried out in the variables J_1, $5J_2$ and x_3 on imposing the aforementioned conditions. To illustrate this point, we begin by rewriting the determinant $|\det(\zeta_{ij})|$ in Eq. (6.9) in terms of the J_i. Introducing $s \equiv 5J_2$, we obtain

$$|\det(\zeta_{ij})| = \frac{1}{27}\left(J_1^3 - \frac{3}{5}sJ_1 - \frac{2}{5^{3/2}}s^{3/2}x_3 \right) \,, \tag{6.20}$$

since $\det(\zeta_{ij})$ is always negative for density maxima. The integral over the variables $5J_2$ and x_3 becomes

$$\int ds\, \chi_5^2(s) \int dx_3\, \frac{1}{2}\Theta_H(1 - x_3^2)\, |\det(\zeta_{ij})|\, \Theta_H(\lambda_3) \tag{6.21}$$

$$= \frac{1}{2^{7/2}3^3\Gamma(5/2)}\left\{ \int_0^{5J_1^2/4} ds \int_{-1}^{+1} dx_3 + \int_{5J_1^2/4}^{5J_1^2} ds \int_{-1}^{\sqrt{5/s}(J_1/2)(5J_1^2/s-3)} dx_3 \right\}$$
$$\times \left(J_1^3 - \frac{3}{5}sJ_1 - \frac{2}{5^{3/2}}s^{3/2}x_3 \right) s^{3/2} e^{-s/2} \,,$$

---

* This section is of a more technical nature and is not essential for the remainder of this section. The main result of this section is Eq. (6.23).

and can be computed straightforwardly. Taking into account two additional multiplicative factors of $3^{3/2}$, one arising from $n_{\rm pk}(\boldsymbol{y})$ and the other from the integral over $\chi_3^2(3\eta^2)\delta_D(\boldsymbol{\eta})$, we find

$$3^3 \int ds\, \chi_5^2(s) \int dx_3\, \frac{1}{2}\Theta_H(1-x_3^2)\left|\det(\zeta_{ij})\right|\Theta_H(\lambda_3) \tag{6.22}$$

$$= \sqrt{\frac{2}{5\pi}}\left[\left(\frac{J_1^2}{2}-\frac{8}{5}\right)e^{-5J_1^2/2}+\left(\frac{31}{4}J_1^2+\frac{8}{5}\right)e^{-5J_1^2/8}\right]+\frac{1}{2}\left(J_1^3-3J_1\right)$$

$$\times\left[\mathrm{Erf}\left(\sqrt{\frac{5}{2}}J_1\right)+\mathrm{Erf}\left(\sqrt{\frac{5}{2}}\frac{J_1}{2}\right)\right]\equiv f(J_1)\,,$$

which is precisely the function $f(J_1)$ defined in [13]. The rest of the calculation is trivial, and we immediately recover their well-known expression for the average peak abundance $\overline{n}_{\rm pk}(\nu_c)$,

$$\overline{n}_{\rm pk}(\nu_c) = \frac{1}{(2\pi)^2 R_\star^3}G_0(\gamma_1,\gamma_1\nu_c)e^{-\nu_c^2/2}\,, \tag{6.23}$$

where the function G_0 is a special case of

$$G_n(\gamma_1,\omega) = \int_0^\infty du\, u^n f(u)\, \frac{e^{-(u-\omega)^2/2(1-\gamma_1^2)}}{\sqrt{2\pi(1-\gamma_1^2)}}\,, \tag{6.24}$$

and $f(u)$ is the function defined in Eq. (6.22). The integration over u must generally be performed numerically. Note that, within the peak formalism, $\overline{n}_{\rm pk}$ is then related to the mean halo abundance through $\overline{n}_{\rm pk}(\nu_c)d\nu_c = \overline{n}_h(\nu_c)d\ln M$.

It is worth noticing that, while the exponential $\exp[-(u-\omega)^2/2(1-\gamma_1^2)]$ decays rapidly to zero, $u^n f(u)$ are rapidly rising. As a result, the integrands are sharply peaked, and the functions $G_n(\gamma_1,w)$ receive most of their contribution around the maximum of the peak. For large values of ω, we find that G_0 and G_1 asymptote to

$$G_0(\gamma_1,\omega) \approx \omega^3 - 3\gamma_1^2\omega + B_0(\gamma_1)\,\omega^2\, e^{-A(\gamma_1)\omega^2} \tag{6.25}$$

$$G_1(\gamma_1,\omega) \approx \omega^4 + 3\omega^2\left(1-2\gamma_1^2\right) + B_1(\gamma_1)\,\omega^3\, e^{-A(\gamma_1)\omega^2}\,. \tag{6.26}$$

The coefficients $A(\gamma_1)$, $B_0(\gamma_1)$ and $B_1(\gamma_1)$ are obtained from the asymptotic expansion of the Error function that appears in Eq. (6.22). We have explicitly

$$A = \frac{5/2}{(9-5\gamma_1^2)}, \qquad B_0 = \frac{432}{\sqrt{10\pi}\,(9-5\gamma_1^2)^{5/2}}, \qquad B_1 = \frac{4B_0}{(9-5\gamma_1^2)}\,. \tag{6.27}$$

The value of ν_c at which $n_{\rm pk}(\nu_c)$ attains its maximum depends sensitively on the value of γ_1. For small $\gamma_1 \lesssim 0.3$, the peak is at $\nu_c \simeq 0$, reflecting the fact that there is significant power on all scales. For large $\gamma_1 \gtrsim 0.7$, most of the power is in a narrow range of wavenumber, so that density maxima are much more likely to reach heights well above typical, 1σ fluctuations of the density field.

6.5 Two-point correlation functions of peaks

While the average number density of peaks has been worked out analytically in [13], progress with the computation of the peak correlation functions, which is difficult to perform rigorously, has been relatively slow. Even though the peak correlation functions do not generally have explicit analytical expressions, they provide much information on ideas such as the peak-background split and its range of validity. Furthermore, in low dimensions they can be solved numerically and, therefore, furnish insights into non-perturbative effects such as small-scale halo exclusion. For illustrative purposes, we will focus on BBKS peaks defined through Eq. (6.9), but the following considerations can be straightforwardly extended to the excursion-set peaks discussed in Sec. 6.8.

6.5.1 The peak-density cross-correlation function

Let us begin with the cross-correlation function $\xi_{\mathrm{pk},\delta}(r)$ between peaks of height ν_c and the Lagrangian density field. This is the Fourier transform of $P_{hm}^L(k)$ considered in Sec. 4.1.2, evaluated for peaks. Further, it corresponds to the average Lagrangian density profile $\langle\delta(\boldsymbol{r})|\text{peak at }\boldsymbol{0}\rangle$ around peaks. Unlike the peak auto-correlation function, the cross-correlation function can be derived in closed form [13], and is given by

$$\xi_{\mathrm{pk},\delta}^L(r) = b_{10}^L \xi_0^{(0)}(R,r) + b_{01}^L \xi_0^{(1)}(R,r) = \left[b_{10}^L - b_{01}^L \nabla^2 \right] \xi_0^{(0)}(R,r) \tag{6.28}$$

in the notation adopted here. We have used the relation $\nabla^2 \xi_0^{(0)} = -\xi_0^{(1)}$ in the second equality. In the notation of [257], $b_{10}^L = b_\nu$ and $b_{01}^L = b_\zeta$, while in the notation of Sec. 2 and Sec. 4, $b_{10}^L = b_1^L$ and $-b_{01}^L$ contributes to the leading Lagrangian higher-derivative bias $b_{\nabla^2\delta}^L$, along with the contribution from the filtering kernel in $\xi_0^{(0)}(R,r)$ derived in Eq. (2.69) (Sec. 2.6). The linear BBKS peak bias parameters are given by

$$b_{10}^L = \frac{1}{\sigma_0}\left(\frac{\nu_c - \gamma_1 \bar{J}_1}{1 - \gamma_1^2} \right) , \qquad b_{01}^L = \frac{1}{\sigma_2}\left(\frac{\bar{J}_1 - \gamma_1 \nu_c}{1 - \gamma_1^2} \right) , \tag{6.29}$$

where $\bar{J}_1 \equiv G_1(\gamma_1, \gamma_1\nu_c)/G_0(\gamma_1, \gamma_1\nu_c)$ is the average curvature of peaks of height ν_c. We will see in Sec. 6.6.1 that these bias parameters can be obtained directly from the peak abundance Eq. (6.9) using a peak-background split argument. Note that b_{10}^L and b_{01}^L satisfy the relation

$$b_{10}^L + \left(\frac{\sigma_1^2}{\sigma_0^2} \right) b_{01}^L = \frac{\nu_c}{\sigma_0} , \tag{6.30}$$

analogous to Eq. (5.100); the additional factor of σ_1^2/σ_0^2 arises from the fact that, unlike in Eq. (5.100), b_{01}^L defined here has dimensions of length^2. Let us also mention that $\xi_0^{(0)}(r)$ in Ref. [257] corresponds to Eq. (7.10) in Ref. [13]. Although the latter equation appears to have an additional factor of $1/3$ that multiplies the factors of $\nabla^2\xi_0^{(0)}$, this is only because Ref. [13] measures r in units of R_1.

6.5.2 The peak auto-correlation function

Unlike the peak-matter cross-correlation function, the auto-correlation function of peaks is a significantly more involved calculation. However, it exhibits interesting features beyond those present in Eq. (6.28), in particular nonlinear bias and stochasticity. To begin, we define the fractional peak overdensity through

$$\delta_{\mathrm{pk}}^L(\boldsymbol{q}) = \frac{n_{\mathrm{pk}}(\boldsymbol{q})}{\bar{n}_{\mathrm{pk}}} - 1 , \tag{6.31}$$

where $\bar{n}_{\mathrm{pk}}$ is the average peak number density, Eq. (6.23). The peak auto-correlation function is then given by the expectation value of the product $\delta_{\mathrm{pk}}^L(\boldsymbol{q}_1)\delta_{\mathrm{pk}}^L(\boldsymbol{q}_2)$, which reads

$$\langle \delta_{\mathrm{pk}}^L(\boldsymbol{q}_1)\delta_{\mathrm{pk}}^L(\boldsymbol{q}_2) \rangle = \frac{1}{\bar{n}_{\mathrm{pk}}^2}\Big\langle \sum_\alpha \sum_\beta \delta_D(\boldsymbol{q}_1 - \boldsymbol{q}_\alpha)\,\delta_D(\boldsymbol{q}_2 - \boldsymbol{q}_\beta) \Big\rangle - 1 \tag{6.32}$$

$$= \frac{1}{\bar{n}_{\mathrm{pk}}^2}\delta_D(\boldsymbol{q}_1 - \boldsymbol{q}_2)\Big\langle \sum_\alpha \delta_D(\boldsymbol{q}_2 - \boldsymbol{q}_\alpha) \Big\rangle + \frac{1}{\bar{n}_{\mathrm{pk}}^2}\Big\langle \sum_{\alpha\neq\beta} \delta_D(\boldsymbol{q}_1 - \boldsymbol{q}_\alpha)\,\delta_D(\boldsymbol{q}_2 - \boldsymbol{q}_\beta) \Big\rangle - 1$$

$$= \frac{1}{\bar{n}_{\mathrm{pk}}}\delta_D(\boldsymbol{q}_1 - \boldsymbol{q}_2) + \frac{1}{\bar{n}_{\mathrm{pk}}^2}\sum_{\alpha\neq\beta}\Big\langle \delta_D(\boldsymbol{q}_1 - \boldsymbol{q}_\alpha)\,\delta_D(\boldsymbol{q}_2 - \boldsymbol{q}_\beta) \Big\rangle - 1$$

$$= \frac{1}{\bar{n}_{\mathrm{pk}}}\delta_D(\boldsymbol{q}_2 - \boldsymbol{q}_1) + \xi_{\mathrm{pk}}^L(r) , \tag{6.33}$$

with $r = |\boldsymbol{q}_2 - \boldsymbol{q}_1|$. The first term on the right-hand side comes from self-pairs, which are usually ignored in the calculation of the 2-point correlation function in real space, but are relevant in Fourier space; the second term is the usual correlation function $\xi_{\mathrm{pk}}^L(r)$ of peaks excluding self-pairs.

Before turning to the calculation of $\xi_{\rm pk}^L(r)$, let us consider the issue of stochasticity in a bit more detail. The contribution $\overline{n}_{\rm pk}^{-1}\delta_D(\boldsymbol{r})$ corresponds to Poisson shot noise, and yields an exactly constant contribution $\overline{n}_{\rm pk}^{-1}$ to the Fourier transform $P_{\rm pk}^L(k)$ of Eq. (6.33). The actual stochastic contribution to the peak power spectrum on large scales can be obtained by using the fact that the deterministic contribution scales as $P_{\rm L}(k) \propto k^{n_s - 1}$ in the large-scale limit, as it does for any physical biased tracer. The stochastic contribution can then be isolated as the constant that remains when sending k to exactly zero:

$$P_{\varepsilon,{\rm pk}}^{\{0\}} \equiv \lim_{k \to 0} P_{\rm pk}^L(k) = \frac{1}{\overline{n}_{\rm pk}} + \int d^3\boldsymbol{r}\, \xi_{\rm pk}^L(r)\,. \tag{6.34}$$

If $\int d^3\boldsymbol{r}\, \xi_{\rm pk}^L(r)$ does not vanish, then peaks show super-Poisson or sub-Poisson noise in the low-k limit of the power spectrum. In fact, simple arguments suggest that peak-peak exclusion should contribute a negative white-noise term, so that $P_{\varepsilon,{\rm pk}}^{\{0\}} < 1/\overline{n}_{\rm pk}$ (see [127, 177, 178] for a discussion of halo exclusion). Thus, peaks provide a nontrivial example of stochasticity beyond the usual Poissonian counts-in-cells prediction [71, 486].

We now turn to the proper 2-point correlation function $\xi_{\rm pk}^L(r)$ of peaks of height ν_c. Let p_2 be the joint probability for the vector of independent components $\boldsymbol{y} \equiv \{\nu, \eta_i, \zeta_{ij}\}$ at positions $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$. Note that, since $\boldsymbol{r}$ provides a preferred direction, it is no longer sufficient to write the peak abundance solely in terms of the five invariants $\boldsymbol{w}$. Then, $\xi_{\rm pk}^L(r)$ formally reads

$$1 + \xi_{\rm pk}^L(r) = \frac{1}{[\overline{n}_{\rm pk}(\nu_c)]^2} \int d^{10}\boldsymbol{y}_1 \int d^{10}\boldsymbol{y}_2\, n_{\rm pk}(\boldsymbol{y}_1) n_{\rm pk}(\boldsymbol{y}_2)\, p_2(\boldsymbol{y}_1, \boldsymbol{y}_2) \tag{6.35}$$

where, for shorthand convenience, subscripts denote quantities evaluated at different Lagrangian positions $\boldsymbol{q}_1, \boldsymbol{q}_2$, where $r = |\boldsymbol{q}_2 - \boldsymbol{q}_1|$, and $n_{\rm pk}(\boldsymbol{y})$ is given by Eq. (6.9). It is now clear why the evaluation of Eq. (6.35) is tedious: one must evaluate the joint probability distribution $p_2(\boldsymbol{y}_1, \boldsymbol{y}_2)$ for the 2×10-dimensional vector of variables, which involves a 20-dimensional covariance matrix.

Several numerical and analytical investigations have been performed to evaluate Eq. (6.35). Early work [459, 487, 488] focused on Monte-Carlo realizations of 1-dimensional Gaussian (and non-Gaussian) random fields to estimate the peak 2-point correlation function down to small separations where peak-peak exclusion becomes important. Ref. [13] worked out analytically the peak 2-point correlation function in the large-scale limit, neglecting spatial derivatives of the matter correlation function. This essentially corresponds to performing a LIMD bias expansion as in Sec. 2.1. Refs. [489, 490, 491] obtained asymptotic expansions appropriate for high peak threshold $\nu_c \gg 1$. Leading-order derivations are presented in [481, 156, 257], whereas [152, 221] obtained the full NLO (1-loop) expression.

In the following section, we will describe in detail a rigorous perturbative approach to deriving $\xi_{\rm pk}^L(r)$ on large scales where the matter correlation function $\xi_{{\rm L},R}(r)$ is small. Before we discuss this approach, let us make a few general comments:

- The peak auto-correlation function generally reaches the value $\xi_{\rm pk}^L = -1$ in the limit $r \to 0$, because the probability of finding two maxima arbitrarily close together is suppressed for the smoothed density field. This small-scale peak exclusion is clearly seen in the one-dimensional calculation of [488, 347]. As discussed above, this is related to the fact that $\int d^3\boldsymbol{r}\, \xi_{\rm pk}^L(r)$ does not generally vanish for density peaks. As a consequence, $P_{\varepsilon,{\rm pk}}^{\{0\}} \neq 1/\overline{n}_{\rm pk}$ so that the noise is not Poissonian. The one-dimensional calculation of [178] has shown that the small-scale peak exclusion can qualitatively reproduce the shot-noise matrix measurements of [177].

- As pointed out in [156], the higher-derivative peak bias b_{01}^L can significantly amplify the contrast of the BAO feature in the clustering of initial density peaks if $b_{01}^L > 0$, as illustrated in Fig. 28. This arises from the fact that $\xi_0^{(1)}(R, r)$ is equal to $-\nabla^2 \xi_0^{(0)}(R, r)$, and the BAO feature is enhanced through the derivative of the correlation function (see also the discussion in Sec. 2.6). The analysis of [152] further includes the gravitational evolution to ascertain the extent to which the initial BAO sharpening survives at late time. A crucial component of this analysis is velocity bias. We will return to these issues in Sec. 6.9.
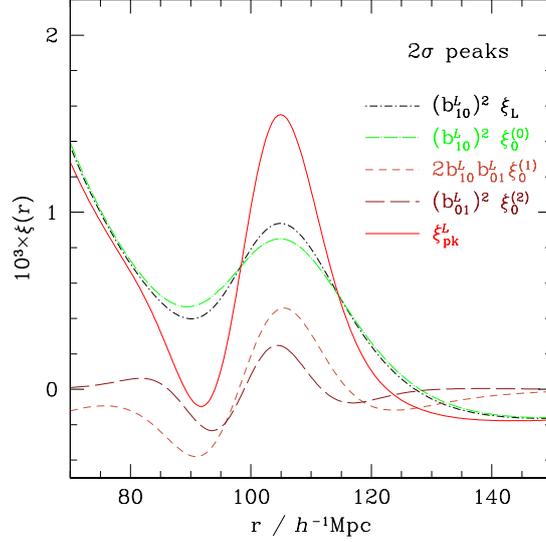
Figure 28: A comparison between the *un-smoothed* Lagrangian matter correlation function $(b_{10}^L)^2 \xi_{\rm L}(r)$ (black, dot-dashed) and the Lagrangian correlation function $\xi_{\rm pk}^L(r)$ of $\nu_c = 2$ peaks (red, solid) around the BAO feature. In order to obtain the peak correlation function, the density field was smoothed with a Gaussian filter on mass scale $M = 3 \times 10^{13}\, h^{-1}\, M_\odot$. The dot-long dashed, short-dashed and long-dashed curves represent the individual contributions $(b_{10}^L)^2 \xi_0^{(0)} = (b_1^L)^2 \xi_{{\rm L},R}$, $2b_{10}^L b_{01}^L \xi_0^{(1)} = -2b_1^L b_{\nabla^2 \delta}^L \nabla^2 \xi_{{\rm L},R}$ and $(b_{01}^L)^2 \xi_0^{(2)} = (b_{\nabla^2 \delta}^L)^2 \nabla^4 \xi_{{\rm L},R}$ to the linear peak correlation function. A nonzero b_{01}^L restores, and even amplifies the acoustic peak otherwise smeared out by the filtering of the matter density field. Note that the second- and higher-order contributions to $\xi_{\rm pk}^L(r)$ are so small on these scales that they do not lead to visible differences. From [152].

6.6 Perturbative peak bias expansion

While [152] showed that the first- and some of the second-order contributions to $\xi_{\rm pk}^L(r)$ (those depending on $\nu \propto \delta_R$ and $J_1 \propto -\nabla^2 \delta_R$) could be obtained from a peak-background split formulated in terms of conditional mass functions, validating thereby the peak-background split calculation of [110] of the LIMD biases b_{i0}^L, they could not determine the physical origin of the other second-order contributions. This was clarified in [221, 196].

Let us go back to the definition of $\xi_{\rm pk}^L(r)$ through Eq. (6.33). Since it describes the 2-point correlation of a scalar variable $\delta_{\rm pk}(\boldsymbol{q})$, which, moreover, only depends on the variables ν, η_i and ζ_{ij}, $\xi_{\rm pk}^L(r)$ can be expanded in two-point correlations of all independent scalars constructed from ν, η_i, ζ_{ij}; that is, *exactly the five quantities in the vector $\boldsymbol{w}$* [Eq. (6.11)]. Note that we have also relied on the absence of coupling between η_i and ζ_{ij} in the BBKS peak prescription. That is, the peak correlation function can be obtained perturbatively by writing the peak density perturbation $\delta_{\rm pk}^L(\boldsymbol{q})$ as

$$
\begin{aligned}
\delta_{\rm pk}^L(\boldsymbol{q}) &= F\left[\nu(\boldsymbol{q}),\, J_1(\boldsymbol{q}),\, 3\eta^2(\boldsymbol{q}),\, 5J_2(\boldsymbol{q}),\, J_3(\boldsymbol{q})\right] \\
&= c_\nu \nu(\boldsymbol{q}) + c_{\nu^2} \nu^2(\boldsymbol{q}) + c_{J_1} J_1(\boldsymbol{q}) + c_{\nu J_1} \nu(\boldsymbol{q}) J_1(\boldsymbol{q}) + \cdots,
\end{aligned}
\tag{6.36}
$$

where in the second line we have written a few example terms up to second order, multiplied by bias coefficients. This corresponds to a specific Lagrangian bias expansion of the type discussed in Sec. 2.3 and 2.4, see Tab. 11 (p. 145). Specifically, all terms involving the tidal field K_{ij} are absent, and the higher-derivative terms only consist of powers of $(\boldsymbol{\nabla}\delta)^2$ and all scalar combinations of $\partial_i \partial_j \delta$, in addition to the terms obtained by expanding the filtering kernel $W_R(k)$ (Sec. 2.6); as discussed there, this is only a subset of all higher-derivative operators that could be present. The physical reason is, again, the simplification made by assuming the density peak constraint and spherical collapse approximation.

Now we are faced with two issues: first, in order to obtain a prediction for the peak correlation function, we need to calculate the values of the bias parameters; second, when calculating the two-point function

144

Peak model (Sec. 6)	General expansion (Sec. 2–4)
b_{ij}^L	$(-1)^j i! j! \, b_{\delta^i (\nabla^2 \delta)^j}^L$ [1]
χ_i^L	$b_{(\boldsymbol{\nabla}\delta)^{2i}}^L$
ω_{lm}^L	higher-derivative bias involving J_2, J_3 constructed from $(\partial_i \partial_j \delta)$

[1] For $j > 0$, there are further contributions to $b_{\delta^i (\nabla^2 \delta)^j}^L$ from the expansion of the filtering kernel (Sec. 2.6).

Table 11: Conversion between peak biases (left) and those appearing in the general bias expansion (right).

using Eq. (6.36) directly, we obtain a large number of zero-lag terms which spoil the perturbative expansion on large scales as described in Sec. 2.10.1. The second problem can be solved by replacing the fields in the second line of Eq. (6.36) with renormalized operators, multiplied by renormalized bias parameters:

$$\delta_{\mathrm{pk}}^L(\boldsymbol{q}) = \sigma_0 b_{10}^L[\nu](\boldsymbol{q}) + \frac{1}{2}\sigma_0^2 b_{20}^L[\nu^2](\boldsymbol{q}) + \sigma_2 b_{01}^L[J_1](\boldsymbol{q}) + \sigma_0 \sigma_2 b_{11}^L[\nu J_1](\boldsymbol{q}) + \cdots . \tag{6.37}$$

Powers of σ_i have been introduced to ensure that the renormalized bias parameters are defined relative to the physical (unnormalized) fields. As shown in [221, 196], the renormalized operators as well as their coefficients can be derived at all orders, given the Gaussian nature of the density field, by making use of orthogonal polynomials $O_{\boldsymbol{n}}^\star\big(\boldsymbol{w}(\boldsymbol{q})\big)$:

$$\delta_{\mathrm{pk}}^L(\boldsymbol{q}) = \sum_{\boldsymbol{n} \neq \{0\}} \sigma^{[\boldsymbol{n}]} b_{\boldsymbol{n}}^L O_{\boldsymbol{n}}^\star\big(\boldsymbol{w}(\boldsymbol{q})\big) . \tag{6.38}$$

Here, $\boldsymbol{n} = \{i_1, i_2, i_3, i_4, i_5\}$ is a list of indices that denote the highest power of each of the 5 variables $\boldsymbol{w} = \{\nu, J_1, \eta^2, J_2, J_3\}$ in the operator $O_{\boldsymbol{n}}^\star$, while $\sigma^{[\boldsymbol{n}]}$ is a shorthand for the product $\sigma_0^{i_1} \sigma_1^{2i_3} \sigma_2^{i_2 + 2i_4 + 3i_5}$. Thus, $\{O_{\boldsymbol{n}}^\star\}_{\boldsymbol{n}}$ corresponds to the entire set of renormalized operators that can be constructed from the 5 invariants contained in $\boldsymbol{w}$. Correspondingly, the coefficients $b_{\boldsymbol{n}}^L$ are physical, renormalized bias parameters which are measurable for example through large-scale correlations, and are given by the *1-point ensemble averages*

$$\sigma^{[\boldsymbol{n}]} b_{\boldsymbol{n}}^L \equiv \frac{1}{\bar{n}_{\mathrm{pk}}} \Big\langle n_{\mathrm{pk}}(\boldsymbol{q}) \, O_{\boldsymbol{n}}[\boldsymbol{w}(\boldsymbol{q})] \Big\rangle . \tag{6.39}$$

Here, $O_{\boldsymbol{n}}^\star$ and $O_{\boldsymbol{n}}$ are dual polynomials satisfying (see Appendix A.3)

$$\langle O_{\boldsymbol{n}}^\star(\boldsymbol{q}) O_{\boldsymbol{n}'}(\boldsymbol{q}) \rangle = \delta_{\boldsymbol{n}\boldsymbol{n}'} , \tag{6.40}$$

and $\delta_{\boldsymbol{n}\boldsymbol{n}'} = \prod_{k=1}^{5} \delta_{i_k i_k'}$. Thus, by making use of a perturbative bias expansion, we have reduced the very complicated problem of calculating the peak two-point function Eq. (6.35) to the *much* simpler problem of calculating the auto- and cross-correlations of the scalars ν, η^2, J_i, and the bias parameters Eq. (6.39) which are one-point expectation values. Moreover, as we will see in Sec. 6.6.2, we can straightforwardly derive higher-order peak correlation functions in this approach as well, while their calculation based on the joint probabilities p_3, p_4, ... would be prohibitive. Note, however, that it is not possible to obtain the non-perturbative stochastic contributions in this way.

In the following subsection, we will derive the specific form of the polynomials, and relate their associated bias parameters to the peak-background split. Note that the polynomials are defined up to multiplicative constants which can be re-absorbed in the bias parameters $b_{\boldsymbol{n}}^L$. In what follows, like in Sec. 2, we will adopt a normalization such that the peak LIMD bias parameters b_{i0}^L correspond to the parameters b_i^L used throughout (see Tab. 11 for the correspondence between the peak bias parameters used here and those used in Sec. 2–4).

6.6.1 Polynomials, bias parameters and the peak-background split

The relevant polynomials are those associated with the distribution of the invariants $\boldsymbol{w}$. Namely, for the variables ν and J_1 distributed according to the bivariate normal distribution $\mathcal{N}(\nu, J_1)$ [Eq. (6.15)], the orthogonal polynomials are the bivariate Hermite $H_{ij}(\nu, J_1)$. For the χ^2 variate $3\eta^2$ with 3 degrees of freedom, these are the generalized Laguerre polynomials $L_q^{(1/2)}(3\eta^2/2)$, which are orthogonal over $[0, \infty)$ with respect to the χ^2-distribution $\chi_3^2(3\eta^2)$ with $k = 3$ degrees of freedom. Finally, for the jointly distributed variables (J_2, J_3), the polynomials are [196]

$$F_{lm}(5J_2, J_3) = (-1)^l \sqrt{\frac{\Gamma(5/2)}{2^{3m}\Gamma(3m + 5/2)}} \, L_l^{(3m+3/2)}(5J_2/2) \, \mathcal{L}_m(x_3) \,, \tag{6.41}$$

where $\mathcal{L}_m(x)$ are Legendre polynomials, and the factor of $(-1)^l$ ensures that the term with highest power of J_2 always has positive sign. The Legendre polynomials $\mathcal{L}_m(x_3)$ appear because they are the orthonormal polynomials associated with the uniform distribution on the interval $[-1, 1]$, i.e. $\Theta_H(1 - x_3^2)/2$ in Eq. (6.14). Note that both L_q and F_{lm} are self-dual, i.e. $L_q^{(1/2)\star} = L_q^{(1/2)}$ and $F_{lm}^\star = F_{lm}$ owing to the factorization of the PDF of $(5J_2, x_3)$. Finally, we slightly modified the notation of [492] so that their F_{l0} agree with the above definition. The physical origin of the appearance of these orthogonal polynomials can be found in the peak-background split: *Long-wavelength background perturbations locally modulate the mean of the distributions* $\mathcal{N}(\nu, u)$, $\chi_3^2(3\eta^2)$ *and* $\chi_5^2(5J_2)\Theta_H(1 - x_3^2)$. *The resulting non-central distributions can then be expanded in the appropriate set of orthogonal polynomials.*

The bias coefficients of the BBKS peaks are given by the ensemble average Eq. (6.39). For the variables (ν, J_1), these are

$$\sigma_0^i \sigma_2^j b_{ij}^L = \frac{1}{\bar{n}_{\rm pk}} \Big\langle n_{\rm pk} \, H_{ij}(\nu, J_1) \Big\rangle \tag{6.42}$$
$$= \frac{1}{\bar{n}_{\rm pk}} \int d^5\boldsymbol{w} \, n_{\rm pk}(\boldsymbol{w}) \, H_{ij}(\nu, J_1) \, p(\boldsymbol{w}) \,,$$

where the peak constraint in $n_{\rm pk}(\boldsymbol{w})$ selects spatial locations corresponding to density peaks. Factors of $1/\sigma_0$ and $1/\sigma_2$ are introduced because bias factors are ordinarily defined relative to the physical field $\delta_R^{(1)}(\boldsymbol{q})$ and its derivatives, e.g. $\nabla^2 \delta_R^{(1)}(\boldsymbol{q})$, rather than the normalized variables ν, J_1. In the particular cases $ij = (10)$ and (01), we recover Eq. (6.29). Explicit expressions for the density bias parameters b_{i0}^L can be found in [110], whereas b_{ij}^L with $j \geq 1$ can be found in [152, 221].

Since the bivariate Hermite polynomials can be generated through the shift $\mathcal{N}(\nu, J_1) \to \mathcal{N}(\nu + \nu_\ell, u \to J_1 + J_{1\ell})$, where ν_ℓ and $J_{1\ell}$ are long-wavelength background perturbations uncorrelated with the (small-scale) fields $\nu(\boldsymbol{q})$ and $J_1(\boldsymbol{q})$, b_{ij}^L are peak-background split biases as in [12]. Thus, we find again that the renormalized, physical bias parameters are exactly those obtained from the peak-background split as discussed for the LIMD biases b_{i0}^L in Sec. 3. Note also that, owing to the fact that ν and J_1 are correlated variables, the b_{ij}^L can also be derived by considering only the long-wavelength perturbation ν_ℓ as done in [152]. Furthermore, since a shift in the mean curvature is equivalent to the replacement $\omega \to \omega - J_{1\ell}$ in Eq. (6.24), the b_{ij}^L can also be obtained as derivatives of the mean peak abundance with respect to ω as explained in [100]. This is in precise analogy to the peak-background split definition of the LIMD bias parameters b_i^L as responses of the mean abundance (Sec. 3). Finally, the same reasoning holds if one exchanges J_1 for the slope $\delta_R' = d\delta_R/dR$ considered in Sec. 5.7.1, because δ' and J_1 are completely correlated for a Gaussian smoothing kernel. Hence, the b_{ij}^L which arise from (δ, δ') can also be written down as ensemble average of bivariate Hermite polynomials [305, 430].

Similarly, the bias parameters associated with the invariant $3\eta^2$ are derived from the generalized Laguerre

polynomials, i.e.

$$\sigma_1^{2q} \chi_q^L = \frac{1}{\overline{n}_{\rm pk}} \Big\langle n_{\rm pk}\, L_q^{(1/2)}(3\eta^2) \Big\rangle \tag{6.43}$$

$$= \frac{(-1)^q}{\overline{n}_{\rm pk}} \int d^5\boldsymbol{w}\, n_{\rm pk}(\boldsymbol{w})\, L_q^{(1/2)}(3\eta^2/2)\, p(\boldsymbol{w}) \;.$$

The factor of $(-1)^q$ ensures that the term with the highest power of J_2 always has a positive sign. The first generalized Laguerre polynomials are $L_0^{(\alpha)}(x) = 1$ and $L_1^{(\alpha)}(x) = -x + \alpha + 1$. Taking into account the peak constraint, the lowest-order bias parameter thus is [221]

$$\chi_1^L = \frac{1}{\sigma_1^2 \overline{n}_{\rm pk}} \int d^5\boldsymbol{w}\, n_{\rm pk}(\boldsymbol{w}) \left(\frac{3}{2}\eta^2 - \frac{3}{2}\right) p(\boldsymbol{w}) \tag{6.44}$$

$$= -\frac{3}{2\sigma_1^2} \;.$$

Note that $3\eta^2/2 - 3/2 \geq -3/2$. Therefore, since $\int d\boldsymbol{w}\, p(\boldsymbol{w}) = \overline{n}_{\rm pk}$, we always have $\chi_1^L \geq (-3/2)\sigma_1^2$ regardless of the peak properties (height, curvature, etc). The effect of a background (long-wavelength) perturbation $\boldsymbol{\eta}_\ell$ on the components of the density gradient is $\eta_i \to \eta_i + \eta_{i,\ell}$. The components $\eta_{i,\ell}$ need not be the same for distinct i. However, owing to invariance under rotations, only the magnitude of the vector $\sum_{i=1}^3 (\eta_{i,\ell})^2 \equiv \eta_\ell^2$ matters. As a result, the background fluctuation effectively perturbs $\chi_3^2(3\eta^2)$ into a non-central χ^2-distribution, with non-centrality parameter $\lambda = \sqrt{\eta_\ell^2}$. Equivalently, the q-th order bias parameters χ_q^L can also be written as the derivatives $\sim \langle \partial^q n_{\rm pk}/\partial(\eta^2)^q \rangle$, again paralleling the case of LIMD biases described in Sec. 3.1.

Finally, the bias parameters ω_{lm}^L, which correspond to the variables J_2 and J_3, are given by

$$\sigma_2^{2l+3m} \omega_{lm}^L = \frac{1}{\overline{n}_{\rm pk}} \Big\langle n_{\rm pk}\, F_{lm}(5J_2, J_3) \Big\rangle \tag{6.45}$$

$$= \frac{1}{\overline{n}_{\rm pk}} \int d^5\boldsymbol{w}\, n_{\rm pk}(\boldsymbol{w})\, F_{lm}(5J_2, J_3) p(\boldsymbol{w}) \;.$$

Here again, the bias coefficients ω_{lm}^L can be derived through a peak-background split upon considering long-wavelength perturbations J_{2l} and J_{3l} to the second- and third-order invariant traces of $\bar{\zeta}_{ij}$. The computation of ω_{10}^L is straightforward, and one obtains [221]

$$\omega_{10}^L = -\frac{5}{2\sigma_2^2} \left(1 + \frac{2}{5}\partial_\alpha \ln G_0^{(\alpha)}(\gamma_1, \gamma_1\nu)\Big|_{\alpha=1}\right) \;, \tag{6.46}$$

where G_0^α is defined in Eq. (6.24). Note that $J_2(\boldsymbol{q})$ quantifies the ellipticity of the peak density profile. In the high-peak limit, $\omega_{10}^L \to 0$ reflecting the fact that the most prominent peaks are nearly spherical.

The previous expressions correspond to restricted expansions in each of the five invariants. The fully general bias parameters $b_{\boldsymbol{n}}^L$ of BBKS peaks take the generic form [196]

$$\sigma_0^i \sigma_1^{2q} \sigma_2^{j+2l+3m} b_{ijqlm}^L = \frac{1}{\overline{n}_{\rm pk}} \Big\langle n_{\rm pk}\, O_{ijqlm}(\boldsymbol{w}) \Big\rangle$$

$$\equiv \frac{1}{\overline{n}_{\rm pk}} \Big\langle n_{\rm pk}\, H_{ij}(\nu, J_1)(-1)^q L_q^{(1/2)}(3\eta^2/2) F_{lm}(5J_2, J_3) \Big\rangle \;. \tag{6.47}$$

This defines the orthogonal polynomials $O_{ijqlm}(\boldsymbol{w})$. Here we have again taken out factors of σ_i, so that the BBKS peak bias factors b_{ijqlm}^L are defined relative to the unnormalized fields $\delta_R(\boldsymbol{q})$, $\boldsymbol{\nabla}\delta_R$ and so on, as is conventional in large-scale structure. Note that, in general, the $b_{\boldsymbol{n}}^L$ do not factorize into a product of b_{ij}^L, χ_q^L and ω_{lm}^L, which correspond to the subsets $b_{ij}^L = b_{ij000}^L$, $\chi_q^L = b_{00q00}^L$, and $\omega_{lm}^L = b_{000lm}^L$ because of the cross-correlation between the variables.

Fig. 29 illustrates the behavior of the BBKS peak bias parameters b_{ij}^L, χ_q^L and ω_{lm}^L up to quadratic order. The bias parameter b_{01}^L (which contributes to $-b_{\nabla^2\delta}^L$) is always positive whereas $\sigma_1^2\chi_1^L$ is negative and equal to $-3/2$ at all peak heights [see Eq. (6.44)].
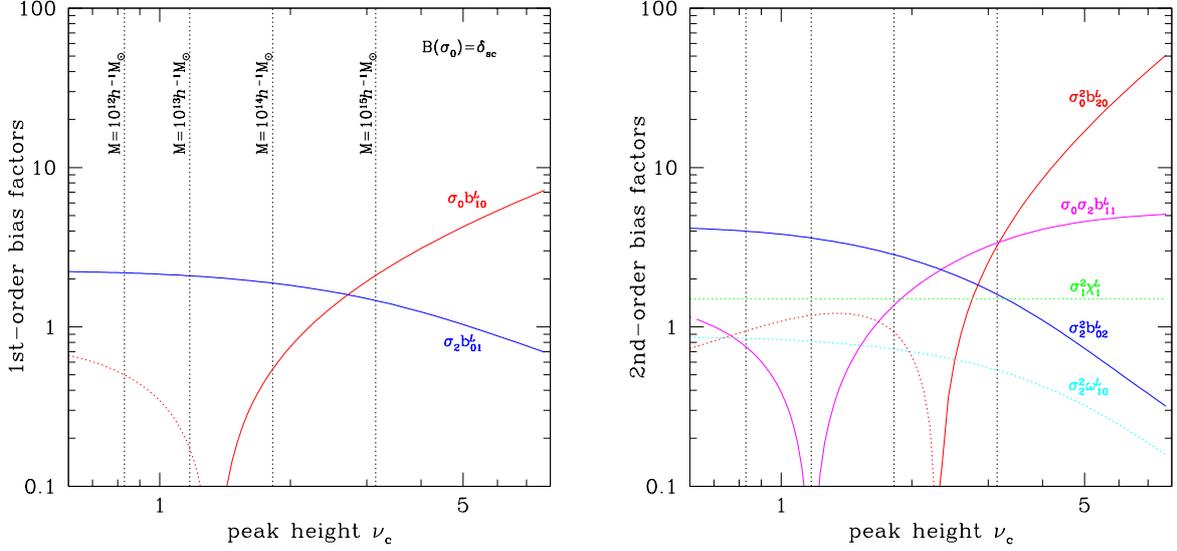
Figure 29: First-order *(left panel)* and second-order *(right panel)* Lagrangian peak bias parameters assuming a constant barrier $B(\sigma_0) = \delta_{\rm cr}$. The bias parameters have been multiplied by the appropriate factors of σ_i, so that they all are dimensionless. The second-order bias factors $\sigma_0^2 b_{20}^L$, $\sigma_0 \sigma_2 b_{11}^L$ and $\sigma_2^2 b_{02}^L$ are associated with the normalized variables $\nu(\boldsymbol{q})$ and $J_1(\boldsymbol{q})$, whereas $\sigma_1^2 \chi_1^L$ and $\sigma_2^2 \omega_{10}^L$ multiply the contributions from $\eta^2(\boldsymbol{q})$ and $J_2(\boldsymbol{q})$ (see text), respectively. Vertical lines mark the peak significance at which the halo mass is $M = 10^{12}$, 10^{13}, 10^{14} and 10^{15} $h^{-1}\,M_\odot$ (from left to right). Dashed curves indicate negative values. Results are shown at $z = 0$ for the ΛCDM cosmology defined in Sec. 1.3. From [221].

6.6.2 *Renormalization and peak correlation functions*

Having obtained the renormalized bias parameters $b_{\boldsymbol{n}}^L$ in the previous section, we can now use Eq. (6.38) to derive the n-point functions of peaks on large scales, by perturbatively expanding in the correlation functions of ν, $J_1, \cdots$. For this, we need the orthogonal polynomials $O_{ijqlm}^\star$ to those that appeared in Sec. 6.6.1. Their properties and construction are described in Appendix A.3. Let us give one example,

$$\left[\nu^2 J_1\right] \equiv H_{21}^\star(\nu, J_1) = \nu^2 J_1 - J_1 - 2\gamma_1 \nu \ . \tag{6.48}$$

As in the case of the standard univariate Hermite polynomials, lower-order terms appear in addition to the leading-order term $\nu^2 J_1$ which exactly cancel the zero-lag terms or self-correlators that appear in correlators of $\nu^2 J_1$ with other operators. Consider for example its auto-correlation function,

$$\left\langle \left[\nu^2(\boldsymbol{q}_1) J_1(\boldsymbol{q}_1)\right] \left[\nu^2(\boldsymbol{q}_2) J_1(\boldsymbol{q}_2)\right] \right\rangle = 2\langle \nu(\boldsymbol{q}_1)\nu(\boldsymbol{q}_2)\rangle^2 \, \langle J_1(\boldsymbol{q}_1) J_1(\boldsymbol{q}_2)\rangle \tag{6.49}$$
$$+ \, 4\langle \nu(\boldsymbol{q}_1)\nu(\boldsymbol{q}_2)\rangle \, \langle \nu(\boldsymbol{q}_1) J_1(\boldsymbol{q}_2)\rangle \, \langle \nu(\boldsymbol{q}_2) J_1(\boldsymbol{q}_1)\rangle \ .$$

We see that the terms of order < 3 in $H_{21}^\star(\nu, J_1)$ exactly cancel all the zero-lag contributions. Only the products of 2-point correlators at finite separation (which are the only nontrivial irreducible correlation functions for Gaussian random fields, see Appendix A) survive. Thus, our expansion in orthogonal polynomials is, in fact, equivalent to the renormalization procedure described in Sec. 2.10.1 (see also [493]). We have correspondingly denoted the operator $H_{21}^\star(\nu, J_1)$ with $[\nu^2 J_1]$. Note that this property of course relies on the Gaussianity of the linear density field.

Similar considerations apply to the other variables considered here. For a single density variable, the polynomials are univariate Hermite polynomials as was already pointed out by [99]. Note that a diagrammatic reasoning along the lines of [141] also explains why the self-correlators disappear. In fact, the Integrated Perturbation Theory (iPT) diagrammatic approach is fully equivalent to an expansion in orthogonal polynomials [494].

We now have all the necessary ingredients to write down the perturbative bias expansion of BBKS peaks according to Eq. (6.38). Up to second order, we have

$$\delta_{\text{pk}}^L(\boldsymbol{q}) = b_{10}^L \delta_R(\boldsymbol{q}) - b_{01}^L \nabla^2 \delta_R(\boldsymbol{q}) \tag{6.50}$$

$$+ \frac{b_{20}^L}{2} \left(\delta_R^2(\boldsymbol{q}) - \sigma_0^2 \right) - b_{11}^L \left(\delta_R(\boldsymbol{q}) \nabla^2 \delta_R(\boldsymbol{q}) + \sigma_1^2 \right)$$

$$+ \frac{b_{02}^L}{2} \left[\left(\nabla^2 \delta_R \right)^2(\boldsymbol{q}) - \sigma_2^2 \right] + \chi_1^L \left[\left(\boldsymbol{\nabla} \delta_R \right)^2(\boldsymbol{q}) - \sigma_1^2 \right]$$

$$+ \omega_{10}^L \left[\frac{3}{2} \left(\partial_i \partial_j \delta_R - \frac{1}{3} \delta_{ij} \nabla^2 \delta_R \right)^2(\boldsymbol{q}) - \sigma_2^2 \right] + \dots .$$

The absence of Lagrangian tidal shear terms ($\propto K_2(\boldsymbol{q})$, $K_3(\boldsymbol{q})$; see Sec. 5.9) follows from the spherical collapse assumption. This series can be used to compute all the N-point correlation functions of BBKS peaks in Lagrangian space. Furthermore, it is also valid in the presence of weak primordial non-Gaussianity (see Sec. 7.4.3). When expanded in this form, $\delta_{\text{pk}}^L(\boldsymbol{q})$ is generally *not* a count-in-cell (measurable) density, but should be understood as a *mean peak field* given a certain realization of the density field and its derivatives, after averaging over small-scale perturbations. This is because we have not included the stochastic contributions induced by the small-scale perturbations in a given realization, as described in Sec. 2.8. Since the stochastic contributions have a finite variance when averaged over a finite volume, Eq. (6.50) is equivalent to Eq. (6.9) only in the limit of infinite "survey" volume, in which case ergodicity ensures that the series coefficients are the average bias parameters given above. As any perturbative expansion, this effective density contrast formally satisfies $\delta_{\text{pk}}^L(\boldsymbol{q}) \geq -1$ only if all the terms in the infinite series expansion are included.

Expressing the right-hand side of Eq. (6.50) as a Fourier transform and collecting all possible combinations of rotational invariants involving exactly n powers of the linear density contrast field $\delta^{(1)}$, the perturbative peak bias expansion takes the form

$$\delta_{\text{pk}}^L(\boldsymbol{q}) = \sum_{n=1}^{\infty} \frac{1}{n!} \int_{\boldsymbol{k}_1} \cdots \int_{\boldsymbol{k}_n} c_n^L(\boldsymbol{k}_1, \dots, \boldsymbol{k}_n) \left[\delta^{(1)}(\boldsymbol{k}_1) \dots \delta^{(1)}(\boldsymbol{k}_n) + \cdots \right] e^{i\boldsymbol{k}_{1\cdots n} \cdot \boldsymbol{q}} , \tag{6.51}$$

where the Lagrangian peak bias functions $c_n^L(\boldsymbol{k}_1, \dots, \boldsymbol{k}_n)$ are defined at the collapse time τ_0, and the ellipsis in the square bracket stands for terms of order $n - 2$, $n - 4$, and so on. These terms are present because each integrand corresponds to a sum of n-th order renormalized operators written in Fourier space, which contain lower-order operators as counter-terms. For $n = 2$ for instance, the term in square brackets is

$$\left[\delta^{(1)}(\boldsymbol{k}_1) \delta^{(1)}(\boldsymbol{k}_2) - \left\langle \delta^{(1)}(\boldsymbol{k}_1) \delta^{(1)}(\boldsymbol{k}_2) \right\rangle \right] \tag{6.52}$$

so that, in the calculation of $\langle \delta_{\text{pk}}^L(\boldsymbol{q}_1) \delta_{\text{pk}}^L(\boldsymbol{q}_2) \rangle$, we obtain using Wick's theorem (see Appendix A.3)

$$\left\langle \left[\delta^{(1)}(\boldsymbol{k}_1) \delta^{(1)}(\boldsymbol{k}_2) - \left\langle \delta^{(1)}(\boldsymbol{k}_1) \delta^{(1)}(\boldsymbol{k}_2) \right\rangle \right] \left[\delta^{(1)}(\boldsymbol{k}_3) \delta^{(1)}(\boldsymbol{k}_4) - \left\langle \delta^{(1)}(\boldsymbol{k}_3) \delta^{(1)}(\boldsymbol{k}_4) \right\rangle \right] \right\rangle e^{i\boldsymbol{k}_{12} \cdot \boldsymbol{q}_1 + i\boldsymbol{k}_{34} \cdot \boldsymbol{q}_2}$$

$$= \left[\left\langle \delta^{(1)}(\boldsymbol{k}_1) \delta^{(1)}(\boldsymbol{k}_3) \right\rangle \left\langle \delta^{(1)}(\boldsymbol{k}_2) \delta^{(1)}(\boldsymbol{k}_4) \right\rangle + \left\langle \delta^{(1)}(\boldsymbol{k}_1) \delta^{(1)}(\boldsymbol{k}_4) \right\rangle \left\langle \delta^{(1)}(\boldsymbol{k}_2) \delta^{(1)}(\boldsymbol{k}_3) \right\rangle \right] e^{i\boldsymbol{k}_{12} \cdot \boldsymbol{q}_1 + i\boldsymbol{k}_{34} \cdot \boldsymbol{q}_2} . \tag{6.53}$$

Thus, only those correlators that involve fields at different Lagrangian positions remain. Note that, since $\langle \delta^{(1)}(\boldsymbol{k}_1) \delta^{(1)}(\boldsymbol{k}_2) \rangle$ is non-zero only if $\boldsymbol{k}_1 + \boldsymbol{k}_2 = 0$, this term is multiplied by a factor of $c_2^L(\boldsymbol{k}_1, -\boldsymbol{k}_1)$ which precisely corresponds to the second-order zero-lag terms in Eq. (6.50). At order $n \geq 3$, terms of the form $\delta^{(1)}(\boldsymbol{k}_1) \langle \delta^{(1)}(\boldsymbol{k}_2) \delta^{(1)}(\boldsymbol{k}_3) \rangle$, and so on, are correspondingly subtracted in Eq. (6.51). Furthermore, we emphasize that $c_n^L(\boldsymbol{k}_1, \cdots, \boldsymbol{k}_n)$ are *not free functions*, but a convenient way to collect all renormalized Lagrangian bias terms at a given order. Eq. (6.51) with the peak constraint is a particular case of a Lagrangian bias expansion in the iPT framework [79, 113, 141].

149

Armed with this result, it is straightforward to write down the n-point functions of peaks. For the cross-correlation functions between peaks and matter in Fourier space, we have (cf. Sec. 4.1.2)

$$P_{\mathrm{pk},m}^L(k) = c_1^L(k)P_{\mathrm{L}}(k) , \qquad B_{\mathrm{pk},mm}^L = c_2^L(\boldsymbol{k}_1, \boldsymbol{k}_2)P_{\mathrm{L}}(k_1)P_{\mathrm{L}}(k_2) , \tag{6.54}$$

where the renormalized Lagrangian peak "bias functions" are

$$c_1^L(k) \equiv \left(b_{10}^L + b_{01}^L k^2\right) W_R(k) \tag{6.55}$$

$$c_2^L(\boldsymbol{k}_1, \boldsymbol{k}_2) \equiv \left\{ b_{20}^L + b_{11}^L \left(k_1^2 + k_2^2\right) + b_{02}^L k_1^2 k_2^2 - 2\chi_1^L \left(\boldsymbol{k}_1 \cdot \boldsymbol{k}_2\right) \right. \tag{6.56}$$

$$\left. + \omega_{10}^L \left[3\left(\boldsymbol{k}_1 \cdot \boldsymbol{k}_2\right)^2 - k_1^2 k_2^2\right] \right\} W_R(k_1)W_R(k_2) .$$

These definitions agree with the renormalized Lagrangian bias functions introduced by [113] in the context of the integrated perturbation theory (iPT) (see also [495, 151]). Specifically, they are the ensemble averages of functional derivatives [113],

$$c_n^L(\boldsymbol{k}_1, \ldots, \boldsymbol{k}_n) = (2\pi)^{3n} \int_{\boldsymbol{k}'} \left\langle \frac{\mathcal{D}^n \delta_X^L(\boldsymbol{k}')}{\mathcal{D}\delta^{(1)}(\boldsymbol{k}_1) \ldots \mathcal{D}\delta^{(1)}(\boldsymbol{k}_n)} \right\rangle \tag{6.57}$$

$$= \sum_{\alpha_1, \ldots, \alpha_n=1}^D \frac{1}{\overline{n}_X} \left\langle \frac{\mathcal{D}^n n_X}{\mathcal{D}y_{\alpha_1} \ldots \mathcal{D}y_{\alpha_n}} \right\rangle \mathcal{U}_{\alpha_1}(\boldsymbol{k}_1) \ldots \mathcal{U}_{\alpha_n}(\boldsymbol{k}_n) ,$$

where $\delta_X^L \equiv n_X/\overline{n}_X - 1$ is the fractional Lagrangian overdensity of a generic biased tracer, $\boldsymbol{y} = \{y_1, \ldots, y_D\}$ are the linear operators from which the Lagrangian bias relation is constructed, and the kernels $\mathcal{U}_\alpha$ are defined as

$$y_\alpha(\boldsymbol{k}) = \mathcal{U}_\alpha(\boldsymbol{k})\,\delta^{(1)}(\boldsymbol{k}) . \tag{6.58}$$

For the BBKS peak constraint, for instance, $\boldsymbol{y} = \{\nu, \eta_k, \zeta_{ij}\}$ and $D = 10$, with corresponding kernels given by

$$\{U_\alpha(\boldsymbol{k})\} = \left\{ \frac{1}{\sigma_0}W_R(k),\ \frac{i}{\sigma_1}k_k W_R(k),\ -\frac{1}{\sigma_2}k_i k_j W_R(k) \right\} . \tag{6.59}$$

For $n = 1$, Eq. (6.57) then gives

$$c_1^L(\boldsymbol{k}) = \frac{1}{\overline{n}_{\mathrm{pk}}} \left[\frac{1}{\sigma_0} \left\langle \frac{\mathcal{D}n_{\mathrm{pk}}}{\mathcal{D}\nu} \right\rangle + \frac{i}{\sigma_1} \sum_i k_i \left\langle \frac{\mathcal{D}n_{\mathrm{pk}}}{\mathcal{D}\eta_i} \right\rangle - \frac{1}{\sigma_2} \sum_{i,j} k_i k_j \left\langle \frac{\mathcal{D}n_{\mathrm{pk}}}{\mathcal{D}\zeta_{ij}} \right\rangle \right] W_R(k) , \tag{6.60}$$

which, upon taking into account invariance under rotations and symmetrizing the arguments, returns Eq. (6.55) [494]. Higher-order Lagrangian bias functions can be computed analogously.

Eq. (6.54) has exactly recovered the result for the two-point function in Sec. 6.5.1. However, we have also easily obtained the corresponding three-point function. This is equivalent to the result derived for the general bias expansion in Eq. (4.12), with additional higher-derivative terms whose bias coefficients are unambiguously predicted by the peak formalism. Furthermore, the peak auto-power spectrum can formally be written as

$$P_{\mathrm{pk}}^L(k) = [c_1^L(k)]^2 P_{\mathrm{L}}(k) + \sum_{n \geq 2} \frac{1}{n!} \int_{\boldsymbol{p}_1} \cdots \int_{\boldsymbol{p}_n} \left[c_n^L(\boldsymbol{p}_1, \cdots, \boldsymbol{p}_n)\right]^2 P_{\mathrm{L}}(p_1) \cdots P_{\mathrm{L}}(p_n)(2\pi)^3 \delta_D(\boldsymbol{p}_{1\ldots n} - \boldsymbol{k}) + \frac{1}{\overline{n}_{\mathrm{pk}}} , \tag{6.61}$$

in agreement with the iPT approach [79]. It is instructive to compare this result, up to order $n = 2$, to the NLO Eulerian halo power spectrum in the general perturbative bias expansion derived in Sec. 4.1.4. In order to compare at the same order in perturbations, we should only keep the terms $\propto [c_1^L(k)]^2$ and $[c_2^L(\boldsymbol{k}_1, \boldsymbol{k}_2)]^2$

in Eq. (6.61). First, notice that, since the renormalized bias expansion Eq. (6.51) is written in terms of the initial density field, the Lagrangian-space power spectrum of peaks does not contain cross-correlations between linear and cubic contributions, as these are completely absorbed in the leading-order contribution via renormalization (see also Sec. 2.10.1). Second, by construction, the BBKS peak number density does not depend on the tidal field, leading to the absence of $b_{K_2}^L$ in Eq. (6.61). Third, there are no contributions from gravitational evolution, and tidal operators such as K^2 and $O_{\rm td}$ can be consistently set to zero. Fourth, the peak power spectrum contains linear and second-order higher-derivative terms encoded in $c_1^L(k)$ and $c_2^L(\boldsymbol{k}_1, \boldsymbol{k}_2)$; in Eq. (4.22), we have only included the linear higher-derivative term. Finally, the loop integrals in Eq. (6.61) are explicitly regularized through the filtering kernels $W_R(k_i)$, while no such kernel is present in Eq. (4.22). On large scales, the difference amounts to a k-independent constant, which in the general perturbative bias expansion is absorbed by the renormalized stochastic amplitude $P_\varepsilon^{\{0\}}$.

Peaks are not the only example of biased tracers that yield higher-derivative bias terms. Excursion-set approaches can produce such terms as soon as the window is not a sharp k-space filter [305], as we have seen in Sec. 5.9. In all cases, the form of the higher-derivative terms is dictated by the fact that the tracer density is a scalar quantity [113] (see Sec. 2.6).

6.7 Bias parameters from cross-correlations at two smoothing scales*

The peak-background split also works with perturbations of long, but finite wavelength, provided one takes into account the correlation between the long and short modes. Therefore, it should be possible to measure the bias parameters using cross-correlations at two different scales R and R_ℓ, where R corresponds to the halo scale and $R_\ell > R$ (the smoothing scale R_ℓ can take any value as long as it is distinct from the halo smoothing scale). This leads to the "Lagrangian cross-correlation" (LCC) moments discussed at the end of Sec. 4.2. Some of this idea can be traced backed to [152], where the authors showed how to recover the higher-derivative peak bias factors from a conditional mass function. The point was first made clear in [305], who demonstrated that the bias factors b_{ij}^L can be computed from one-point measurements rather than computationally more expensive n-point correlation functions.

This approach was implemented by [303, 304] to halos extracted from N-body simulations in order to test the predictions of peak theory. Namely, halos were traced back to their proto-halo patch in the initial conditions (since one is interested in measuring Lagrangian biases); the linear density field was smoothed on a scale $R_\ell \gtrsim R$, and the quantity $H_n(\nu_\ell = \delta_\ell/\sigma_{0l})$ was computed (for $n = 1, 2$) at the location of each proto-halo. In the peak approach, the overdensity of proto-halos is given by $1 + \delta_h^L = n_{\rm pk}/\overline{n}_{\rm pk}$. Therefore, the average of $H_n(\nu_\ell)$ over all proto-halos reads

$$M_n^{\rm LCC}(R_\ell) \equiv \frac{1}{N_h} \sum_{p=1}^{N_h} H_n[\nu_\ell(\boldsymbol{q}_p)] = \int_{-\infty}^{+\infty} d\nu_\ell\, \mathcal{N}(\nu_\ell) \big\langle 1 + \delta_h^L \big| \nu_\ell \big\rangle H_n(\nu_\ell) \tag{6.62}$$

$$= \frac{1}{\overline{n}_{\rm pk}} \int d^5\boldsymbol{w}\, n_{\rm pk}(\boldsymbol{w})\, (-\epsilon_\nu)^n \left(\frac{\partial}{\partial \nu} + \frac{\epsilon_{J_1}}{\epsilon_\nu} \frac{\partial}{\partial J_1} \right)^n p(\boldsymbol{w})$$

$$= \sum_{i+j=n} \frac{n!}{i!j!} (\sigma_0 \epsilon_\nu)^i (\sigma_2 \epsilon_{J_1})^j b_{ij}^L \,,$$

where $\big\langle 1 + \delta_h^L \big| \nu_\ell \big\rangle$ is the average number density of proto-halos (relative to the mean) in the presence of a long-wavelength density perturbation, $p \in \{1, \cdots, N_h\}$ is the index labeling the proto-halos, and $\epsilon_X = \langle \nu_\ell(\boldsymbol{x}) X(\boldsymbol{x}) \rangle$ denotes the cross-correlation between ν_ℓ and the variables $X = (\nu, J_1)$ defined at the halo smoothing scale. Relations between bias factors of a given order (which arise owing to their close connection with Hermite polynomials, see e.g. [305]) can then be used to extract a measurement of each b_{ij}^L (see [303, 304] and Sec. 4.5 for an overview of the results).

Note that the expression Eq. (6.62) is slightly different from that obtained in the context of the general

---

* This section is of a more technical nature and is not essential for the remainder of this section.

perturbative bias expansion, Eq. (4.40) in Sec. 4.2: instead of the moment $\sigma_0 \epsilon_\nu = \langle \nu_\ell \delta_R \rangle = \langle \delta_\ell \delta_R \rangle / \sigma(R_\ell)$, the general bias expansion yields a proportionality factor of $\langle \delta_\ell \delta \rangle / \sigma(R_\ell)$. The difference is explained by the fact that the peak approach works with the density field smoothed on the peak scale $R(M)$, while the general bias expansion does not perform such a smoothing. The difference between the two is, as expected, absorbed by higher-derivative bias terms. Specifically, noting that Eq. (4.40) assumes that $R_\ell \gg R(M)$ in order for the perturbative approach to be valid, we have

$$\langle \delta_\ell \delta_R \rangle = \int_{\boldsymbol{k}} W_{R_\ell}(k) W_R(k) P_{\mathrm{L}}(k) \approx \int_{\boldsymbol{k}} W_{R_\ell}(k) \left[1 - c_W R^2 k^2 + \cdots \right] P_{\mathrm{L}}(k) = \langle \delta_\ell \delta \rangle + c_W R^2 \langle \delta_\ell \nabla^2 \delta \rangle + \cdots ,$$
$$(6.63)$$

where c_W is a constant depending on the filter employed (for example, $c_W = 1/10$ for a real-space tophat filter). The assumption that $R_\ell \gg R(M)$ implies that the smoothing kernel $W_R(k)$ can be expanded in this integral, since the contribution of modes of order $k \gtrsim 1/R$ is highly suppressed by $W_{R_\ell}(k)$. The last term is precisely a higher-derivative term of the order indicated in Eq. (4.40). Thus, once the smoothing kernel $W_R(k)$ is expanded, the result from the general bias expansion agrees with the peak approach. The only difference is that higher-derivative biases contain the contributions from the smoothing kernel in the general bias expansion, while the latter remain implicit in the peak expressions.

As shown by [306], this approach can be generalized to measure the higher-derivative peak bias factors. The main difference is the appearance of other polynomials and distributions. Consider the bias factors χ_k^L associated with $\eta^2(\boldsymbol{x})$ for illustration. In analogy with Eq. (6.62), the ensemble average of $L_n^{(1/2)}(3\eta_\ell^2)$ at the proto-halo positions is

$$\frac{1}{N_h} \sum_{p=1}^{N_h} L_n^{(1/2)} \left(\frac{3\eta_\ell^2(\boldsymbol{q}_p)}{2} \right) = \int_0^\infty d(3\eta_\ell^2) \, \chi_3^2(3\eta_\ell^2) \left\langle 1 + \delta_h \big| 3\eta_\ell^2 \right\rangle L_n^{(1/2)} \left(\frac{3\eta_\ell^2}{2} \right) . \qquad (6.64)$$

The calculation is somewhat more intricate than for Hermite polynomials because the bivariate Normal distributions (which arise from the conditioning on scale R_ℓ) are replaced by bivariate χ^2-distributions. For $n = 1$ for instance, a little algebra shows that the average of $L_1^{(1/2)}(3\eta_\ell^2/2)$ at the location of proto-halos is

$$\frac{1}{N_h} \sum_{i=1}^{N_h} L_1^{(1/2)} \left(\frac{3\eta_\ell^2(\boldsymbol{q}_p)}{2} \right) = -\epsilon^2 \sigma_1^2 \chi_1^L , \qquad (6.65)$$

where the cross-correlation coefficient ϵ is

$$\epsilon^2 \equiv \frac{\langle \eta^2 \eta_\ell^2 \rangle - \langle \eta^2 \rangle \langle \eta_\ell^2 \rangle}{\sqrt{\left(\langle \eta^4 \rangle - \langle \eta^2 \rangle^2 \right) \left(\langle \eta_\ell^4 \rangle - \langle \eta_\ell^2 \rangle^2 \right)}} = \left(\frac{\sigma_{1\times}^2}{\sigma_{1s} \sigma_{1\ell}} \right)^2 . \qquad (6.66)$$

Here, $\eta_\ell^2 = (\nabla \delta_\ell)^2 / \sigma_{1\ell}^2$ is the long-wavelength perturbation and the subscript "$\times$" denotes the splitting of filtering scales, that is, one filter is on scale R while the second is on scale R_ℓ. Note that, unlike the average of $H_n(\nu_\ell)$, the right-hand side of Eq. (6.65) involves only one bias factor because the variable η^2 does not correlate with the others.

This approach has been applied to measure the Lagrangian LIMD biases [303, 304], the higher-derivative Lagrangian bias χ_1^L, or equivalently $b_{(\boldsymbol{\nabla}\delta)^2}^L$ [306], and the tidal-shear bias $b_{K_2}^L$ [326], by stacking the tidal invariant K_2 defined in Eq. (5.101) at the positions of proto-halos. For a summary of these measurements, see Sec. 4.5.1.

6.8 Excursion-set peaks

In contrast to the excursion-set approach, in which the basic ingredient is random walks at all points in space (see Sec. 5), the peak formalism focuses on the subset of points that correspond to initial density maxima, i.e. peaks. In fact, these two approaches are not irreconcilable as was shown in [396]. Similar ideas

can already be found in the early work of [496]. However, before we discuss the relation between peak theory and excursion sets, we will first address the equivalent of the *cloud-in-cloud* problem in the context of peak theory. The cloud-in-cloud problem in the excursion-set formalism is discussed in detail in Sec. 5.2.3.

So far, in order to make the connection with the abundance of dark matter halos, we have implicitly assumed that there is an one-to-one correspondence between halos of mass M collapsing at redshift z and maxima of the initial density field smoothed on the scale $R(M)$ and of height $\delta_{\rm cr}(z)$. This is a good approximation at least for relatively massive objects (see Sec. 5.1). Therefore, the halo mass function (the number density of halos per logarithmic mass bin $d\ln M$) is given by [497, 496, 410]

$$\overline{n}_h(M)d\ln M = \overline{n}_{\rm pk}(\nu_c)d\nu_c \equiv \frac{\overline{\rho}_m}{M} f_{\rm pk}(\nu_c)d\nu_c \;, \qquad (6.67)$$

where $f_{\rm pk}(\nu_c)d\nu_c$ is the analog of the excursion-set multiplicity function, that is, the fraction of peaks (instead of random walks) at the smoothing scale $R(M)$ with height in the range $[\nu_c, \nu_c + d\nu_c]$. In the case of BBKS peaks (cf. Sec. 6.4),

$$f_{\rm pk}(\nu_c) = \frac{M}{\overline{\rho}_m} \overline{n}_{\rm pk}(\nu_c) = \left(\frac{V}{V_\star}\right) G_0(\gamma_1, \gamma_1 \nu_c) \frac{e^{-\nu_c^2/2}}{\sqrt{2\pi}} \;, \qquad (6.68)$$

plays the role of the multiplicity function. Here, $V = M/\overline{\rho}_m$ is the Lagrangian volume occupied by a halo of mass M, while $V_\star = (2\pi)^{3/2}R_1^3$ is the characteristic volume occupied by a peak on the smoothing scale R. Note that $f_{\rm pk} = f_{\rm pk}(\delta_{\rm cr}, \sigma_i)$ truly is a function of $\delta_{\rm cr}$ and the spectral moments σ_i. Unfortunately, the prescription Eq. (6.67) does not ensure that a peak of height ν_c identified at the filtering scale R is not embedded in a bigger collapsed object.

To remedy this problem, one considers trajectories $\delta(R)$ as a function of R at the location of initial density maxima in analogy with the excursion-set formalism [496, 429, 396]. To ensure that peaks-in-peaks are not counted, we must enforce the constraint that a peak identified in the density field smoothed on scale R is not included in a peak identified on a larger smoothing scale. This *peak-in-peak* effect is difficult to handle because of the correlated nature of the walks $\delta(R)$ for realistic filters such as tophat or Gaussian. Namely, one should in principle consider an infinite number of constraints for each filtering scale between R and infinity. As discussed in Sec. 5.7, this condition can be approximated by the milder requirement that the density contrast of the maxima satisfies

$$\delta(R) > \delta_{\rm cr}, \qquad \delta(R + dR) < \delta_{\rm cr} \;, \qquad (6.69)$$

as proposed by [429]. These two inequalities enforce the condition that maxima of height $\delta_{\rm cr}$ are just at the edge of disappearing. As in Sec. 5.7, they can be combined into the constraint

$$\delta_{\rm cr} < \sigma_0 \nu < \delta_{\rm cr} - \frac{d\delta}{dR}dR \equiv \delta_{\rm cr} - \delta'(R)dR \;, \qquad (6.70)$$

where throughout this section a prime denotes a derivative with respect to the filtering scale. For the condition Eq. (6.70) to make sense, we must have $\delta'(R) \leq 0$ or, equivalently, the random walk $\delta(R)$ must up-cross the threshold $\delta_{\rm cr}$. The analysis of [421] shows that this up-crossing is, to a good approximation, a first-crossing down to peak height of order unity. The combination of the BBKS peak constraint with this up-crossing condition has been dubbed *excursion-set peaks* or ESP by [396].

The ESP approach still suffers from some of the limitations inherent to the excursion-set approach: in particular, as it is solely based on the quantities at the position $\boldsymbol{x}_{\rm pk}$ of the host halo, information about the halo merger history (the presence of local density maxima within the Lagrangian patch, for example) is incomplete. To overcome this, the peak-patch approach of Ref. [435] (see Sec. 5.1) relies on a more sophisticated description of the collapse which considers the behavior of δ (and the other relevant variables) in a neighborhood of $\boldsymbol{q}_{\rm pk}$ and on various filtering scales. This issue becomes important when considering lower-mass halos, where the dynamics increasingly deviates from spherical collapse. Another issue is that any Lagrangian-based prescription of halo abundance and clustering requires spectral moments of the linear power spectrum [see Eq. (6.1)], which are not always finite for the real-space tophat filter. To alleviate this

problem, Refs. [303, 304, 306] have used a Gaussian filter whenever a calculation using tophat filter is not convergent. In practice, the Gaussian filter radius R_G was determined from the tophat filter radius R_T through the requirement

$$\langle \delta_G | \delta_T \rangle = \frac{\langle \delta_G(\boldsymbol{q}) \delta_T(\boldsymbol{q}) \rangle}{\sqrt{\langle \delta_G^2 \rangle \langle \delta_T^2 \rangle}} = \delta_{\mathrm{cr}} \,, \qquad (6.71)$$

where δ_G and δ_T are the density field smoothed with a Gaussian and real-space tophat filter, respectively. This ensures that a peak with $\delta_T = \delta_{\mathrm{cr}}$ also satisfies $\langle \delta_G \rangle = \delta_{\mathrm{cr}}$ on average. Clearly, it would be desirable to have a less ad hoc treatment along the lines of [480] for instance.

The number of virialized objects $\overline{n}_{\mathrm{ESP}}(R)$ per smoothing interval dR is then equal to the number of trajectories that both satisfy the peak constraint and up-cross the threshold δ_{cr} on scale R. For convenience, we introduce the normalized variable $\mu \equiv -\delta'/\sigma_{\delta'}$ with $\sigma_{\delta'} = \sqrt{\langle \delta'^2 \rangle}$, so that the up-crossing condition becomes $\mu \geq 0$. We then have

$$\overline{n}_{\mathrm{ESP}}(R)\,\Delta R = \int_0^\infty d\mu \int_{\nu_c}^{\nu_c + (\sigma_{\delta'}/\sigma_0)\mu \Delta R} d\nu \int_{-\infty}^{+\infty} dJ_1 \cdots \int_{-J_2^{3/2}}^{+J_2^{3/2}} dJ_3 \frac{3^{3/2}}{R_1^3} \big| \det(\zeta_{ij}) \big| \delta_D \big[\boldsymbol{\eta}(\boldsymbol{q}) \big] \Theta_H(\lambda_3)\, p(\boldsymbol{w}) \quad (6.72)$$

$$\approx \frac{3^{3/2}}{R_1^3} \int_0^\infty d\mu \int_{-\infty}^{+\infty} dJ_1 \cdots \int_{-J_2^{3/2}}^{+J_2^{3/2}} dJ_3 \frac{\sigma_{\delta'}}{\sigma_0} \mu \big| \det(\zeta_{ij}) \big| \delta_D \big[\boldsymbol{\eta}(\boldsymbol{q}) \big] \Theta_H(\lambda_3) \delta_D \big[\nu(\boldsymbol{q}) - \nu_c \big] p(\boldsymbol{w})\, \Delta R \,,$$

where the number density of ESP peaks is expressed as a function of R and, in the second line, we have assumed that ΔR is infinitesimal. Further, in the ESP approach the rotational invariants in Eq. (6.11) are supplemented by μ, i.e. $\boldsymbol{w} \equiv \{\nu(\boldsymbol{q}), J_1(\boldsymbol{q}), \mu(\boldsymbol{q}), 3\eta^2(\boldsymbol{q}), 5J_2(\boldsymbol{q}), J_3(\boldsymbol{q})\}$. Using $d\nu_c/dR = -\nu_c \sigma_0'/\sigma_0$ and Eq. (6.17), we can read off the number density of ESP peaks as a function of ν_c from the variable transformation $\overline{n}_{\mathrm{ESP}}(\nu_c) = \overline{n}_{\mathrm{ESP}}(R) dR/d\nu_c$. Therefore, we can define the ESP equivalent of the *local* number density Eq. (6.9) of BBKS peaks of height ν_c as

$$n_{\mathrm{ESP}}(\boldsymbol{q}) = \left(\frac{1}{\nu_c \gamma_{\nu\mu}} \right) \mu\, \Theta_H \big[\mu(\boldsymbol{q}) \big] \times n_{\mathrm{pk}}(\boldsymbol{q}) \,, \qquad (6.73)$$

upon taking advantage of the definition of the cross-correlation between ν and μ: $\gamma_{\nu\mu} = -\sigma_0'/\sigma_{\delta'}$. For a Gaussian filter, μ is precisely equal to $J_1 R \sigma_2/\sigma_{\delta'}$, in which case both δ' and the peak curvature J_1 are perfectly correlated (see also the discussion on assembly bias in Sec. 5.10). Hence, in this specific case, the condition $\lambda_3 \geq 0$ automatically ensures $\mu \geq 0$. This implies that ESP and BBKS peaks are the same for a Gaussian filter.

Eq. (6.73) can be generalized to moving, deterministic or stochastic collapse barriers, $\delta_{\mathrm{cr}} \to B$, upon making the replacements $\delta_D(\nu - \nu_c) \to \delta_D(\nu - \tilde{B})$ and $\Theta_H(\mu) \to \Theta_H(\mu + \tilde{B}')$, where $\tilde{B} \equiv B/\sigma_0$. The second condition constrains the slope of the (correlated) walk to be steeper than that of the barrier at up-crossing. When the barrier is stochastic however, the following two physical pictures are possible, as was pointed out by [306]: either each walk "sees" a moving barrier whose shape changes from peak to peak, or each walk "sees" a constant flat barrier whose height varies as a function of R. In the second case, the condition $\mu \geq -B'$ must be replaced by $\mu \geq 0$. Here, we will follow [306] and adopt the second interpretation. This somewhat simplifies the calculation, while yielding only a percent-level difference.

As discussed in Sec. 5.9, each variable which the collapse threshold B depends on beyond the density adds a dimension to the first-crossing problem, and introduces scatter in the (σ, δ)-plane. To account for the scatter in the collapse threshold, Ref. [303, 306] considered the square-root barrier Eq. (5.107), $\tilde{B} = \nu_c + \beta$, where the stochastic variable β follows a distribution $p(\beta)$ which is represented by a lognormal. This furnishes a good description of the actual collapse thresholds as a function of halo mass [318]. Using the definition of
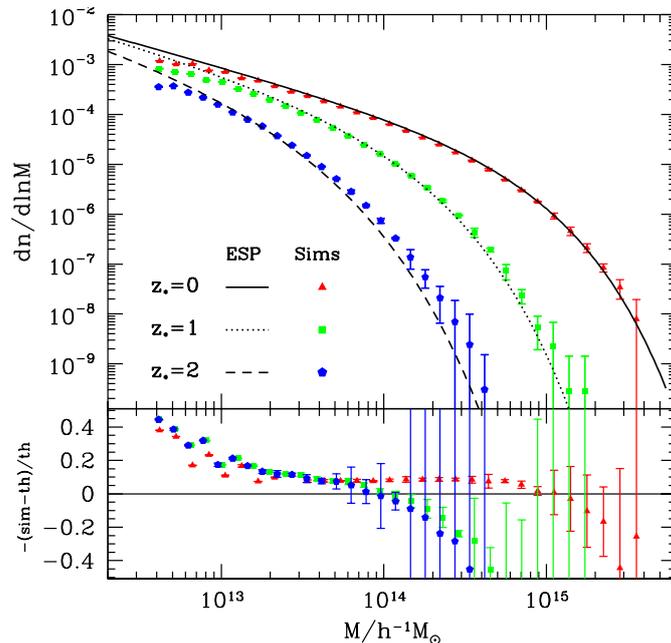
Figure 30: *Top panel*: Logarithmic mass function of SO halos (with $\Delta_{\rm SO} = 200$) extracted from N-body simulations. Different symbols refer to different redshifts as indicated in the figure. The solid, dotted and dashed curves represent the ESP prediction at $z = 0$, 1 and 2. *Bottom panel*: Fractional deviation of the simulations from the ESP prediction. In both panels, error bars denote the scatter among realizations. From [499].

$f(J_1)$, Eq. (6.22), the multiplicity function of ESP peaks reads[19]

$$f_{\rm ESP}(\nu_c) = V \overline{n}_{\rm ESP}(\nu_c) = \left(\frac{V}{V_\star} \right) \frac{1}{\nu_c \gamma_{\nu\mu}} \int d\beta \, p(\beta) \int_0^\infty d\mu \, \mu \int_0^\infty dJ_1 \, f(J_1) \mathcal{N}(\nu_c + \beta, J_1, \mu) \,, \qquad (6.74)$$

where $\mathcal{N}(\nu, J_1, \mu)$ denotes a trivariate normal distribution with vanishing mean and unit variances. In addition to the terms appearing in the bias expansion Eq. (6.50) of BBKS peaks discussed earlier in this section, the expansion for ESP peaks will include terms depending on δ'.

However, a dependence on tidal fields, $K_2(\boldsymbol{q})$, $K_3(\boldsymbol{q})$ as defined in Sec. 5.9, will arise only if deviations from the spherical collapse are modeled explicitly. Such extensions of the peak constraint (along the line of [121] and, recently, [498]) are desirable since they would provide a physical description as well as predictions for the scatter β. As we shall see in Sec. 7.4, a first-principle description of the scatter in Lagrangian bias models is in fact essential for the non-Gaussian bias consistency relation.

Using Eq. (6.67), the halo mass function predicted by the ESP approach eventually is

$$\overline{n}_h(M) = \overline{\rho}_m \, f_{\rm ESP}(\nu_c) \frac{d\nu_c}{dM} = -\frac{1}{3} M R \left(\frac{\sigma_{\delta'} \gamma_{\nu\mu} \nu_c}{\sigma_0} \right) V^{-1} f_{\rm ESP}(\nu_c) \,. \qquad (6.75)$$

One should bear in mind that, when more than one filters are employed to calculate the rotational invariants (to ensure the convergence), then the multiplicative factor of R in the above expression is the smoothing radius of the normalized density field ν. For instance, in the ESP implementation of [303, 306], a tophat filter is used to define ν and μ, whereas a Gaussian filter is used for the variables J_1, η^2, J_2 and J_3 in order to ensure the convergence of σ_1 and σ_2. In this case, R would be the tophat radius.

---

[19]This expression can be simplified with Bayes' theorem, $\mathcal{N}(\nu, J_1, \mu) = \mathcal{N}(\nu, J_1)\mathcal{N}(\mu|\nu, J_1)$, so that the integral $\int_0^\infty d\mu \, \mu \, \mathcal{N}(\mu|\nu, J_1)$ analogous to Eq. (5.82) takes a compact form similar to Eq. (5.85).

Predictions for the halo mass function based on Eq. (6.75) were presented in [500], while [501, 502] discussed the implications for halo mass accretion histories. Ref. [303] included a halo mass definition which uses a tophat filter in real space together with the mean dependence and scatter of the critical collapse threshold on the halo mass. Comparison with N-body simulations shows that the ESP model provides a good fit to the mass function of spherical-overdensity (SO) halos (identified with a constant overdensity threshold $\Delta_{\mathrm{SO}} = 200$) if one assumes a simple square-root barrier $B(s) = \delta_{\mathrm{cr}} + \beta\sqrt{s}$ with lognormal scatter $p(\beta)$ [303, 306]. Fig. 30 indeed shows that the ESP prediction fares reasonably well at redshift zero, which was used to calibrate the free parameters describing the square-root barrier. Notwithstanding, it underestimates the abundance of massive halos at higher redshift. Note, however, that the figure is generated with the same mean $\langle\beta\rangle$ and variance $\mathrm{Var}(\beta)$ at all redshifts [499], even though these were inferred from halos which virialized at $z = 0$ only. It is likely that the mean and variance in the linear collapse threshold depend on redshift. Including this dependence may improve the agreement with simulations at higher redshift.

Finally, the results described in Sec. 6.7 on the estimation of Lagrangian bias parameters directly apply to the ESP model, provided X is replaced by $X = (\nu, J_1, \mu)$ and n_{pk} by n_{ESP}.

6.9 Gravitational evolution of Lagrangian density peaks

In this section, we show how the peak statistics in Lagrangian space can be related to the Eulerian (late-time) statistics by approximately modeling gravitational evolution. Specifically, we will assume that the halo centers are test particles that do not interact with each other and that locally flow with the dark matter. We will begin with a derivation of the velocity bias of peaks, and use that in conjunction with the continuity equation to derive the time evolution of the linear LIMD and higher-derivative peak bias from initial Lagrangian space to Eulerian space ($b_{01}^L \to b_{01}^E$). Next, we will discuss how a phase-space treatment can predict the time evolution of the Lagrangian peak correlation function at any order. We will employ the Zel'dovich approximation for simplicity. Because of this, we do not recover the correct gravitational evolution discussed in Sec. 2.3–2.4 at second and higher order [503, 504, 505]. Nevertheless, it illustrates the possibilities and challenges for including gravitational evolution in the peak formalism.

6.9.1 Velocity bias

The peak model is an explicit realization of biased tracers that have biased velocities. The velocity bias arises from the correlation between linear velocities and density gradient. In particular, the linear-order, 3-dimensional velocity dispersion of Lagrangian density peaks is [13]

$$\sigma_{v,\mathrm{pk}}^2(\tau) \equiv \langle \boldsymbol{v}_{\mathrm{pk}}^2(\boldsymbol{q}, \tau)\rangle = (\mathcal{H}f)^2 \sigma_{-1}^2 \left(1 - \gamma_0^2\right) , \tag{6.76}$$

where γ_0 is defined in Eq. (6.3). Thus, the velocity dispersion of peaks is smaller than that of the dark matter filtered on the same scale, $\langle \boldsymbol{v}_R^2\rangle = (\mathcal{H}f)^2 \sigma_{-1}^2$. It is because large-scale flows are more likely to be directed toward peaks than to be oriented randomly [13, 506, 474, 475]. An important consequence is that the linear peak velocities are biased according to the relation

$$\boldsymbol{v}_{\mathrm{pk}}(\boldsymbol{q}, \tau) = \boldsymbol{v}_R(\boldsymbol{q}, \tau) - (\mathcal{H}f)\frac{\sigma_0^2}{\sigma_1^2}\boldsymbol{\nabla}\delta_R(\boldsymbol{q}, \tau) , \tag{6.77}$$

which can be inferred from a calculation of peak pairwise velocity statistics or the redshift-space linear power spectrum in the distant-observer limit [156, 257]. Furthermore, this also implies that the linear peak displacement $\boldsymbol{s}_{\mathrm{pk}}$ is biased [152, 174]. Namely, $\boldsymbol{s}_{\mathrm{pk}}$ is given in Fourier space by

$$\boldsymbol{s}_{\mathrm{pk}}(\boldsymbol{k}, \tau) = \left(1 - \frac{\sigma_0^2}{\sigma_1^2}k^2\right) W_R(k)\boldsymbol{s}_{(1)}(\boldsymbol{k}, \tau) , \tag{6.78}$$

where $\boldsymbol{s}_{(1)}$ is the linear displacement field [see Eq. (B.16)]. This defines the time-independent linear peak velocity bias $c_{v,\mathrm{pk}}(k)$ [156, 257]

$$c_{v,\mathrm{pk}}(k) = \left(1 - \frac{\sigma_0^2}{\sigma_1^2}k^2\right) W_R(k) , \tag{6.79}$$

such that $s_{\rm pk}(\boldsymbol{k},\tau) = c_{v,\rm pk}(k)s_{(1)}(\boldsymbol{k},\tau)$. In the notation of Sec. 2.7, $\beta_{\nabla^2 \boldsymbol{v}} = \sigma_0^2/\sigma_1^2$ plus a contribution from the expansion of the filtering kernel $W_R(k)$. Note that $c_{v,\rm pk}$ depends on k and on the smoothing scale R, but not on the peak height ν nor on time; equivalently, $\beta_{\nabla^2 \boldsymbol{v}}$ only depends on the radius R. Since $\sigma_0/\sigma_1 \propto R$, which is the only scale in the problem, we expect this velocity bias to increase with halo mass. As we can see, $s_{\rm pk}$ differs from the matter displacement $s_{(1)}$ by the filtering kernel, which arises from the fact that halo internal motions do not affect the motion of the host halo, and by the factor of $(1 - k^2\sigma_0^2/\sigma_1^2)$.

It is important to distinguish between *local* and *statistical* velocity bias. The former means that the velocities of biased tracers are locally different than that of the dark matter; that is, a local observer would measure a relative velocity between the tracer and the dark matter. The latter implies that the tracers locally flow with the dark matter, but their velocities are statistically biased. This can occur essentially as a selection effect, if the tracers reside in special locations where the velocity is smaller or larger than at randomly chosen locations. In the peak approach, both effects are present in the form of the window function $W_R(k)$ (local part) and the factor $1 - (\sigma_0/\sigma_1)^2 k^2$ (statistical part).

Let us now discuss in more detail the relation between Eq. (6.76) and Eq. (6.77). Clearly, taking the ensemble average of the square of Eq. (6.77) yields Eq. (6.76), as it should. However, Ref. [257] also pointed out that, for Eq. (6.79),

$$
\begin{aligned}
(\mathcal{H}f)^{-2}\sigma_{v,\rm pk}^2 &= \frac{1}{2\pi^2}\int_0^\infty dk\, P_{\rm L}(k)\, c_{v,\rm pk}^2(k) \\
&= (\mathcal{H}f)^{-2}\langle \boldsymbol{v}_R \cdot \boldsymbol{v}_{\rm pk}\rangle = \frac{1}{2\pi^2}\int_0^\infty dk\, P_{\rm L}(k)\, c_{v,\rm pk}(k)W_R(k)\ .
\end{aligned}
\tag{6.80}
$$

That is, the variance of peak velocities, involving an integral over $c_{v,\rm pk}^2$, is equal to the peak–smoothed dark matter velocity variance, which involves $c_{v,\rm pk}$ only. This identity indicates that, at the peak position, the velocities of the peak and the filtered matter distribution are the same. This follows from our assumption that the peaks locally flow with the smoothed dark matter.

A further requirement on the peak velocity bias is that, in real space, it has to be a higher-derivative operator which maps a vector (velocity) field onto another vector field. Homogeneity and isotropy then require that it is built from powers of the Laplacian ∇^2, or equivalently, powers of k^2 in Fourier space. Therefore, we generically expect the lowest-order k-dependence to scale as $c_{v,\rm pk}(k) = (1 - R_v^2 k^2)W_R(k)$, for some constant R_v. This second requirement, together with Eq. (6.80), then uniquely yields the velocity bias Eq. (6.79). Note that, for instance, the choice $c_{v,\rm pk}(k) = [1 - (\sigma_{-1/2}/\sigma_0)^2\, k]W_R(k)$ also satisfies Eq. (6.80), but does not arise from a local operator in real space.

6.9.2 Linear evolution: continuity equation

Having discussed the origin of the peak velocity bias, we proceed forward and compute the time evolution of the linear peak bias using the continuity argument invoked in [152]. This generalizes the discussion of [115] to a linearly biased displacement/velocity field, and parallels the discussion in Sec. 2.7. Integrating the linear continuity equation $\partial\delta_{\rm pk}/\partial\tau = -\boldsymbol{\nabla} \cdot \boldsymbol{v}_{\rm pk}$, and evaluating the result in Fourier space, the first-order Eulerian peak bias function reads

$$
c_1^E(k,\tau) = c_{v,\rm pk}(k) + \frac{D(\tau_0)}{D(\tau)}c_1^L(k,\tau_0)\ ,
\tag{6.81}
$$

where the first-order Lagrangian peak bias function is defined in Eq. (6.55) and we have momentarily restored the time-dependence of c_1^L for clarity. This relation is a particular case of Eq. (2.82). Note however that, unlike in Sec. 2.3 and Sec. 2.7, τ parametrizes the trajectory of the peak-patch from the Lagrangian ($\tau = 0$) initial conditions) to Eulerian (collapse epoch at $\tau = \tau_0$) space (though the peak trajectory can formally be extended beyond τ_0). Therefore,

$$
c_1^E(k,\tau)\delta_R(\tau) = c_1^E(k,\tau)\frac{D(\tau)}{D(\tau_0)}\delta_R(\tau_0) \stackrel{\tau \to 0}{=} c_1^L(k,\tau_0)\delta_R(\tau_0)\ .
\tag{6.82}
$$

That is, we recover the Lagrangian bias Eq. (6.55) in the limit $\tau \to 0$ when peaks are defined relative to the density field linearly extrapolated to τ_0, which is our convention throughout this section. The Eulerian peak bias c_1^E is scale-independent in the limit $kR \ll 1$, as expected from "local bias theorems" [104, 105]. Eq. (6.81) represents the fact that peaks stream towards (or move apart from) each other in high (low) density environments. This effect is higher order in derivatives, as any velocity bias (statistical or local) has to be.

On splitting Eq. (6.81) into its higher-derivative ($\propto k^2$) and local (k-independent) terms, we obtain

$$b_{10}^E(\tau) \equiv 1 + \frac{D(\tau_0)}{D(\tau)}\, b_{10}^L(\tau_0), \qquad b_{01}^E(\tau) \equiv -\frac{\sigma_0^2}{\sigma_1^2} + \frac{D(\tau_0)}{D(\tau)}\, b_{01}^L(\tau_0)\,. \tag{6.83}$$

The first relation is the usual relation Eq. (2.32) for the Eulerian, linear bias [15]. The second relation implies that, in the idealized limit $\tau \to \infty$, the higher-derivative bias parameter approaches the negative, R-dependent constant $-\sigma_0^2/\sigma_1^2$ as structure grows and $D^{-1}(\tau)$ shrinks. The higher-derivative contribution to the linear peak bias function $c_1^E(k,\tau)$ thus persists at late time if the linear velocities are statistically biased [152]. Note that since b_{01}^L is always positive (see Fig. 29), the contribution from the linear velocity bias suppresses the amplitude of $b_{01}^E(\tau_0)$. In fact, $b_{01}^E(\tau_0)$ even becomes negative for peaks corresponding to $M \lesssim 5 \times 10^{13}\, h^{-1}\, M_\odot$.

A temporally constant velocity bias might seem to be at odds with the prediction of a two-fluid calculation (dark matter and halos) which yields a decaying velocity bias [118]. However, as we discuss in Sec. 2.7, these two results simply correspond to different assumptions about the time dependence of the force difference between matter and halos; Eq. (6.83) assumes a relative force that is constant in time, while the calculation of [118] assumed an instantaneous force in the initial conditions. Recent work has shown conclusive evidence for a statistical bias of the form Eq. (6.79) in two-point statistics of the initial halo velocities [259, 174]. Since, in the linear regime, the acceleration is parallel to the initial velocity, this implies that the linear velocity bias $c_{v,\text{pk}}$ does not decay, but remains constant throughout time. In other words, the gravitational force acting on initial density peaks is biased at the linear level [174, 507] (see the discussion in Sec. 2.7). The recent measurements of [174], which are not volume weighted and thus less plagued by numerical artifacts (see [173] for a discussion), appear consistent with this interpretation.

Finally, we have thus far assumed that the peak velocity is equal to that of the smoothed dark matter component evaluated at the peak position. In practice however, the halo velocity is commonly computed as the center-of-mass velocity of all the particles belonging to the halo. Assuming that effects related to the finite extent of dark matter halos introduce a *time dependence* in the window, i.e. $W_R \to W(\boldsymbol{k}, \tau)$, the halo velocity bias at any time $\tau \geq \tau_i$ reads [329]

$$c_v(k,\tau) = W(k,\tau) + \big(c_v(k,\tau_i) - W(k,\tau_i)\big)\left(\frac{D(\tau_i)}{D(\tau)}\right)^{3/2} - D^{-3/2}(\tau)\int_{\tau_i}^{\tau} d\tau'\, W'(k,\tau') D^{3/2}(\tau')\,, \tag{6.84}$$

where $W' = \partial W/\partial \tau$. There are thus two contributions to $c_v(k,\tau)$: one from W [i.e. the smoothing $W_R(k)$ in Eq. (6.79)] and one from W'. The second contribution proportional to W' dominates for $\tau \simeq \tau_i$, but is negligible at late time $\tau \gg \tau_i$. Therefore, any measurement of a higher-derivative term, apart from that induced by the window function W_R, in the velocities of proto-halos should be attributed to a statistical velocity bias. This furnishes a clean way of disentangling statistical biases from effects induced by the time dependence of the halo profile.

6.9.3 Evolution at higher order

We now compute the evolution of the peak two-point statistics using the phase-space distribution of peaks, that is, the joint distribution of peak velocities. Assuming that each halo center is in one-to-one correspondence with a peak, the Eulerian comoving position and proper velocity of a halo can in general be expressed through the mapping

$$\boldsymbol{x}_{\text{pk}}(\tau) = \boldsymbol{q}_{\text{pk}} + \boldsymbol{s}(\boldsymbol{q}_{\text{pk}}, \tau)\,, \qquad \boldsymbol{v}_{\text{pk}}(\tau) = a(\tau)\, \frac{\partial}{\partial \tau} \boldsymbol{s}(\boldsymbol{q}_{\text{pk}}, \tau)\,, \tag{6.85}$$

where $\boldsymbol{q}_{\rm pk}$ is the initial position of the halo center, and $\boldsymbol{s}(\boldsymbol{q},\tau)$ is the displacement. For simplicity, we will work within the Zel'dovich approximation [508, 509], but the results can be extended to include higher-order Lagrangian displacements. In this approximation, the peaks displacement is given by Eq. (6.78),

$$\boldsymbol{s}_{\rm pk}(\boldsymbol{q},\tau) = -\frac{D(\tau)}{D(\tau_0)}\left(\frac{\boldsymbol{\nabla}}{\nabla^2}\delta_R + \frac{\sigma_0^2}{\sigma_1^2}\boldsymbol{\nabla}\delta_R\right)(\boldsymbol{q},\tau_0) \tag{6.86}$$

$$\frac{\partial}{\partial\tau}\boldsymbol{s}_{\rm pk}(\boldsymbol{q},\tau) = -f\mathcal{H}\frac{D(\tau)}{D(\tau_0)}\left(\frac{\boldsymbol{\nabla}}{\nabla^2}\delta_R + \frac{\sigma_0^2}{\sigma_1^2}\boldsymbol{\nabla}\delta_R\right)(\boldsymbol{q},\tau_0)\,, \tag{6.87}$$

where we have explicitly written the dependence on τ_0 to emphasize the fact that the linear fields are normalized relative to the collapse epoch.

Consider now an ensemble of realizations of some Lagrangian, biased point process. The correlation function $\xi_{\rm pk}^L(r,\tau)$ is related to the zeroth moment of the joint probability $p_2(\tilde{\boldsymbol{v}}_1,\tilde{\boldsymbol{v}}_2;\boldsymbol{r},\tau|\text{pk})$ to have a pair of peaks separated by a distance $\boldsymbol{r}$ and with normalized velocities $\tilde{\boldsymbol{v}}_1$ and $\tilde{\boldsymbol{v}}_2$ [510, 152],

$$\overline{n}_{\rm pk}^2\left[1 + \xi_{\rm pk}^L(r,\tau)\right] = \int d^3\tilde{\boldsymbol{v}}_1 d^3\tilde{\boldsymbol{v}}_2\, p_2(\tilde{\boldsymbol{v}}_1,\tilde{\boldsymbol{v}}_2;\boldsymbol{r},\tau|\text{pk})\,, \tag{6.88}$$

where we have introduced the scaled velocity $\tilde{\boldsymbol{v}} = \boldsymbol{v}/\mathcal{H}f\sigma_v$, where $\sigma_v \equiv \sigma_{-1}(z)$ is proportional to the 3-dimensional variance of the matter velocity field, such that $\langle\tilde{\boldsymbol{v}}^2\rangle = 1$. When the peak motions are governed by Eq. (6.87), $p_2(\tilde{\boldsymbol{v}}_1,\tilde{\boldsymbol{v}}_2;r,\tau|\text{pk})$ can be easily related to the joint probability distribution at the initial time $\tau_i \ll \tau_0$,

$$p_2(\tilde{\boldsymbol{v}}_1,\tilde{\boldsymbol{v}}_2;\boldsymbol{r},\tau|\text{pk}) = \int d^3\boldsymbol{q}\,\delta_D(\boldsymbol{q} + \sigma_v[\tilde{\boldsymbol{v}}_2 - \tilde{\boldsymbol{v}}_1] - \boldsymbol{r})\,p_2(\tilde{\boldsymbol{v}}_1,\tilde{\boldsymbol{v}}_2;\boldsymbol{q},\tau_i|\text{pk})\,. \tag{6.89}$$

Here, $\boldsymbol{q}$ is the Lagrangian separation vector. Note that the 2-point correlation function depends only on r, even though the probability $p_2(\tilde{\boldsymbol{v}}_1,\tilde{\boldsymbol{v}}_2;\boldsymbol{r},\tau|\text{pk})$ also depends on $\tilde{\boldsymbol{v}}_i \cdot \boldsymbol{r}$.

We have implicitly assumed that each observed tracer corresponds to a unique Lagrangian patch. Like in the peak-patch picture of [435] (see Sec. 5.1), the merging history depends on the details of the small-scale matter distribution inside the patch. Here, however, we are not following the merging of sub-clumps, but rather the motion and collapse of the patch as a whole. Given that the number of patches is conserved by definition, Eq. (6.89) is equivalent to Liouville's theorem, which states that the phase space density of conserved tracers is conserved.

Eq. (6.89) is especially useful when one knows how to calculate distributions in the initial conditions. For a bias prescription that can be specified through a finite number of constraints, e.g. for the BBKS peaks considered here,

$$p_2(\tilde{\boldsymbol{v}}_1,\tilde{\boldsymbol{v}}_2;\boldsymbol{q},\tau_i|\text{pk}) = \int d^{13}\boldsymbol{u}_1\int d^{13}\boldsymbol{u}_2\,n_{\rm pk}(\boldsymbol{u}_1)n_{\rm pk}(\boldsymbol{u}_2)\,p_2(\boldsymbol{u}_1,\boldsymbol{u}_2;\boldsymbol{q},\tau_i)\,, \tag{6.90}$$

where $\boldsymbol{u}$ is a vector containing the 13 variables that describe both the BBKS peak constraint and the initial relative velocity (i.e., $\boldsymbol{u} = \{\nu,\eta_k,\zeta_{ij},\tilde{v}_l\}$), and p_2 is the joint-probability of $\boldsymbol{u}_1 = \boldsymbol{u}(\boldsymbol{x}_1)$ and $\boldsymbol{u}_2 = \boldsymbol{u}(\boldsymbol{x}_2)$. Note that the number density $n_{\rm pk}$ of BBKS peaks, Eq. (6.9), does not depend on $\tilde{v}_i$ as this is prohibited by the equivalence principle (Sec. 2.10.2). For Gaussian initial conditions, p_2 is a multivariate Gaussian. In this case, expressing the Dirac delta in Eq. (6.89) as the Fourier transform of a uniform distribution and integrating out the velocities, the 2-point correlation function of discrete tracers can eventually be written as

$$\overline{n}_{\rm pk}^2\left[1 + \xi_{\rm pk}(r,\tau)\right] = \int_{\boldsymbol{k}}\int d^3\boldsymbol{q}\,e^{i\boldsymbol{k}\cdot(\boldsymbol{r}-\boldsymbol{q})}\int d^{10}\boldsymbol{y}_1\int d^{10}\boldsymbol{y}_2\,n_{\rm pk}(\boldsymbol{y}_1)n_{\rm pk}(\boldsymbol{y}_2)\,p_2(\boldsymbol{y}_1,\boldsymbol{y}_2;\boldsymbol{q},\tau_i) \tag{6.91}$$

$$\times\exp\left(-\frac{1}{2}\sigma_v^2\,\boldsymbol{k}^\top C\,\boldsymbol{k} + i\sigma_v\boldsymbol{k}\cdot\Delta\boldsymbol{s}\right)\,,$$

where $p_2(\boldsymbol{y}_1,\boldsymbol{y}_2;\boldsymbol{q},z_i)$ is the joint PDF for $\boldsymbol{y} = \{\nu,\eta_k,\zeta_{ij}\}$ at Lagrangian position $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$, $\Delta s_i = \langle(s_1 - s_2)_i|\text{pk}\rangle$ is the relative peak displacement, and $C_{ij} \equiv \langle(s_1 - s_2)_i(s_1 - s_2)_j|\text{pk}\rangle - \Delta s_i\Delta s_j$ is its covariance

matrix. Note that the Gaussian integral over velocities introduces the displacement covariance matrix, rather than its inverse.

Eq. (6.91) is the exact result within the Zel'dovich approximation. Therefore, it is important to realize that, even though momentum conservation is ensured at leading order only, our approximation is invariant under generalized Galilei transformations [198, 199, 200] — that is, uniform, but time-dependent boosts, see Sec. 2.10.2 for the explicit definition — since relative displacements are unchanged. This ensures that the effect of very-long-wavelength perturbations vanishes in the equal-time, 2-point peak correlation function Eq. (6.91) [511, 201, 203, 271].

Following [152], the integrand of Eq. (6.91) can be expanded in powers of the linear power spectrum $P_L(k)$ such that, at NLO in the Zel'dovich approximation, we have

$$\xi_{\rm pk}(r,\tau) = \left(\frac{D(\tau)}{D(\tau_0)}\right)^2 \int_{\boldsymbol{k}} e^{-\frac{1}{3}k^2\sigma_{v,\rm pk}^2(z)} \left[c_1^E(k,\tau)\right]^2 P_L(k)\, e^{i\boldsymbol{k}\cdot\boldsymbol{r}} + \xi_{\rm MC}(r,\tau) + \mathcal{O}(3)\,. \qquad (6.92)$$

Here, c_1^E is defined in Eq. (6.81), and $\mathcal{O}(3)$ denotes the NNLO (next-to-next-leading order, or two- and higher-loop) contributions, the linear power spectrum P_L is evaluated at the collapse time τ_0, and the NLO mode-coupling term $\xi_{\rm MC}(r,\tau)$ is given by

$$\xi_{\rm MC}(r,\tau) \equiv \int_{\boldsymbol{k}} P_{\rm MC}(k,\tau)e^{i\boldsymbol{k}\cdot\boldsymbol{r}}$$

$$P_{\rm MC}(k,\tau) \equiv \frac{1}{2}\left(\frac{D(\tau)}{D(\tau_0)}\right)^4 \int_{\boldsymbol{k}_1}\int_{\boldsymbol{k}_2} \left[c_2^E(\boldsymbol{k}_1,\boldsymbol{k}_2,\tau)\right]^2 P_L(k_1)P_L(k_2)\,(2\pi)^3\delta_D(\boldsymbol{k}-\boldsymbol{k}_{12})\,. \qquad (6.93)$$

The quadratic (Eulerian) peak bias function c_2^E is given by

$$c_2^E(\boldsymbol{k}_1,\boldsymbol{k}_2;\tau) = \frac{1}{2}\left[1 + \left(\frac{k_1}{k_2}+\frac{k_2}{k_1}\right)\mu + \mu^2\right]c_{v,\rm pk}(k_1)c_{v,\rm pk}(k_2) + \frac{D(\tau_0)}{D(\tau)}c_1^L(k_1)c_{v,\rm pk}(k_2)\left(1+\frac{k_1}{k_2}\mu\right) \quad (6.94)$$

$$+ \frac{D(\tau_0)}{D(\tau)}c_1^L(k_2)c_{v,\rm pk}(k_1)\left(1+\frac{k_2}{k_1}\mu\right) + \left(\frac{D(\tau_0)}{D(\tau)}\right)^2 c_2^L(\boldsymbol{k}_1,\boldsymbol{k}_2)\,.$$

Therefore, the peak constraint yields a 2-point correlation function (Fig. 31) consistent with that inferred in Lagrangian PT (e.g. [510, 117, 512, 79]) with the additional feature that the Lagrangian bias functions contain higher-derivative terms, and the displacement to the Eulerian position is modified owing to a velocity bias. Note also that, although [152] did not consider Lagrangian tidal shear bias, it is straightforward to include this contribution into c_2^L following [513].

It should be noted that, unlike semi-analytic methods which resum part of the perturbative expansion (such as e.g. RPT [73]), our perturbative solution is the series expansion of the exact, Zel'dovich approximated correlation function. Therefore, even though our truncated perturbative expansion formally violates Galilei-invariance at any order (see [203] for a related discussion), it will automatically satisfy Galilei-invariance on the scales where the expansion has converged towards the full Zel'dovich result [514]. At NLO, the convergence is achieved at the 1% level across the BAO. Of course, the mode-coupling is not given correctly due to the Zel'dovich approximation. To remedy this problem, one has to consider higher-order displacements in LPT [135, 503, 512].

We now discuss the relation of this result to the NLO halo power spectrum in the general bias expansion derived in Sec. 4.1.4. First, the bias functions c_1^E and c_2^E contain a subset of the local bias terms obtained in the general bias expansion, as discussed in Sec. 6.6.2. However, they contain additional higher-derivative terms, which are higher order following the perturbative counting in Sec. 4.1.4 (see the caveats of the counting mentioned there). On the other hand, due to the Zel'dovich treatment of nonlinear evolution and the absence of Lagrangian tidal bias, Eqs. (6.92)–(6.93) do not contain contributions from the coupling of cubic and linear terms. Further, unlike the results in Sec. 4.1.4, where the loop momentum attains arbitrarily high values, in Eqs. (6.92)–(6.93) the integrals are cut off on the scale $k \sim R^{-1}$ by the filtering kernel $W_R(k)$. Finally, unlike the contributions $\mathcal{I}^{[O,O']}(k)$ appearing in Sec. 4.1.4, whose low-k limits are subtracted, here
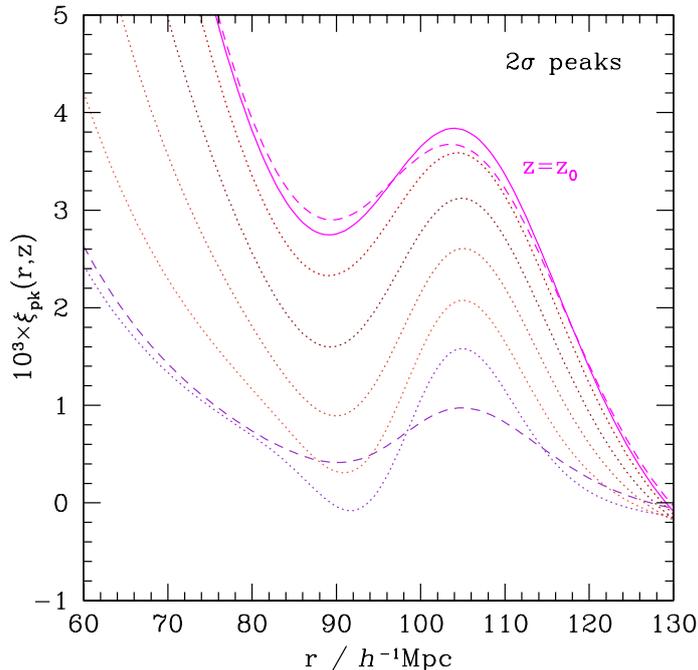
Figure 31: Redshift evolution of the correlation function of $\nu_c = 2$ (2σ) peaks collapsing at $z_0 = 0.3$ as predicted by Eq. (6.92). The curves from bottom to top represent $\xi_{\rm pk}(r, z)$ at redshift $z = \infty$, 5, 2, 1, 0.5 (dotted curves) and $z = z_0$ (solid curve). Only the correlation function at the collapse epoch ($z = z_0$) can be measured in real data. For comparison, the dashed curves show the correlation function at $z = \infty$ and $z = z_0$ in a LIMD bias approximation. From [152].

the mode-coupling term contributes a non-vanishing white noise in the limit $k \to 0$ which adds to the Poisson noise $1/\overline{n}_{\rm pk}$. There are infinitely many such white-noise contributions induced by higher-order terms in the bias expansion. All these non-vanishing contributions in the limit $k \to 0$ renormalize $1/\overline{n}_{\rm pk}$ into an effective white noise term, which can generally be super- or sub-Poissonian depending on the halo mass considered. This effect physically arises from small-scale exclusion and, therefore, cannot be modeled perturbatively. Note however that, in the peak approach, the "renormalized" white noise amplitude in the limit $k \to 0$ can be obtained from a numerical evaluation of $\int d^3 r \xi_{\rm pk}^L(r)$ through Eq. (6.34) (see the discussion in Sec. 4.5.3 and [178, 347]).

Furthermore, the contribution proportional to $c_1^L c_2^L \partial^{-1}\delta \cdot \partial\delta$, where c_n^L are Lagrangian bias functions, shifts the position of the BAO peak, as was first pointed out in [269, 515]; this term is of course also present in the NLO contribution to the two-point function derived in the general bias expansion (Sec. 4.1.4). The higher-derivative correction to the second-order bias function does not change the strength of the effect significantly, which is dominated by the quadratic LIMD bias b_{20}^L. Importantly, this nonlinear shift can be either accounted for in a forward analysis [269, 516], or reversed using a suitable reconstruction of the linear displacement field [517]. Alternatively, the "linear point", which lies midway between the dip at $r \sim 90 \, h^{-1}$ Mpc and the BAO peak at $r \sim 105 \, h^{-1}$ Mpc, is weakly affected by nonlinearities and could also be used for distance measurements [518, 519, 520].

Before concluding, let us illustrate how gravitational motions from the initial to final time wash out most of the Lagrangian higher-derivative bias induced by the peak constraint, by considering the peak correlation function $\xi_{\rm pk}(r, z)$ around the BAO scale (see Sec. 6.5.2). Fig. 31 shows the redshift evolution of $\xi_{\rm pk}(r, z)$, as predicted by Eq. (6.92), for $\nu_c = 2$ BBKS peaks from the initial conditions $\tau_i = 0$ ($z = \infty$) (bottom dotted curve) until halo collapse at $\tau = \tau_0$. The latter is assumed to take place at $z_0 \equiv z(\tau_0) = 0.3$ (top solid curve). For illustration purposes, $\xi_{\rm pk}(r)$ is also shown at the intermediate redshift values $z = 5$, 2, 1 and 0.5. For comparison, the bottom and top dashed curves represent the initial and final correlation

161

function in the LIMD approximation, in which all the peak bias parameters are set to zero except for b_{N0}^L, the velocity bias has been turned off so that $\sigma_{v,\mathrm{pk}} = \sigma_{v,\mathrm{dm}}$, and filtering is absent (cf. Fig. 4 in Sec. 2.1). As the redshift decreases, gravitational instability generates coherent motions which amplify the large-scale amplitude of the peak correlation function, together with random motions which increasingly smear out the initial BAO feature. Although the randomness in the large-scale flows is less important for the peaks than for the LIMD-biased tracers (owing to the velocity bias), the final correlation function of peaks is noticeably more similar to that of the LIMD-restricted tracer than it was initially. Still, mild differences remain at $z = z_0$ between the peak and LIMD predictions, especially around the BAO feature. This residual higher-derivative bias, which strongly depends on the halo mass, is a subtle combination of the higher-derivative terms in the Lagrangian bias functions $c_n^L(\boldsymbol{k}_1, \ldots, \boldsymbol{k}_n)$ and of the velocity bias.

7 Bias and primordial non-Gaussianity

So far, we have made the assumption of adiabatic Gaussian initial perturbations, which forms a key prediction of the inflationary paradigm [521]. Moreover, the cosmic microwave background has placed stringent upper limits on deviations from Gaussianity in the initial conditions, known under the term *primordial non-Gaussianity (PNG)*. Nevertheless, there is still significant unconstrained parameter space for PNG which offers rich insights into the physics of inflation (see Sec. 7.2, and [522, 523, 524] for reviews). This provides strong motivation to study the impact of PNG on large-scale structure.

The connection between the clustering of galaxies on large scales and the statistics of the initial conditions has been studied since the 1980s, including the seminal papers [525, 526, 422, 527, 528, 529, 530] (see also [70] for an overview). In the 2000s, it was realized that non-Gaussianity in the initial conditions is likely to be only a small correction to the non-Gaussianity induced by nonlinear gravitational evolution [531]. Refs. [67, 62] then developed techniques to disentangle primordial and late-time non-Gaussianities in the bispectrum of the large-scale structure. Finally, this field experienced another breakthrough with the discovery of the strong effect of PNG of the local type (*local PNG*) on the clustering of galaxies on large scales [68] (although this effect was already implicitly contained in the results of [525, 526]). We now know that galaxy clustering can provide independent constraints on the magnitude of local PNG that are competitive with those from the CMB and, in the long run, may even give the best constraints (Sec. 7.6). Moreover, constraints from galaxy bias are based on small-scale modes $k \gtrsim 0.3\,h\,\mathrm{Mpc}^{-1}$ (of order of the formation scale R_*^{-1} of tracers), while the CMB probes modes with $k \lesssim 0.1\,h\,\mathrm{Mpc}^{-1}$. Given the possibility of a scale dependence in the PNG, as predicted from several models of the early Universe, the two should therefore be seen as complementary probes of PNG. The effect of PNG on large-scale structure has recently been reviewed in [428, 523, 532]. Here, we significantly expand their discussion of bias in the presence of PNG, and include results derived in the past several years.

We begin in Sec. 7.1 by describing the effect of PNG in the framework of the general bias expansion described in Sec. 2, starting with the simplest type, local PNG, in Sec. 7.1.1–7.1.2. For clarity, we refer to the additional terms introduced in the bias expansion in the presence of PNG as *non-Gaussian bias*. Local PNG exemplifies the essential physics of the effect of PNG on galaxy clustering, without the complications of more general non-Gaussian initial conditions, which are then considered in Sec. 7.1.3–7.1.4. The discussion of additional terms appearing on smaller scales (Sec. 7.1.5) concludes the treatment of PNG in this context. Sec. 7.1.6 provides a brief summary of the contributions to the general bias expansion at leading order in the amplitude of PNG, complementing the summary in Sec. 2.11, and also gives the contributions to the two-

and three-point functions of galaxies in their rest frame, extending the results given in Sec. 4.1.1. We briefly discuss the implications for early-Universe physics of detecting and constraining the various signatures of PNG in the clustering of galaxies in Sec. 7.2. In order to obtain quantitative constraints on PNG however, we need predictions for the new bias coefficients that appear in the presence of PNG. The PBS argument of Sec. 3 can be used to derive these, as described in Sec. 7.3. We then show in Sec. 7.4 how, for any given "microscopic" Lagrangian model of tracers, the effect of PNG on the n-point functions in Lagrangian space can be fully predicted. This includes the thresholding (Sec. 2.1), excursion set (Sec. 5) and peak models (Sec. 6), which are considered in turn in Sec. 7.4.1–7.4.3.

Finally, the comparison of the theoretical predictions of halo bias in the presence of PNG with N-body simulations is discussed in Sec. 7.5. We present current and forecasted observational constraints on local PNG using galaxy clustering in Sec. 7.6, generalizing the forecast in Sec. 4.1.3.

7.1 Primordial non-Gaussianity in the general bias expansion

In this section, we describe in detail how the general bias expansion described in Sec. 2.5–2.8 can be extended to include non-Gaussian initial conditions. For this, we will restrict to the leading contributions in the limit of weak PNG, which are described by the three- and four-point functions of the initial conditions. As we will see, these are phenomenologically by far the most important contributions given current constraints on large-scale PNG.

7.1.1 Primordial non-Gaussianity of the local type

PNG deals with the statistical properties of the initial conditions for structure formation. These are usually phrased in terms of the primordial Bardeen potential ϕ [533], which is, in turn, directly related to the curvature perturbation in comoving gauge $\mathcal{R}$ by $\phi = (3/5)\mathcal{R}$ for modes that enter the horizon during the matter-dominated epoch [534]. The Newtonian potential $\Phi(\mathbf{k})$ is related to $\phi(\mathbf{k})$ during matter domination by the transfer function $T(k)$ which satisfies $T(k) = 1$ for $k \ll k_{\rm eq}$, where $k_{\rm eq} \simeq 0.02\,h\,{\rm Mpc}^{-1}$ is the wavenumber that entered the horizon at matter-radiation equality. Thus, ϕ is related to the linear density field by

$$\delta^{(1)}(\mathbf{k}, \tau) = \mathcal{M}(k, \tau)\phi(\mathbf{k}) \quad \text{where} \quad \mathcal{M}(k, \tau) = \frac{2}{3}\frac{k^2 T(k) D_{\rm md}(\tau)}{\Omega_{m0} H_0^2}\,. \tag{7.1}$$

Here, $D_{\rm md}(\tau)$ is the linear growth factor normalized to $a(\tau)$ during the matter-dominated epoch. To be precise, following our discussion in Sec. 2.9, $\delta^{(1)}$ defined in Eq. (7.1) is the matter density contrast in synchronous-comoving gauge. In the following, we will drop the explicit time dependence on δ and $\mathcal{M}$, as well as the bias parameters, but include it again in Sec. 7.1.6.

In all previous sections, we have assumed that $\phi = \phi_G$ is a Gaussian random field, which is completely described by its power spectrum $P_\phi(k)$. The simplest way to generate non-Gaussian initial conditions is to perform a local, nonlinear transformation of such a Gaussian field ϕ_G, $\phi(\mathbf{x}) = f(\phi_G(\mathbf{x}))$. Since the root-mean-squared value of ϕ is less than 10^{-4}, a Taylor expansion converges rapidly, and so we only keep the leading nonlinear term. This leads to

$$\phi(\mathbf{x}) = \phi_G(\mathbf{x}) + f_{\rm NL}\left[\phi_G^2(\mathbf{x}) - \langle\phi_G^2\rangle\right] + \mathcal{O}(\phi_G^3)\,. \tag{7.2}$$

This is the definition of *local quadratic PNG*, parametrized by the parameter $f_{\rm NL}$ as first introduced in [535]. The term $-f_{\rm NL}\langle\phi_G^2\rangle$ ensures that $\langle\phi\rangle = 0$.

At leading order in $f_{\rm NL}$, the only poly-spectrum beyond the power spectrum is the bispectrum,

$$B_\phi(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = 2f_{\rm NL}[P_\phi(k_1)P_\phi(k_2) + 2 \text{ perm.}]\,. \tag{7.3}$$

Note that if we want to derive the N-point functions of ϕ at order $f_{\rm NL}^2$, we also have to include the next term $g_{\rm NL}\phi^3$ appearing in the expansion Eq. (7.2), as both are of the same perturbative order. From Eq. (7.1), we immediately obtain the leading *primordial contribution* to the matter bispectrum,

$$B^{(1)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) \equiv \mathcal{M}(k_1)\mathcal{M}(k_2)\mathcal{M}(k_3)B_\phi(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)\,. \tag{7.4}$$

That is, this is the synchronous-comoving-gauge matter bispectrum at early times $\tau \to 0$. A particularly important regime for the effects of PNG on galaxy clustering is the *squeezed limit*, where two wavenumbers are much larger than the third. In this limit, the matter bispectrum Eq. (7.4) becomes

$$B^{(1)}(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3) \overset{k_1 \simeq k_2 \gg k_3}{=} 4f_{\mathrm{NL}}\mathcal{M}(k_3)P_\phi(k_3)P_{\mathrm{L}}(k_S) + \mathcal{O}\left(\frac{k_3^2}{k_S^2}\right)$$

$$= 4f_{\mathrm{NL}}\mathcal{M}^{-1}(k_3)P_{\mathrm{L}}(k_3)P_{\mathrm{L}}(k_S) + \mathcal{O}\left(\frac{k_3^2}{k_S^2}\right), \quad \text{where} \quad \boldsymbol{k}_S \equiv \boldsymbol{k}_1 + \frac{1}{2}\boldsymbol{k}_3. \quad (7.5)$$

In the second line, we have used the relation between $\phi(\boldsymbol{k})$ and $\delta^{(1)}(\boldsymbol{k})$, Eq. (7.1).

7.1.2   General bias expansion with local PNG

Let us go back to the expression for the galaxy-matter cross-power spectrum in the large-scale limit $(k \to 0)$, Eq. (4.2) or Eq. (2.121) in Sec. 2.10:

$$\langle \delta_g(\boldsymbol{k})\delta(\boldsymbol{k}') \rangle \Big|_{\mathrm{LO}} \overset{\mathrm{Gauss}}{=} b_1 \langle \delta^{(1)}(\boldsymbol{k})\delta^{(1)}(\boldsymbol{k}') \rangle. \quad (7.6)$$

In the case of Gaussian initial conditions, the NLO correction to this expression (or 1-loop power spectrum) derived in Sec. 4.1.4 involves the second- and third-order density field induced by gravitational evolution as well as nonlinear bias contributions. The NLO contribution is suppressed on large scales by $(k/k_{\mathrm{NL}})^{3+n}$, where n is the effective power spectrum index and $k_{\mathrm{NL}}(\tau) \gtrsim 0.2\,h\,\mathrm{Mpc}^{-1}$ is the nonlinear scale [Eq. (4.25) in Sec. 4.1.4; higher-derivative corrections are similarly suppressed by $(kR_*)^2$].

Now let us consider the leading correction to the galaxy-matter cross-power spectrum induced by local-type PNG, which corresponds to the following two contributions:

$$\langle \delta_g(\boldsymbol{k})\delta(\boldsymbol{k}') \rangle \Big|_{\mathrm{PNG}} = \frac{1}{2}b_2 \left\langle \left(\delta^{(1)}\right)^2(\boldsymbol{k})\delta^{(1)}(\boldsymbol{k}') \right\rangle + b_{K^2} \left\langle \left(K_{ij}^{(1)}\right)^2(\boldsymbol{k})\delta^{(1)}(\boldsymbol{k}') \right\rangle, \quad (7.7)$$

since now the bispectrum of the linear (initial) density field $\delta^{(1)}$ no longer vanishes. The notation $(\delta^{(1)})^2(\boldsymbol{k})$ denotes the Fourier transform of the square of the density field, which is given by a convolution in Fourier space, and analogously for the tidal field $(K_{ij}^{(1)})^2(\boldsymbol{k})$. Now, for reasons that will become clear shortly, we re-introduce the artificial smoothing of the density field appearing in the bias operators on the scale $R = \Lambda^{-1}$, which was employed in Sec. 2.10 to make the distinction between large (perturbative, or background) and small (non-perturbative, or peak) scales rigorous. In the end, all observables should be independent of Λ, and we can take the limit $\Lambda^{-1} \to 0$. Correspondingly, we denote the "bare" bias coefficients with c_O. Eq. (7.7) becomes

$$\langle \delta_g(\boldsymbol{k})\delta(\boldsymbol{k}') \rangle \Big|_{\mathrm{PNG}} = \frac{1}{2}c_{2,\Lambda} \left\langle \left(\delta_\Lambda^{(1)}\right)^2(\boldsymbol{k})\,\delta^{(1)}(\boldsymbol{k}') \right\rangle + c_{K^2,\Lambda} \left\langle \left(K_{ij,\Lambda}^{(1)}\right)^2(\boldsymbol{k})\,\delta^{(1)}(\boldsymbol{k}') \right\rangle, \quad (7.8)$$

We now evaluate the first contribution using Eq. (7.4), yielding [536, 100]

$$\left\langle \left(\delta_\Lambda^{(1)}\right)^2(\boldsymbol{k})\,\delta^{(1)}(\boldsymbol{k}') \right\rangle' = \int_{\boldsymbol{p}} B^{(1)}(\boldsymbol{p}, \boldsymbol{k} - \boldsymbol{p}, \boldsymbol{k}')W_\Lambda(p)W_\Lambda(|\boldsymbol{k} - \boldsymbol{p}|)$$

$$\approx 4f_{\mathrm{NL}} \int_{\boldsymbol{p}} |W_\Lambda(p)|^2 P_{\mathrm{L}}(p)\,\mathcal{M}^{-1}(k)P_{\mathrm{L}}(k) = 4f_{\mathrm{NL}}\sigma^2(\Lambda)\,\langle \phi(\boldsymbol{k})\delta^{(1)}(\boldsymbol{k}') \rangle'. \quad (7.9)$$

In the second approximate equality, we have assumed that $k \ll p \sim \Lambda$, where the integral peaks, so that $B^{(1)}$ is evaluated in the squeezed limit Eq. (7.5). Further, $\sigma^2(\Lambda)$ is the variance of the density field smoothed on the cutoff scale $R = \Lambda^{-1}$. We will discuss the sub-leading terms in k/Λ in Sec. 7.1.5, which are similar to the higher-derivative bias contributions discussed in Sec. 2.6; physically, the squeezed limit is a good assumption as long as the scales on which galaxy statistics are measured are much larger than the scales over which

galaxies form. The second term in Eq. (7.8), $\langle (K_{ij,\Lambda}^{(1)})^2(\boldsymbol{k})\delta^{(1)}(\boldsymbol{k}')\rangle$, precisely yields Eq. (7.9) multiplied by a factor of 2/3.

Perhaps surprisingly, we see that $(\delta_\Lambda)^2$ and $(K_{ij,\Lambda})^2$ need to be renormalized at *leading order* in perturbation theory in the presence of local PNG, since their correlation with $\delta^{(1)}$ on large scales is strongly cutoff-dependent, as it is proportional to $\sigma^2(\Lambda)$. Moreover, the contribution in Eq. (7.9) is not small on large scales. Rather, it becomes even larger than the Gaussian leading-order contribution in Eq. (7.6) for sufficiently small k; note that, on large scales $k < k_{\rm eq} \simeq 0.02\,h\,{\rm Mpc}^{-1}$, $\mathcal{M}^{-1}(k) \propto (k/H_0)^{-2}$. This result is clearly not satisfactory, but can be cured by adding counter-terms to δ^2, K^2 which absorb this unphysical contribution and yield renormalized operators:

$$[\delta^2](\boldsymbol{k}) = (\delta_\Lambda^2)(\boldsymbol{k}) - 4f_{\rm NL}\sigma^2(\Lambda)\phi(\boldsymbol{k}) + [\text{Gaussian counter-terms}] \tag{7.10}$$

$$[K^2](\boldsymbol{k}) = (K_\Lambda^2)(\boldsymbol{k}) - \frac{8}{3}f_{\rm NL}\sigma^2(\Lambda)\phi(\boldsymbol{k}) + [\text{Gaussian counter-terms}]\,. \tag{7.11}$$

What is crucial about these counter-terms induced by PNG is that they both involve the Bardeen potential ϕ directly, without any derivatives. Such a term cannot be generated by gravitational evolution, as we have seen in Sec. 2.5, since the gravitational potential is not locally observable; only density and tidal effects proportional to second spatial derivatives of ϕ are.

In the presence of local PNG, therefore, we need to add $f_{\rm NL}\phi$ to the list of operators that appear in the bias expansion [537, 538],

$$\delta_g\Big|_{\rm PNG} \supset b_\phi\, f_{\rm NL}[\phi]\,. \tag{7.12}$$

While $[\phi]$ is strictly a renormalized operator, it does not receive any counter-terms at leading order in $f_{\rm NL}$. Adding this new *linear-order* operator to the bias expansion leads to a leading-order large-scale galaxy-matter cross-power spectrum given by

$$P_{gm}^{\rm LO}(k) \equiv \langle \delta_g(\boldsymbol{k})\delta(\boldsymbol{k}')\rangle'_{\rm LO} = b_1\langle \delta^{(1)}(\boldsymbol{k})\delta^{(1)}(\boldsymbol{k}')\rangle' + b_\phi f_{\rm NL}\langle \phi(\boldsymbol{k})\delta^{(1)}(\boldsymbol{k}')\rangle'$$
$$= \left[b_1 + b_\phi f_{\rm NL}\mathcal{M}^{-1}(k)\right] P_{\rm L}(k)\,, \tag{7.13}$$

which now includes the leading terms on large scales in a proper cutoff-independent way. Similarly, we find for the leading-order galaxy auto-power spectrum in the presence of local PNG

$$P_{gg}^{\rm LO}(k) = \left[b_1 + b_\phi f_{\rm NL}\mathcal{M}^{-1}(k)\right]^2 P_{\rm L}(k) + P_\varepsilon^{\{0\}}\,, \tag{7.14}$$

where $P_\varepsilon^{\{0\}} \equiv \lim_{k\to 0}\langle \varepsilon(\boldsymbol{k})\varepsilon(\boldsymbol{k}')\rangle'$ is the leading-order stochastic contribution which is scale-independent on large scales and thus of the same form as in the Gaussian case. This shows that PNG of the local type leads to a *scale-dependent bias* $\propto k^{-2}$ in the large-scale two-point function of tracers [68, 536] (see Fig. 32 on p. 186). This is a unique smoking-gun signal for PNG, as neither any local process nor nonlinear gravitational evolution can generate such a scale dependence, as we have explained above. The term proportional to $(c_{2,\Lambda}/2)\langle(\delta_\Lambda^2)(\boldsymbol{k})\delta_\Lambda(\boldsymbol{k}')\rangle$ in Eq. (7.8) on the other hand must be seen as an artifact of the bare bias expansion which is absorbed in the renormalized operator $[\delta^2]$. Correspondingly, the new bias parameter b_ϕ has nothing to do with b_2 a priori. We will show how b_ϕ can be derived using a generalized PBS argument in Sec. 7.3.

The physical interpretation of Eqs. (7.13)–(7.14) is the following. If the initial potential fluctuations have a bispectrum given by Eq. (7.3), then long-wavelength potential perturbations modulate the initial amplitude of small-scale perturbations. Consider a long-wavelength mode $\phi(\boldsymbol{k}_\ell)$, and a patch much smaller than the wavelength of this mode. If we measure the local power spectrum of small-scale density perturbations within such a patch at position $\boldsymbol{x}$, denoted as $P_{\rm L}(\boldsymbol{k}_S|\boldsymbol{x})$, then Eq. (7.5) states that this is modulated by

$$P_{\rm L}(\boldsymbol{k}_S|\boldsymbol{x}) = \left[1 + 4f_{\rm NL}\phi(\boldsymbol{k}_\ell)e^{i\boldsymbol{k}_\ell\cdot\boldsymbol{x}}\right] P_{\rm L}(k_S)\,, \tag{7.15}$$

where $P_{\rm L}(k)$ is the matter power spectrum averaged over the entire initial conditions. To see this, simply multiply this relation by $\phi(\boldsymbol{k}'_\ell)e^{i\boldsymbol{k}'_\ell\cdot\boldsymbol{x}}$, take the expectation value, and integrate over $d^3\boldsymbol{x}$. This recovers

the initial matter bispectrum in the squeezed limit, Eq. (7.5). We thus see that the amplitude of small-scale perturbations in the initial conditions is modulated by long-wavelength potential perturbations via $4f_{\mathrm{NL}}\phi(\boldsymbol{k}_\ell)$. Since the abundance of halos and galaxies is sensitive to the amplitude of initial fluctuations, their abundance is correspondingly modulated by $4f_{\mathrm{NL}}\phi(\boldsymbol{k}_\ell)$ [539], leading to the scale-dependent bias in Eq. (7.13). In Sec. 2.5, we were able to absorb the amplitude of small-scale perturbations completely through stochastic contributions such as $P_\varepsilon^{\{0\}}$ since, in the Gaussian case, the small-scale initial perturbations are statistically the same everywhere. This changes in the non-Gaussian case, and the relevant quantity that determines the effect on bias is the squeezed-limit bispectrum, as we have seen.

The term $b_\phi\phi(\boldsymbol{k})$ is the leading PNG contribution to bias, but it is not the only one. The non-Gaussianity present in the initial conditions couples to gravitational evolution, and modifies the growth of matter perturbations as well (see Fig. 6 on p. 34). Changing the statistics of the small-scale initial conditions at $\boldsymbol{q} = \boldsymbol{x}_{\mathrm{fl}}(0)$ will modify the subsequent evolution along the fluid trajectory $\boldsymbol{x}_{\mathrm{fl}}(\tau)$ [267, 540]. At lowest order in derivatives and linear order in f_{NL}, the bias operators we need to include consist of all combinations of $f_{\mathrm{NL}}\phi$ with the Gaussian operators we have listed in Sec. 2.5. A crucial point here is that $\phi = \phi(\boldsymbol{x}_{\mathrm{fl}}(0)) = \phi(\boldsymbol{q})$ is to be evaluated at the *Lagrangian* position $\boldsymbol{q}$ that corresponds to the Eulerian position $(\boldsymbol{x}, \tau)$ [538, 541],

$$\phi(\boldsymbol{q}) = \phi(\boldsymbol{x}) - s^i(\boldsymbol{x}, \tau)\partial_i\phi(\boldsymbol{x}) + \cdots, \tag{7.16}$$

where we have expanded to second order in perturbations. This is because the coupling is present in the initial conditions, and not induced by evolution. In fact, this holds for any imprints present in the initial conditions, including isocurvature modes between baryons and CDM (see Sec. 8.2). In summary, working in Eulerian frame, the basis of operators introduced in Sec. 2.5.3, Eq. (2.64), needs to be augmented by the following additional terms up to third order in case of local-type PNG:

$$
\begin{aligned}
&1^{\mathrm{st}} && f_{\mathrm{NL}}\phi(\boldsymbol{q}) && (7.17)\\
&2^{\mathrm{nd}} && f_{\mathrm{NL}}\mathrm{Tr}[\Pi^{[1]}(\boldsymbol{x})]\phi(\boldsymbol{q}) &&\\
&3^{\mathrm{rd}} && f_{\mathrm{NL}}\mathrm{Tr}[(\Pi^{[1]}(\boldsymbol{x}))^2]\phi(\boldsymbol{q}),\ f_{\mathrm{NL}}(\mathrm{Tr}[\Pi^{[1]}(\boldsymbol{x})])^2\phi(\boldsymbol{q}),
\end{aligned}
$$

where $\Pi^{[1]}_{ij} = K_{ij} + (\delta_{ij}/3)\delta \propto \partial_i\partial_j\Phi$. We have emphasized the distinction between Eulerian and Lagrangian arguments. The continuation to higher orders in perturbation theory is now obvious. When deriving the statistics of galaxies at a given order in perturbation theory, one also expands $\phi(\boldsymbol{q})$ following Eq. (7.16) as we will see in Sec. 7.1.6. This however does not lead to additional bias parameters, as the amplitude of the displacement terms is controlled by the corresponding bias parameter (for example, b_ϕ). The fully Lagrangian basis can be analogously constructed out of $\phi(\boldsymbol{q})$ and Eq. (2.61).

Eq. (7.17) only gives the leading non-Gaussian contribution $\mathcal{O}(f_{\mathrm{NL}})$. At $\mathcal{O}(f_{\mathrm{NL}}^2)$, one needs to add $\phi^2(\boldsymbol{q})$ to the list, and, following our discussion after Eq. (7.3), also needs to keep cubic non-Gaussian terms $\propto g_{\mathrm{NL}}$. However, all these terms have an extremely small observable effect on LSS statistics [540]. For example, one would have to measure the bispectrum of tracers with all three k modes of order aH, which is unlikely to yield a significant detection in the foreseeable future, unless the cubic non-Gaussianity amplitude is very large, $g_{\mathrm{NL}} \gtrsim 10^4$. For this reason, we restrict to linear order in f_{NL} in the bias expansion throughout. Of course, even this will generate terms of order f_{NL}^2 and f_{NL}^3 in the tracer power spectrum and bispectrum, respectively. These terms should be kept as they become important if a *single* k mode becomes of order aH.

The bias terms listed in Eq. (7.17) only apply to the case of PNG of the local type. However, the generalization to other forms of the squeezed-limit bispectrum is straightforward, and only amounts to replacing $\phi(\boldsymbol{q})$ with a nonlocal transformation of the potential. We will describe this next.

7.1.3 Beyond local-type PNG

We now consider the generalization of the results of the last section to arbitrary quadratic non-Gaussianity, that is, PNG described to leading order by a three-point function or bispectrum. The impact of an arbitrary three-point function on the two-point function of galaxies was first considered in [542], who used the thresholding model described in Sec. 2.1 (based on [525]; see Sec. 7.4.1). Ref. [543] explored the effects of

multi-field inflation on bias (we will discuss this in Sec. 7.1.4). The leading scale-dependent bias induced by PNG on large scales for a generic primordial bispectrum was derived by [544], who generalized the physical argument of [539], while [545] derived the same in the high-peak limit, and [495] adopted the iPT approach (Sec. 6.6.2). Finally, [98, 546, 547] identified a missing contribution to the scale-dependent bias from PNG when the PNG is not of the local type; we will return to this issue in Sec. 7.3.

Our treatment here continues to follow the philosophy of the general perturbative bias expansion (Sec. 2.5) and is most closely related to that in [540], which in turn had precursors in [537, 100]. The only assumptions made in this treatment are (i) gravity is described by General Relativity, and (ii) the initial conditions are given by weakly non-Gaussian adiabatic perturbations ϕ. No assumptions are made regarding universality of the galaxy or halo mass function, or the high-peak limit.

General quadratic PNG is often parametrized by a generalization of the local expansion Eq. (7.2) [544, 546],

$$\phi(\boldsymbol{k}) = \phi_G(\boldsymbol{k}) + \int_{\boldsymbol{k}_1} \int_{\boldsymbol{k}_2} K_{\mathrm{NL}}(\boldsymbol{k}_1, \boldsymbol{k}_2) \phi_G(\boldsymbol{k}_1) \phi_G(\boldsymbol{k}_2) (2\pi)^3 \delta_D(\boldsymbol{k} - \boldsymbol{k}_{12}) \,, \tag{7.18}$$

where $\phi_G(\boldsymbol{k})$ is a Gaussian field, and $K_{\mathrm{NL}}(\boldsymbol{k}_1, \boldsymbol{k}_2)$ is a kernel. The kernel K_{NL} is not uniquely determined by the bispectrum of ϕ [544, 546], which raises the worry that a prediction for the non-Gaussian bias based only on B_ϕ is ambiguous. Fortunately, this is not the case. Instead of adopting the expansion Eq. (7.18) to derive the effects of PNG on galaxy bias, we will build the bias expansion upon the n-point statistics of ϕ directly.

As discussed in Sec. 7.1.2, since we are interested in the clustering of galaxies on scales much larger than the scales ($\sim R_*$) on which galaxy formation happens, the kinematic regime of the three-point function that is relevant for bias is the *squeezed limit* where one mode $\boldsymbol{k}_3 = \boldsymbol{k}_\ell$ is much smaller than the other two modes $\boldsymbol{k}_1, \boldsymbol{k}_2$. The bispectrum can then be written as

$$\begin{aligned} B_\phi(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_\ell) &= A(\boldsymbol{k}_S, \boldsymbol{k}_\ell) P_\phi(k_\ell) P_\phi(k_S) + \mathcal{O}\left(\frac{k_\ell^2}{k_S^2}\right) \\ &= \sum_{J=0,2,4,\dots} A_J(k_S, k_\ell) \mathcal{L}_J(\hat{\boldsymbol{k}}_S \cdot \hat{\boldsymbol{k}}_\ell) P_\phi(k_\ell) P_\phi(k_S) + \mathcal{O}\left(\frac{k_\ell^2}{k_S^2}\right) \,, \end{aligned} \tag{7.19}$$

where $\boldsymbol{k}_S \equiv \boldsymbol{k}_1 + \boldsymbol{k}_\ell/2 = -\boldsymbol{k}_2 - \boldsymbol{k}_\ell/2$ as above and $\mathcal{L}_J$ are the Legendre polynomials. In the squeezed limit, only even multipoles can contribute to the bispectrum for symmetry reasons. This can be seen by considering the local small-scale power spectrum of ϕ, $P_\phi(\boldsymbol{k}_S|\boldsymbol{x})$, analogously to Eq. (7.15): since this is the power spectrum of a real scalar field, we have to have $P_\phi(\boldsymbol{k}_S|\boldsymbol{x}) = P_\phi(-\boldsymbol{k}_S|\boldsymbol{x})$. This requires the absence of odd multipole contributions in Eq. (7.19) [548, 540]. In the following, we will assume that B_ϕ is scale-invariant and write

$$A_J(k_S, k_\ell) = 4 a_J (k_\ell/k_S)^\alpha \,, \tag{7.20}$$

where a_J and α are dimensionless constants. General scale-invariant bispectrum shapes can be approximated by considering a sum of several contributions $\{a_{J,i}, \alpha_i\}$ [549]; note that α can be a real number. The local form Eq. (7.3) is a special case of Eq. (7.19) with $J = 0$, $\alpha = 0$ and $a_0 = f_{\mathrm{NL}}$. The factor of 4 in Eq. (7.20) is motivated by the kernel expansion Eq. (7.18); the dimensionless coefficients a_J quantify the quadratic coupling strength of the field ϕ.

Let us consider the case $J = 0$ first, and look again at the galaxy-matter cross-power spectrum before renormalization. As in the case of local PNG, we need to deal with the term $(1/2)c_{2,\Lambda}\delta_\Lambda^2(\boldsymbol{x})$ in the bare bias expansion, which now becomes [544], in analogy with Eq. (7.9),

$$\left\langle \left(\delta_\Lambda^{(1)}\right)^2 (\boldsymbol{k}) \delta^{(1)}(\boldsymbol{k}') \right\rangle = 4 a_0 \sigma_{-\alpha/2}^2(\Lambda) \, k^\alpha \langle \phi_\Lambda(\boldsymbol{k}) \delta^{(1)}(\boldsymbol{k}') \rangle \,, \tag{7.21}$$

where the spectral moment $\sigma_{-\alpha/2}(\Lambda)$ is defined in Tab. 4 [see also Eq. (6.1); note that this definition is valid for non-integer n]. Again, we find a cutoff-dependence that needs to be renormalized. In this case, due to

the factor k^α in Eq. (7.21), we have to add a different counter-term to obtain the renormalized operator $[\delta^2]$,

$$[\delta^2](\boldsymbol{k}) = (\delta_\Lambda^2)(\boldsymbol{k}) - 4a_0\sigma_{-\alpha/2}^2(\Lambda)k^\alpha\phi(\boldsymbol{k}) + [\text{Gaussian counter-terms}]. \qquad (7.22)$$

The same again holds for the renormalized operator $[K^2]$, up to a factor of $2/3$. The reasoning of Sec. 7.1.2 now goes through in exactly the same way. We have to introduce an additional bias operator defined by

$$\psi(\boldsymbol{q}) \equiv \int_{\boldsymbol{k}} k^\alpha\phi(\boldsymbol{k})e^{i\boldsymbol{k}\cdot\boldsymbol{q}}. \qquad (7.23)$$

The Eulerian basis of bias operators in general PNG with $J = 0$ is then directly obtained from Eq. (7.17) by replacing $f_{\text{NL}}\phi \to a_0\psi$, yielding

$$J = 0 : \qquad 1^{\text{st}} \qquad a_0\psi(\boldsymbol{q}) \qquad\qquad (7.24)$$
$$2^{\text{nd}} \qquad a_0\text{Tr}[\Pi^{[1]}(\boldsymbol{x})]\psi(\boldsymbol{q})$$
$$3^{\text{rd}} \qquad a_0\text{Tr}[(\Pi^{[1]}(\boldsymbol{x}))^2]\psi(\boldsymbol{q}), \ a_0(\text{Tr}[\Pi^{[1]}(\boldsymbol{x})])^2\psi(\boldsymbol{q}).$$

The physical reason behind the new operator $\psi(\boldsymbol{q})$ is the same as for local PNG: Eq. (7.19) for $J = 0$ states that the local initial matter power spectrum is rescaled by a single long-wavelength mode $\psi(\boldsymbol{k}_\ell)$ via

$$P_{\text{L}}(\boldsymbol{k}_S|\boldsymbol{x}) = \left[1 + 4a_0\psi(\boldsymbol{k}_\ell)e^{i\boldsymbol{k}_\ell\cdot\boldsymbol{x}}k_S^{-\alpha}\right]P_{\text{L}}(k_S). \qquad (7.25)$$

This modulation is mediated by $\psi(\boldsymbol{k}_\ell)$, which now appears in the bias expansion. From Eq. (7.24), the leading-order cross- and auto-correlations of galaxies are then given by

$$P_{gm}^{\text{LO}}(k) = \left[b_1 + b_\psi a_0 k^\alpha \mathcal{M}^{-1}(k)\right]P_{\text{L}}(k)$$
$$P_{gg}^{\text{LO}}(k) = \left[b_1 + b_\psi a_0 k^\alpha \mathcal{M}^{-1}(k)\right]^2 P_{\text{L}}(k) + P_\varepsilon^{\{0\}}. \qquad (7.26)$$

We see that, on large scales, there is a scale-dependent bias $\propto k^{-2+\alpha}$. For $\alpha = 0$, we clearly recover the case of local-type PNG. For $\alpha = 2$, the non-Gaussian contribution to the galaxy two-point function is scale-independent on large scales and thus degenerate with the Gaussian term (linear bias b_1). This is not surprising, since for $\alpha = 2$, $\psi \propto \nabla^2\phi$, which on large scales is simply proportional to the matter density. This applies to the equilateral and orthogonal shapes of PNG generated during single-field inflation, and raises the question of whether these shapes can be constrained using the scale-dependent bias. A scale-dependence in the bias $\propto T^{-1}(k)$ does arise due to the transfer function on smaller scales $k \gtrsim k_{\text{eq}}$. On those scales, the transfer function is no longer constant. However, for adiabatic perturbations, we can expand the inverse of the transfer function as $T^{-1}(k) = 1 + t_1(k/k_{\text{eq}})^2 + t_2(k/k_{\text{eq}})^4 + \cdots$, where t_i are factors of order unity. This correspondingly leads to a scale-dependent bias of

$$b_\psi f_{\text{NL}}^{\alpha=2}k^2\mathcal{M}^{-1}(k) \sim f_{\text{NL}}^{\alpha=2}R_*^2 H_0^2 T^{-1}(k) \sim \text{const} + f_{\text{NL}}^{\alpha=2}R_*^2 H_0^2\left(\frac{k^2}{k_{\text{eq}}^2} + \frac{k^4}{k_{\text{eq}}^4} + \cdots\right), \qquad (7.27)$$

where we have introduced $f_{\text{NL}}^{\alpha=2} = a_0(\alpha = 2)$ and assumed that $b_\psi \sim R_*^2$ is controlled by the characteristic scale of the tracer R_*, the Lagrangian radius in case of halos (this will be justified in Sec. 7.3). These terms scale in the same way with k, but are much smaller than the Gaussian higher-derivative bias $b_{\nabla^2\delta}\nabla^2\delta, \ldots$ (Sec. 2.6) unless $f_{\text{NL}}^{\alpha=2} \gtrsim 10^3$ [540]. However, this degeneracy can be broken if the scale dependence of the bias at $k \gtrsim k_{\text{eq}}$ can be measured with sufficient precision, so that the different scales R_* and k_{eq}^{-1} involved in the higher-derivative and PNG contributions can be disentangled. Ref. [550] show that constraints on $f_{\text{NL}}^{\alpha=2}$ from the scale-dependent bias can still be obtained; however, the partial degeneracy with the higher-derivative and NLO (1-loop) contributions to the galaxy power spectrum increases the uncertainty on $f_{\text{NL}}^{\alpha=2}$ by a factor of ~ 40 over the case where all higher-derivative and nonlinear bias contributions are absent.

Let us now briefly consider the case $J = 2$, i.e. a primordial bispectrum that is anisotropic in the squeezed limit [551]. Such a form can, among other non-standard physics, indicate the presence of a massive spin-2

(tensor) field during inflation (see Sec. 7.2). In this case, $[\delta^2]$ receives a somewhat different counter-term [540]

$$[\delta^2](\boldsymbol{x}) = \delta^2(\boldsymbol{x}) - \frac{16}{105} a_2 \sigma^2_{-\alpha/2}(\Lambda) \psi_{ij}(\boldsymbol{q}) K^{ij}(\boldsymbol{x}) \,, \tag{7.28}$$

where $\psi_{ij}(\boldsymbol{q})$ is a traceless tensor field derived from the Bardeen potential via

$$\psi_{ij}(\boldsymbol{q}) \equiv \int_{\boldsymbol{k}} \frac{3}{2} \left(\frac{k^i k^j}{k^2} - \frac{1}{3} \delta^{ij} \right) k^\alpha \phi(\boldsymbol{k}) e^{i \boldsymbol{k} \cdot \boldsymbol{q}} \,. \tag{7.29}$$

This type of PNG corresponds to an *anisotropic* modulation of the local small-scale power spectrum by long-wavelength potential perturbations via [552]

$$P_{\rm L}(\boldsymbol{k}_S | \boldsymbol{x}) = \left[1 + 4 a_2 \psi_{ij}(\boldsymbol{k}_\ell) \frac{k_S^i k_S^j}{k_S^{2+\alpha}} e^{i \boldsymbol{k}_\ell \cdot \boldsymbol{x}} \right] P_{\rm L}(k_S) \,. \tag{7.30}$$

If the small-scale matter power spectrum can be measured directly, this effect can also be used to search for spin-2 fields by constructing an estimator for ψ_{ij} based on the local power spectrum $P_{\rm L}(\boldsymbol{k}_S | \boldsymbol{x})$ [553].

Up to cubic order, the following terms then enter the bias expansion for PNG with $J = 2$:

$$J = 2 : \quad 1^{\rm st} \quad\quad - \tag{7.31}$$

$$2^{\rm nd} \quad\quad a_2 \Pi^{[1]}_{ij}(\boldsymbol{x}) \psi^{ij}(\boldsymbol{q})$$

$$3^{\rm rd} \quad\quad a_2 \,{\rm Tr}[\Pi^{[1]}(\boldsymbol{x})] \Pi^{[1]}_{ij}(\boldsymbol{x}) \psi^{ij}(\boldsymbol{q}) \,, \; \psi^{ij}(\boldsymbol{q}) \Pi^{[2]}_{ij}(\boldsymbol{x}) \,.$$

Note that the leading non-Gaussian contribution to the bias expansion, $\Pi^{[1]}_{ij} \psi^{ij} = \psi_{ij} K^{ij}$, is a second-order term. For this reason, anisotropic PNG with $J = 2$ does not lead to a prominent large-scale signature in the galaxy two-point function. It does however produce a scale dependence $\propto k^{-2+\alpha}$ in the quadrupole of the galaxy bispectrum [540] (see Sec. 7.1.6), which for $\alpha < 2$ is a smoking-gun signature of inflationary physics just as the scale-dependent bias from local PNG is. Further, the same effect produces a strong scale-dependence in the two-point function of galaxy *shapes* which are expected to keep a memory of an anisotropic local initial power spectrum of small-scale fluctuations [552, 554].

Finally, in the presence of the next-higher spin contribution to PNG, $J = 4$, we need to introduce a tensor field ψ_{ijkl} which is trace-free and symmetric with respect to all indices [540]. Thus, symmetry dictates that the leading contribution (at lowest order in derivatives) is at cubic order,

$$J = 4 : \quad 1^{\rm st} \quad\quad - \tag{7.32}$$

$$2^{\rm nd} \quad\quad -$$

$$3^{\rm rd} \quad\quad a_4 \Pi^{[1]}_{ij}(\boldsymbol{x}) \Pi^{[1]}_{kl}(\boldsymbol{x}) \psi^{ijkl}(\boldsymbol{q}) \,.$$

This leads to a characteristic scale-dependent signature only in the galaxy 4-point function (trispectrum).

7.1.4 Stochasticity from PNG

We now turn to the contributions from PNG to the stochastic terms in the bias expansion, discussed for Gaussian initial conditions in Sec. 2.8, beginning with the isotropic case, $J = 0$. By assumption, the stochastic variables ε, ε_O introduced there only depend on the statistics of the small-scale initial perturbations. As long as the coupling between long and short modes is completely captured by Eq. (7.25), all effects are accounted for in our non-Gaussian basis Eq. (7.24). In this case, the list of terms in the Gaussian case, Eq. (2.87) in Sec. 2.8, only needs to be augmented by terms of the same type multiplied by $a_0 \psi(\boldsymbol{q})$,

$$J = 0 : \quad 1^{\rm st} \quad\quad - \tag{7.33}$$

$$2^{\rm nd} \quad\quad a_0 \varepsilon_\psi(\boldsymbol{x}) \psi(\boldsymbol{q})$$

$$3^{\rm rd} \quad\quad a_0 \varepsilon_{\psi\delta}(\boldsymbol{x}) \psi(\boldsymbol{q}) \,{\rm Tr}[\Pi^{[1]}(\boldsymbol{x})] \,.$$

Here, the fields ε_ψ, $\varepsilon_{\psi\delta}$ are characterized by their 1-point moments just as their counterparts for Gaussian initial conditions. For anisotropic PNG, for example $J = 2$, a stochastic term $\varepsilon_\psi^{ij}\psi_{ij}$ is the leading contribution. However, since the small-scale modes do not know of any preferred direction at leading order, this contribution has to be a higher-derivative contribution: $\varepsilon_\psi^{ij}(\boldsymbol{k}) = k^i k^j \hat{\varepsilon}_\psi(\boldsymbol{k})$, where $\hat{\varepsilon}_\psi$ approaches white noise on large scales. This further suppresses its numerical importance.

One can show that Eq. (7.33) together with Eq. (2.87) comprise the complete set of stochastic fields as along as *the initial conditions are derived from a single statistical field, corresponding to a single set of random phases* (App. B of [540]). This is the case for the *ansätze* Eq. (7.2) and Eq. (7.18) considered above. However, if the initial Bardeen potential ϕ consists of a superposition of two or more independent fields with different amounts of non-Gaussianity, for example

$$\phi(\boldsymbol{q}) = \varphi_G(\boldsymbol{q}) + \sigma_G(\boldsymbol{q}) + f_{\mathrm{NL}}^\sigma \sigma_G^2(\boldsymbol{q}), \tag{7.34}$$

where φ_G and σ_G are independent Gaussian random fields, then Eq. (7.24) together with Eq. (7.33) are not sufficient. Such models are characterized by a large trispectrum $\langle \phi(\boldsymbol{k}_1)\phi(\boldsymbol{k}_2)\phi(\boldsymbol{k}_3)\phi(\boldsymbol{k}_4)\rangle$ (Fourier transform of the four-point correlation function) in the collapsed limit, where the k_i are comparable in magnitude but either $|\boldsymbol{k}_{13}|$ or $|\boldsymbol{k}_{24}|$ is much smaller [555, 556]; the trispectrum amplitude in this limit is customarily parametrized by a parameter τ_{NL}. A contribution of this type leads to a cutoff-dependent loop contribution in the galaxy two-point function which is not absorbed by $\psi(\boldsymbol{q})$ or any other term in the basis Eq. (7.24). Thus, for this type of PNG it is necessary to add an independent field $\hat{\psi}$, with associated bias parameter $b_{\hat{\psi}}$, to the bias expansion, where $\hat{\psi}$ is uncorrelated with the Gaussian part of the initial conditions ϕ [543, 556, 540] [in contrast with ψ, which is completely correlated with ϕ, Eq. (7.23)]. This potentially large contribution to galaxy clustering on large scales was already pointed out by [557].

Since $\hat{\psi}$ is uncorrelated with long-wavelength modes of ϕ and only correlates with itself, it does not affect the galaxy-matter cross-correlation, but gives a non-vanishing contribution to the galaxy power spectrum

$$P_{gg}^{\mathrm{LO}}(k) \supset \tau_{\mathrm{NL}}(b_{\hat{\psi}})^2 P_{\hat{\psi}}(k). \tag{7.35}$$

For the local-type example given in Eq. (7.34), we have $P_{\hat{\psi}}(k) = P_\phi(k)$. Thus, this term will be one of the dominant terms on sufficiently large scales. In this case, the leading-order correlation coefficient between matter and galaxies becomes

$$r_{gm}^{\mathrm{LO}}(k) \equiv \left.\frac{P_{gm}(k)}{\sqrt{P_{gg}(k)P_{mm}(k)}}\right|_{\mathrm{LO}} \overset{f_{\mathrm{NL}}\neq 0}{=} \frac{b_\psi P_{\delta\psi}(k)}{\sqrt{\left[(b_\psi)^2 P_\psi(k) + (\tau_{\mathrm{NL}}/f_{\mathrm{NL}}^2)(b_{\hat{\psi}})^2 P_{\hat{\psi}}(k)\right]P_{\mathrm{L}}(k)}}, \tag{7.36}$$

where $P_{\psi\delta}(k) \equiv \langle\psi(\boldsymbol{k})\delta^{(1)}(\boldsymbol{k}')\rangle'$. This is equal to unity if and only if $\tau_{\mathrm{NL}}b_{\hat{\psi}} = 0$, otherwise the correlation coefficient between matter and galaxies is less than one. Note the very different scale dependence of the correlation coefficient compared to the Gaussian case, Eq. (2.35), discussed in Sec. 2.3. Hence, by measuring the correlation coefficient between galaxies and matter on large scales, we can determine whether the collapsed limit of the four-point function exceeds the value predicted for initial conditions sourced by a single degree of freedom.

7.1.5 Beyond the squeezed limit

The leading non-Gaussian contribution to the galaxy two-point function for general scale-invariant quadratic PNG is given by Eq. (7.26), which was derived by evaluating the primordial bispectrum in the squeezed limit. This limit however is only valid if modes with wavelength of order the separation k do not contribute appreciably to the variance $[\sigma(R_*)]^2$ of the small-scale density field on the scale R_* relevant for the formation of the tracer. Empirically, one finds that the squeezed limit is no longer a good approximation when k approaches the peak of the matter power spectrum ($k \sim 0.02\,h\,\mathrm{Mpc}^{-1}$) [544]. We can deal with this issue perturbatively by expanding the initial matter bispectrum beyond the leading squeezed limit.

Assuming a scale-free bispectrum of primordial potential perturbations as given in Eq. (7.19), we obtain (see [558] for details),

$$B^{(1)}(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_\ell) = 4 \sum_{J=0,2,4,\ldots} \left\{ a_J^{\{0\}} \left(\frac{k_\ell}{k_S} \right)^\alpha + a_J^{\{2\}} \left(\frac{k_\ell}{k_S} \right)^{\alpha+2} f_J^{\{2\}}(k_S) + \cdots \right\} P_\phi(k_\ell) P_{\rm L}(k_S). \qquad (7.37)$$

The subleading term is no longer scale-free, since it involves derivatives of the transfer function $dT(k)/dk$. However, given a primordial power spectrum and bispectrum, the subleading coefficient $a_J^{\{2\}}$ and function $f_J^{\{2\}}(k_S)$ are uniquely determined. Moreover, if α is the only (non-analytic) scaling dimension in the full bispectrum, which holds for commonly considered shapes such as local, equilateral, orthogonal, quasi-single-field, and the folded template (see Sec. 7.2), then the leading correction beyond the squeezed limit is suppressed by $(k_\ell/k_S)^2$ relative to the leading term, which is the scaling we assume here. The reasoning in the following does not rely on this assumption however. Note that even if, at leading order, only the $J=0$ term exists, terms of higher order in J will in general appear at subleading order.

The key point to notice is that the correction beyond the squeezed limit is of the same general form as the leading term. Thus, specializing to a $J=0$ contribution, we merely have to introduce another field $\tilde{\psi}$ in the bias expansion, where $\tilde{\psi}$ is defined as a nonlocal transformation of ϕ, Eq. (7.23) with $\alpha \to \alpha + 2$. Adding $\tilde{\psi}$ as counter-term to $[\delta^2]$ in Eq. (7.28) then removes the cutoff dependence up to order $(k/\Lambda)^4$ as desired [558]. At leading order, this leads to a galaxy auto-power spectrum given by

$$P_{gg}^{\rm LO}(k)\Big|_{J=0} = \left[b_1 + \left(a_0^{\{0\}} b_\psi k^\alpha + a_0^{\{2\}} b_{\tilde\psi} k^{2+\alpha} \right) \mathcal{M}^{-1}(k) \right]^2 P_{\rm L}(k) + P_\varepsilon^{\{0\}}. \qquad (7.38)$$

Correspondingly, for the $J > 0$ contributions in Eq. (7.37) a similar reasoning applies, as described in Sec. 7.1.3. These become relevant for loop corrections to the two-point function as well as the galaxy three- and higher-point functions. We will see in Sec. 7.3 that, if we consider halos which approximately follow a universal mass function, then $b_{\tilde\psi} \sim R_*^2 b_\psi$, where R_* is the Lagrangian radius of the halos. This can serve as a rough estimate of the magnitude of the beyond-squeezed-limit contributions.

At this point, it is also worth noting that there are other terms which contribute to the galaxy power spectrum Eq. (7.38) at a similar order as the beyond-squeezed-limit terms written in Eq. (7.38). First, the Gaussian higher-derivative terms combine with the leading non-Gaussian term to a contribution

$$-2 b_\psi b_{\nabla^2 \delta} k^2 \langle \psi(\boldsymbol{k}) \delta(\boldsymbol{k}') \rangle'.$$

Second, in the same way as higher-derivative terms, such as $b_{\nabla^2 \delta} \nabla^2 \delta$, appear in the general bias expansion for Gaussian initial conditions (Sec. 2.6), we expect related terms to be present for the field ψ in the non-Gaussian case, i.e. $b_{\nabla^2 \psi} \nabla_q^2 \psi(\boldsymbol{q})$ at leading order. For example, if the abundance of galaxies depends on the statistics of initial fluctuations within a region of scale R_*, then such a term is obtained after formally expanding the spatial convolution, with $b_{\nabla^2 \psi} \propto R_*^2$. All these terms scale as $R_*^2 k^2$ times the leading PNG contribution and are thus comparable to the beyond-squeezed-limit contribution given in Eq. (7.38). Note that this is in contrast to the scale dependence induced by the transfer function entering through $\mathcal{M}^{-1}(k)$, which is controlled by $k_{\rm eq}^{-1} \gg R_*$ [see the discussion around Eq. (7.27)].

In Sec. 7.4.2, we will show that if we make the strong assumption that galaxy formation is an exactly local function of the initial density field smoothed on a single scale R, then both the beyond-squeezed-limit contributions as well as terms of the type $b_{\nabla^2 \psi} \nabla_q^2 \psi(\boldsymbol{q})$ are uniquely determined by an integral over the bispectrum.

7.1.6 Summary

In Sec. 7.1.1–7.1.5, we have derived the contributions to the general galaxy bias expansion that appear in the presence of PNG, focusing on the leading, quadratic (in the Bardeen potential) PNG parametrized by the bispectrum of potential perturbations. We now summarize the complete set of bias contributions up to cubic order in perturbations. This complements the expansion given in Eq. (2.135) for Gaussian initial

conditions (Sec. 2.11). We first set $\tau_{\rm NL} = 0$, to obtain

$$
\begin{aligned}
\delta_g(\boldsymbol{x})\Big|_{a_0,\,a_2,\,a_4} = a_0\Bigg\{ & b_\psi \psi(\boldsymbol{q}) + b_{\psi\delta}\psi(\boldsymbol{q})\delta(\boldsymbol{x}) + b_{\psi\delta^2}\psi(\boldsymbol{q})\delta^2(\boldsymbol{x}) + b_{\psi K^2}\psi(\boldsymbol{q})(K_{ij})^2(\boldsymbol{x}) \\
& + \varepsilon_\psi(\boldsymbol{x})\psi(\boldsymbol{q}) + \varepsilon_{\psi\delta}(\boldsymbol{x})\psi(\boldsymbol{q})\delta(\boldsymbol{x}) + b_{\nabla^2\psi}\nabla_q^2\psi(\boldsymbol{q}) \Bigg\} \\
& + a_2\Bigg\{ b_{\psi K}\psi^{ij}(\boldsymbol{q})K_{ij}(\boldsymbol{x}) + b_{\psi K\delta}\psi^{ij}(\boldsymbol{q})K_{ij}(\boldsymbol{x})\delta(\boldsymbol{x}) + b_{\psi\Pi^{[2]}}\psi^{ij}(\boldsymbol{q})\Pi_{ij}^{[2]}(\boldsymbol{x}) \\
& + \varepsilon_{\psi K}(\boldsymbol{x})\psi^{ij}(\boldsymbol{q})K_{ij}(\boldsymbol{x}) \Bigg\} + a_4 b_{\psi KK}\psi^{ijkl}(\boldsymbol{q})K_{ij}(\boldsymbol{x})K_{kl}(\boldsymbol{x})\,, \quad (7.39)
\end{aligned}
$$

where we restrict the expansion to linear order in the parameters a_J, as discussed at the end of Sec. 7.1.2. Here, we have emphasized that the fields ψ, ψ_{ij}, ψ_{ijkl} are evaluated at the Lagrangian position $\boldsymbol{q}$ corresponding to the Eulerian position $(\boldsymbol{x}, \tau)$. Further, these three fields are defined as nonlocal transformations of the primordial Bardeen potential ϕ,

$$
\begin{aligned}
\psi(\boldsymbol{q}) &\equiv \int_{\boldsymbol{k}} k^\alpha \phi(\boldsymbol{k}) e^{i\boldsymbol{k}\cdot\boldsymbol{q}} \\
\psi_{ij}(\boldsymbol{q}) &\equiv \int_{\boldsymbol{k}} \frac{3}{2}\left(\frac{k_i k_j}{k^2} - \frac{1}{3}\delta_{ij}\right) k^\alpha \phi(\boldsymbol{k}) e^{i\boldsymbol{k}\cdot\boldsymbol{q}} \\
\psi_{ijkl}(\boldsymbol{q}) &\equiv \int_{\boldsymbol{k}} \frac{35}{8}\mathcal{P}_{ijkl}(\hat{\boldsymbol{k}}) k^\alpha \phi(\boldsymbol{k}) e^{i\boldsymbol{k}\cdot\boldsymbol{q}}\,, \quad (7.40)
\end{aligned}
$$

where $\mathcal{P}_{ijkl}(\hat{\boldsymbol{k}})$ denotes the complete trace-free projection operator with respect to $\boldsymbol{k}$ [see Eq. (A.25) for the explicit expression].

All operators appearing in Eq. (7.39) are understood to be renormalized; for clarity, we have omitted the square brackets. The first two lines, $\propto a_0$, contain the contributions for isotropic PNG. The first line contains the generalization of the deterministic LIMD bias expansion, while the second line gives the stochastic and higher-derivative contributions. Here, we have made the same approximation as in Eq. (2.135), and included only the leading, linear higher-derivative term. Note that in the case of PNG, this also includes beyond-squeezed-limit contributions (Sec. 7.1.5) which in general obey a different scaling than other higher-derivative contributions.

The third and fourth lines, $\propto a_2$ and a_4, contain the deterministic and stochastic contributions for anisotropic PNG ($J = 2$ and 4). Due to symmetry, these terms start at second and third order in perturbations, respectively. Correspondingly, we have not included a higher-derivative contribution, although the same caveats regarding beyond-squeezed-limit terms mentioned above also apply here.

Finally, we give contributions $\propto \tau_{\rm NL}$, which are present if the initial conditions are a superposition of several random fields, as discussed in Sec. 7.1.4. Note that, by definition, $\hat{\psi}$ is not correlated with the other long-wavelength perturbations ψ, δ, K_{ij}, and, at this order, is completely specified by its power spectrum $P_{\hat{\psi}}(k)$. We obtain, in precise analogy with the terms $\propto a_0$ in Eq. (7.39),

$$
\begin{aligned}
\delta_g(\boldsymbol{x})\Big|_{\tau_{\rm NL}} = \tau_{\rm NL}\Bigg\{ & b_{\hat{\psi}}\hat{\psi}(\boldsymbol{q}) + b_{\hat{\psi}\delta}\hat{\psi}(\boldsymbol{q})\delta(\boldsymbol{x}) + b_{\hat{\psi}\delta^2}\hat{\psi}(\boldsymbol{q})\delta^2(\boldsymbol{x}) + b_{\hat{\psi}K^2}\hat{\psi}(\boldsymbol{q})(K_{ij})^2(\boldsymbol{x}) \\
& + \varepsilon_{\hat{\psi}}(\boldsymbol{x})\hat{\psi}(\boldsymbol{q}) + \varepsilon_{\hat{\psi}\delta}(\boldsymbol{x})\hat{\psi}(\boldsymbol{q})\delta(\boldsymbol{x}) + b_{\nabla^2\hat{\psi}}\nabla_q^2\hat{\psi}(\boldsymbol{q}) \Bigg\}\,. \quad (7.41)
\end{aligned}
$$

We can further restrict Eq. (7.39) to the most commonly considered case of local PNG by setting

$a_0 \to f_{\rm NL}$, $\psi \to \phi$, and $a_2 = 0$, $a_4 = 0$, $\tau_{\rm NL} = 0$. This yields

$$\delta_g(\boldsymbol{x})\Big|_{f_{\rm NL}} = f_{\rm NL}\Bigg\{ b_\phi \phi(\boldsymbol{q}) + b_{\phi\delta}\phi(\boldsymbol{q})\delta(\boldsymbol{x}) + b_{\phi\delta^2}\phi(\boldsymbol{q})\delta^2(\boldsymbol{x}) + b_{\phi K^2}\phi(\boldsymbol{q})(K_{ij})^2(\boldsymbol{x})$$
$$+ \varepsilon_\phi(\boldsymbol{x})\phi(\boldsymbol{q}) + \varepsilon_{\phi\delta}(\boldsymbol{x})\phi(\boldsymbol{q})\delta(\boldsymbol{x}) + b_{\nabla^2\phi}\nabla_q^2\phi(\boldsymbol{q})\Bigg\} . \tag{7.42}$$

Ref. [540] was the first to give the complete expressions Eqs. (7.39)–(7.42). Initial studies of bias expansions in the presence of local-type PNG only considered the first term, $b_\phi\phi$. Ref. [538] extended the expansion by including all terms of the form $\phi^n(\boldsymbol{q})\delta^m(\boldsymbol{x})$. They did not, however, include the tidal term $\propto \phi(K_{ij})^2$, nor the stochastic and higher-derivative terms in the second line of Eq. (7.42). This also applies to [541]. Ref. [559] included the Gaussian tidal term $(K_{ij})^2$, but did not write down its non-Gaussian counterpart $\phi(K_{ij})^2$ as they restricted to a second-order bias expansion. They also did not consider the stochastic term $\varepsilon_\phi\phi$ which is however relevant at second order (see below).

After having given the general bias expansion in the presence of quadratic PNG, we now turn to galaxy statistics, generalizing the results of Sec. 4.1.1. Specifically, we provide a succinct summary of the leading-order galaxy power spectrum and bispectrum in the rest frame (without RSD and other projection effects). We further make explicit the time dependence of all quantities.

First, the leading-order cross- and auto-power spectra of galaxies are given by

$$P_{gm}^{\rm LO}(k,\tau) = [b_1(\tau) + \Delta b_1(k,\tau)]\, P_{\rm L}(k,\tau)$$
$$P_{gg}^{\rm LO}(k,\tau) = [b_1(\tau) + \Delta b_1(k,\tau)]^2\, P_{\rm L}(k,\tau) + P_\varepsilon^{\{0\}}(\tau) + \tau_{\rm NL}[b_{\hat\psi}(\tau)]^2 P_{\hat\psi}(k) , \tag{7.43}$$

where

$$\Delta b_1(k,\tau) \equiv a_0 b_\psi(\tau) k^\alpha \mathcal{M}^{-1}(k,\tau) \stackrel{\rm local\ PNG}{=} f_{\rm NL} b_\phi(\tau)\mathcal{M}^{-1}(k,\tau) \tag{7.44}$$

is the *scale-dependent bias* induced by PNG. This is, of course, merely a rephrasing of the *nonlocal bias* $b_\psi(\tau)\psi(\boldsymbol{q})$. Here, we have let α denote the squeezed-limit scaling index of the $J = 0$ contribution in Eq. (7.19). For example, for local and equilateral type PNG [560, 561, 562], we have $\alpha = 0$, and 2, respectively. Quasi-single field inflation [563, 564] yields $0 \le \alpha \le 3/2$.

Note that we have not included the beyond-squeezed limit and higher-derivative corrections, which typically scale as $R_*^2 k^2 \Delta b_1(k) P_{\rm L}(k)$; see Sec. 7.1.5 for a discussion. Further, while the set of cubic operators listed above is sufficient to derive the NLO (1-loop) contribution from PNG to the galaxy power spectrum, these terms are highly suppressed and we thus refrain from writing the full expression here; they involve three new bias parameters: $a_0 b_{\psi\delta}$, $a_2 b_{\psi K}$, $a_2 b_{\psi\Pi^{[2]}}$. In order to obtain an order-of-magnitude estimate, one can replace $P_{\rm L}(p)$ in the standard NLO terms, Eq. (4.22) on p. 82 in Sec. 4.1.4, with $p^\alpha \mathcal{M}^{-1}(p)P_{\rm L}(p)$. Even for $\alpha = 0$, this leads to terms which are suppressed by $\sim 10^{-4} a_J$ relative to the standard NLO terms, and would thus only be relevant if $a_J \gtrsim 100$.

Next, we give the complete expression for the rest-frame galaxy bispectrum at leading order, including all relevant terms from Eq. (7.39) [540]. Here, there are considerably more terms, since all quadratic combinations of ψ and, for $J = 2$, ψ_{ij} appear. It is useful to decompose the separable contributions to the galaxy bispectrum in terms of Legendre polynomials $\mathcal{L}_l(\hat{\boldsymbol{k}}_i \cdot \hat{\boldsymbol{k}}_j)$ of the cosine between two wavevectors:

$$B_{ggg}^{\rm LO}(k_1,k_2,k_3,\tau) = b_1^3 B^{(1)}(k_1,k_2,k_3,\tau) + \sum_{l=0,1,2}\Big[P_{\rm L}(k_1,\tau)P_{\rm L}(k_2,\tau)\mathcal{B}^{[l]}(k_1,k_2,\tau)\,\mathcal{L}_l(\hat{\boldsymbol{k}}_1\cdot\hat{\boldsymbol{k}}_2) + 2\text{ perm.}\Big] . \tag{7.45}$$

The first contribution here is simply the leading-order matter bispectrum Eq. (7.4). Note that Eq. (7.45) is *not restricted to the squeezed limit, but valid for all configurations*. Hence, all multipoles l including odd ones appear here. As in Sec. 2.8, we define the large-scale stochastic amplitudes

$$P_{\varepsilon_O \varepsilon_{O'}}^{\{0\}}(\tau) \equiv \lim_{k\to 0}\langle \varepsilon_O(\boldsymbol{k},\tau)\varepsilon_{O'}(\boldsymbol{k}',\tau)\rangle' \tag{7.46}$$

174

and $B_{\varepsilon}^{\{0\}}(\tau)$ [Eq. (2.86)]. While we restrict to B_{ggg} here, the results for B_{ggm} and B_{gmm} can be obtained analogously.

The monopole consists of deterministic and stochastic contributions given by

$$\mathcal{B}^{[0]}(k_1, k_2, \tau) = \big(b_1(\tau) + \Delta b_1(k_1, \tau)\big)\big(b_1(\tau) + \Delta b_1(k_2, \tau)\big)$$
$$\times \left[\frac{34}{21}b_1(\tau) + b_2(\tau) + \frac{b_{\psi\delta}(\tau)}{b_\psi(\tau)}\big(\Delta b_1(k_1, \tau) + \Delta b_1(k_2, \tau)\big)\right]$$
$$+ \tau_{\mathrm{NL}}b_1(\tau)b_{\hat\psi}(\tau)b_{\hat\psi\delta}(\tau)\left(\frac{P_{\hat\psi}(k_1)}{P_{\mathrm{L}}(k_1, \tau)} + \frac{P_{\hat\psi}(k_2)}{P_{\mathrm{L}}(k_2, \tau)}\right) + \mathcal{B}_\varepsilon^{[0]}(k_1, k_2, \tau)$$

$$\mathcal{B}_\varepsilon^{[0]}(k_1, k_2, \tau) = \frac{B_\varepsilon^{\{0\}}(\tau)}{3P_{\mathrm{L}}(k_1, \tau)P_{\mathrm{L}}(k_2, \tau)}$$
$$+ \left[(b_1(\tau) + \Delta b_1(k_1, \tau))\left(P_{\varepsilon\varepsilon_\delta}^{\{0\}}(\tau) + \frac{P_{\varepsilon\varepsilon_\psi}^{\{0\}}(\tau)}{b_\psi(\tau)}\Delta b_1(k_1, \tau)\right)\frac{1}{P_{\mathrm{L}}(k_2, \tau)} + \{1 \leftrightarrow 2\}\right]$$
$$+ \tau_{\mathrm{NL}}b_{\hat\psi}(\tau)P_{\varepsilon\varepsilon_{\hat\psi}}^{\{0\}}(\tau)\frac{P_{\hat\psi}(k_1) + P_{\hat\psi}(k_2)}{P_{\mathrm{L}}(k_1, \tau)P_{\mathrm{L}}(k_2, \tau)}. \tag{7.47}$$

The dipole term, whose only contributions at this order come from expansions of the Lagrangian position q around the Eulerian position x [cf. Eq. (7.16)], is given by

$$\mathcal{B}^{[1]}(k_1, k_2, \tau) = \big(b_1(\tau) + \Delta b_1(k_1, \tau)\big)\big(b_1(\tau) + \Delta b_1(k_2, \tau)\big)\left[\frac{k_1}{k_2}\big(b_1(\tau) + \Delta b_1(k_1, \tau)\big) + \frac{k_2}{k_1}\big(b_1(\tau) + \Delta b_1(k_2, \tau)\big)\right]$$
$$+ \tau_{\mathrm{NL}}b_1(\tau)[b_{\hat\psi}(\tau)]^2\left(\frac{k_1}{k_2}\frac{P_{\hat\psi}(k_1)}{P_{\mathrm{L}}(k_1, \tau)} + \frac{k_2}{k_1}\frac{P_{\hat\psi}(k_2)}{P_{\mathrm{L}}(k_2, \tau)}\right). \tag{7.48}$$

That is, $\mathcal{B}^{[1]}$ is induced by the fact that the fields ψ, $\hat\psi$ are evaluated at the Lagrangian position. It is completely determined by the bias parameters b_1, b_ψ, $b_{\hat\psi}$ that appear in the galaxy power spectrum, and can hence serve as a clean, independent cross-check of the scale-dependent bias measured in the power spectrum [540].

Finally, the quadrupole

$$\mathcal{B}^{[2]}(k_1, k_2, \tau) = \frac{4}{3}\big(b_1(\tau) + \Delta b_1(k_1, \tau)\big)\big(b_1(\tau) + \Delta b_1(k_2, \tau)\big)\left[b_{K^2}(\tau) + \frac{2}{7}b_1(\tau)\right] \tag{7.49}$$
$$+ a_2 b_{\psi K}(\tau)\big(b_1(\tau) + \Delta b_1(k_1, \tau)\big)\big(b_1(\tau) + \Delta b_1(k_2, \tau)\big)\left[k_1^{\alpha_2}\mathcal{M}^{-1}(k_1, \tau) + k_2^{\alpha_2}\mathcal{M}^{-1}(k_2, \tau)\right],$$

contains contributions from tidal bias [$\propto b_{K^2}$, see Sec. 4.1.1] and from anisotropic PNG with $J = 2$ ($\propto a_2 b_{\psi K}$). Note the different prefactor in the second line, which is due to the definition of the field ψ_{ij} multiplied by the bias $b_{\psi K}$ [Eq. (7.40)]. Here, α_2 denotes the scaling index of the $J = 2$ contribution in Eq. (7.19). The bispectrum quadrupole thus allows for $J = 0$ and $J = 2$ contributions to be disentangled. As we will explain in the next section, this means that galaxy clustering can in principle probe the presence of additional fields in the early Universe as well as determine their spin.

Finally, while we have restricted to the two- and three-point functions here, higher n-point functions can also be used to constrain PNG. For example, the estimator constructed in [553] (out of the matter density field) essentially corresponds to measuring the collapsed trispectrum.

7.2 Probing inflation with galaxy clustering

In the previous section, we have seen that the squeezed-limit bispectrum and collapsed trispectrum of the initial conditions can lead to a rich array of signatures in the clustering of galaxies on large scales. We now briefly summarize which aspects of the physics of inflation are probed by these signatures.

Single-field inflation: In single-field inflation, one single scalar degree of freedom drives the expansion during the inflationary quasi-de Sitter phase, and leads to the primordial curvature perturbations that provide the seeds of large-scale structure. Interestingly, Ref. [565] showed that in single-field inflation and in the attractor regime, the squeezed-limit bispectrum is universal and corresponds to a local PNG with $f_{\mathrm{NL}} \propto n_s - 1$, where $n_s - 1 \approx 0.04$ is the deviation from the scale-invariance of the power spectrum of primordial curvature perturbations. This prediction, known as *consistency relation*, would suggest that there is a definite nonzero value of local f_{NL} predicted in single-field inflation that LSS surveys could target. On the other hand, multi-field inflationary models, which we discuss below, generically produce larger values of local f_{NL}. While $n_s - 1$ is small for the simplest single-field models, oscillations in the primordial power spectrum can significantly enhance this effect [566, 567, 568, 569].

However, the calculation of the bispectrum in [565] was done in a particular gauge (comoving gauge), and evaluated during the epoch of inflation, raising the question of how this effect transfers to late times and how it manifests in actual observations of galaxy clustering. Several recent papers [570, 95, 571, 145, 185, 572] have shown that the single-field consistency relation is equivalent to the statement that there is *no physical coupling* of long-wavelength potential perturbations ϕ_ℓ to small-scale perturbations. That is, *in single-field inflation, there is no scale-dependent bias in the rest frame of galaxies*; the leading effect of a large-scale perturbation enters as $\partial_i \partial_j \phi_\ell$, precisely as argued in Sec. 2.9. The apparent contradiction with Refs. [573, 574], who argue that terms in second-order relativistic perturbation theory (see [575] for a recent review) actually lead to an effective f_{NL} of order 1, is most likely resolved by including the mapping to local observables defined in terms of proper length and time units. Unlike the squeezed-limit bispectrum and hence value of f_{NL}, which depend on the coordinate system in which they are calculated, the absence of long-/short-mode coupling in the rest frame of comoving observers is a physical, gauge-invariant statement.

It is important to stress however that variants of single-field inflation that are *not* in the attractor regime, due for example to a non-Bunch-Davies state [576, 577, 578] or an initial kick of the inflaton [579], do lead to a local-type bispectrum *over a finite range of values of* k_ℓ/k_S. That is, while they lead to a physical scaling of $(k_\ell/k_S)^2$, as discussed above, for sufficiently small values of k_ℓ/k_S (that is, in comoving coordinates the bispectrum obeys the consistency relation), there can be an intermediate regime where k_ℓ/k_S is small and yet the bispectrum shows a scaling with $\alpha < 2$ (see [524] for a discussion). This also applies to resonant non-Gaussianity [566, 567]. Thus, the simple separable ansatz Eq. (7.19) is not sufficient for such models. A full derivation of the observable scale-dependent bias due to PNG in these models has been performed recently in [569], who showed that for models that satisfy CMB bounds as well as theoretical consistency, the scale-dependent bias is unlikely to be observable. Previously, Ref. [568] had reached different conclusions, as they included the unphysical contribution from the consistency relation.

Note that the mapping from the galaxy rest frame to the observer's frame on Earth by way of tracing photon geodesics through the perturbed spacetime, which includes gravitational redshift, Doppler effect, and lensing, does lead to an apparent scale-dependent bias in the observed galaxy clustering. We review these contributions in Sec. 9. However, these effects are present even for perfectly Gaussian initial conditions, and unrelated to inflationary physics. They can easily be calculated given a fiducial cosmological model and the luminosity function of the sampled galaxy population. Thus, the observation of any residual scale-dependent bias $\propto k^{-2}$ in galaxy clustering—after all light-propagation effects have been taken into account—would rule out single-field inflation in the attractor regime. Moreover, the projected constraints on f_{NL} from planned future galaxy surveys are at the level of $\sigma(f_{\mathrm{NL}}) \sim 1$ (Sec. 7.6) and thus expected to improve upon current CMB constraints ($|f_{\mathrm{NL}}| \lesssim 5$ [580]) significantly.

By definition, in single-field inflation there is only a single source of random phases which precludes any large-scale stochasticity between galaxies and matter of the type described in Sec. 7.1.4. A detection of the latter would thus similarly rule out single-field inflation. Detecting this signature requires observations of both the matter-galaxy cross-power spectrum, for example from weak gravitational lensing, and the galaxy power spectrum over the same volume. No detailed study of the detectability of such a signature has been published yet.

Finally, single-field inflation does generally produce *nonlocal* non-Gaussianities with $\alpha = 2$ [565, 576, 581]. In particular, the precise shapes are of the equilateral and orthogonal types [582]. While canonical single-field

slow-roll inflation leads to (physical) amplitudes of $f_{\rm NL}^{\rm eql} \sim 0.01$ [583], these non-Gaussianities are enhanced in more general single-field models where the inflaton has a small sound speed c_s. In this case, one has $f_{\rm NL}^{\rm eql} \simeq 1/c_s^2$ [581]. Note that, since these types of PNG have $\alpha = 2$ in the squeezed limit [Eq. (7.26)], they do not lead to a large-scale scale-dependent bias. In fact, improving upon CMB constraints on $f_{\rm NL}^{\rm eql}$ with large-scale structure will be quite challenging [263].

Multifield inflation: By definition, these models involve additional light degrees of freedom beyond a single scalar, and generically produce local PNG with $f_{\rm NL}$ of order 1 or larger (see [584] for a review). Moreover, if some of the fields have a mass that is comparable to the Hubble rate during inflation, then a scale-dependent bias [Eq. (7.26)] $\propto k^{\alpha-2}$ with $0 \leq \alpha \leq 3/2$ is induced, and a measurement of α allows for a measurement of the mass m of the field, provided $m < 3H/2$ [563, 564, 585, 586]. For higher-mass fields, oscillatory features are present in the bispectrum [587, 588, 589]. Further, the large-scale stochasticity in the relation between galaxies and mass described in Sec. 7.1.4 is a unique feature of multifield models (see [543, 556] for concrete examples).

Higher-spin fields: An anisotropic squeezed-limit bispectrum [$J = 2, 4, \cdots$ in Eq. (7.19)] leads to further unique signatures in galaxy clustering, as we have seen in Sec. 7.1.3. This type of PNG can signal the presence of higher-spin fields during inflation [588, 590, 591]; "solid inflation" [592, 593] and "Chromo-natural inflation" [594] are models of a different type which also lead to anisotropic PNG. In addition, inflationary models with *anisotropic* non-Bunch-Davies (i.e. non-vacuum) state [595, 596] as well as those with large primordial magnetic fields [597, 598] in general also produce anisotropic squeezed-limit bispectra.

In summary, we see that large-scale galaxy statistics provide numerous possibilities for probing the statistics of the initial conditions and putting constraints on inflationary models. However, in order to derive constraints on $f_{\rm NL}, a_J, \tau_{\rm NL}, \cdots$, we need predictions for the corresponding bias parameters $b_\psi, b_{K\psi}, b_{\hat\psi}, \cdots$. The peak-background split provides a possibility for deriving such predictions from semi-analytical methods and simulations.

7.3 Non-Gaussian bias parameters from the peak-background split

In Sec. 3.1, we were able to derive the renormalized LIMD bias parameters $b_N \equiv N! \, b_{\delta^N}$ as a response of the abundance of halos to a change in the background density via the PBS argument. We now present similar identifications for the scale-dependent bias from PNG, specifically the parameters $b_{\phi\delta^N}$, $b_{\psi\delta^N}$ ($N \geq 0$) which appear in the expansions Eq. (7.17) and Eq. (7.24) for local and nonlocal PNG, respectively (we restrict to isotropic PNG with $J = 0$ throughout). Similar to Sec. 3, we begin with the general case which applies to galaxies and other tracers, and then specialize to the case of halos following a universal mass function.

In order to derive $b_{\phi\delta^N}$, $b_{\psi\delta^N}$, we need to consider the response of the galaxy number density to a change in the amplitude of initial density fluctuations, which is where the scale-dependent bias physically originates, through Eq. (7.15) and Eq. (7.25). The simplest way to parametrize such a dependence is to rescale initial perturbations by a factor of $1 + 2\epsilon k^{-\alpha}$ from their fiducial value, where ϵ is an infinitesimal parameter [corresponding to $a_0\psi$, Eq. (7.25)]. 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 and shape by rescaling the initial density perturbations by $(1 + 2\epsilon k^{-\alpha})$. We then define the Lagrangian non-Gaussian PBS bias $b_{\psi\delta^N}^L$ by generalizing the expression Eq. (3.4) to

$$b_{\psi\delta^N}^L \equiv \frac{1}{N!} \frac{1}{\overline{n}_g(0)} \frac{\partial^{N+1}\overline{n}_g(\Delta, \epsilon)}{\partial\Delta^N \partial\epsilon}\bigg|_{\Delta=0, \epsilon=0}. \tag{7.50}$$

This includes b_ψ as a special case for $N = 0$, and b_ϕ for $N = 0$, $\alpha = 0$. Note that the distinction between Lagrangian and Eulerian bias parameters matters only for $N \geq 1$, as the ϵ-derivative is always defined with respect to the initial fluctuations. For $N \geq 1$, the mapping between Lagrangian $b_{\psi\delta^N}^L$ and Eulerian $b_{\psi\delta^N}^E$ can be derived in analogy to the Gaussian case of Sec. 2.1; specifically, $b_{\psi\delta^N}^E$ is given by $b_{\psi\delta^N}^L$ plus corrections involving the $b_{\psi\delta^m}^E$ with $0 \leq m < N$.

The bias parameters defined in Eq. (7.50) can be understood as follows. The mean comoving galaxy number density is a function of the mean comoving matter density $\overline{\rho}_m$ and the initial power spectrum of

density fluctuations (and higher n-point functions for PNG). $b^L_{\psi\delta^N}$ then denotes a specific derivative of this function with respect to $\overline{\rho}_m$ and ϵ at fiducial values of $\overline{\rho}_m$ and the primordial power spectrum normalization. Thus, through a simple generalization, the PBS argument can be used to derive the leading bias parameters in the non-Gaussian case as well. Specifically, the leading effect of local PNG ($\alpha = 0$), quantified by b_ϕ, is controlled by the response of the mean number density of galaxies to a rescaling of the amplitude $\mathcal{A}_s$ of initial fluctuations,[20]

$$b_\phi = 4 \frac{1}{\overline{n}_g} \frac{\partial \overline{n}_g}{\partial \ln \mathcal{A}_s} = 2 \frac{1}{\overline{n}_g} \frac{\partial \overline{n}_g}{\partial \ln \sigma_8} \, . \tag{7.51}$$

Here, we have used that a scale-independent rescaling of the initial conditions can be equally parametrized through $\mathcal{A}_s$ or σ_8. This relation, and more generally Eq. (7.50), can be evaluated for any given prescription that predicts the mean abundance of galaxies or halos. We discuss the application of Eq. (7.51) to halos identified in N-body simulations, as performed in [315, 599], in Sec. 7.5. For now, we begin with some general considerations.

Let us assume that the number density of the given galaxy sample depends on the amplitude of initial fluctuations chiefly on a particular scale R. Then, we can conclude that:

- For *local* PNG, the particular scale R is irrelevant for $b_{\phi\delta^N}$, as all perturbations $\delta(\boldsymbol{k})$ are rescaled uniformly by a long-wavelength potential perturbation.

- For *nonlocal separable* bispectra with index α as in Eq. (7.19), the scale of the small-scale perturbations that govern the abundance of galaxies does matter for the bias parameters. Note that $[b_{\psi\delta^N}] = (\text{Mpc})^\alpha$. By dimensional analysis, we thus expect that $b_{\psi\delta^N} \propto R^\alpha$. If the galaxy number density depends on the amplitude of initial perturbations on several different scales, then b_ψ becomes a linear combination of these different dependencies with relative weights controlled by α, i.e. the shape of the primordial bispectrum.

- *Non-separable* bispectra first need to be approximated by a linear combination of separable shapes (see e.g. [549]). In that case, the galaxy statistics in general involve several different b_{ψ_i} with different α_i.

Thus, a given galaxy population responds differently to different shapes of primordial non-Gaussianity, i.e. b_ψ and $b_{\psi O}$ depend on the galaxy sample as well as the shape of the primordial bispectrum.

We now turn to the special case of dark matter halos. Specifically, we consider a slight generalization of the universal mass function discussed in Sec. 3.3. We write the mean abundance of halos as

$$\overline{n}_h = \overline{n}_h \left(\overline{\rho}_m, \sigma \right) J \, , \quad \text{where} \quad \sigma \equiv \sigma(R[M]) \, , \quad J \equiv \left| \frac{d \ln \sigma}{d \ln M} \right| \, . \tag{7.52}$$

That is, $\overline{n}_h$ is given as a function of the mean density of the Universe and the variance of the density field smoothed on the scale $R(M)$, as well as its derivative with respect to scale. The Jacobian J transforms from an interval in $\ln \sigma$ to an interval in $\ln M$, and we thus assume that $\overline{n}_h$ is linearly proportional to it (as is the case for the universal mass function introduced in Sec. 3.3). Under the rescaling $\delta(\boldsymbol{k}) \to [1 + 2\epsilon k^{-\alpha}]\delta(\boldsymbol{k})$, the small-scale root-mean-square density fluctuation σ transforms to lowest order as

$$\sigma \to \left[1 + 2\epsilon \frac{\sigma^2_{-\alpha/2}}{\sigma^2} \right] \sigma \, , \tag{7.53}$$

where $\sigma_{-\alpha/2} \equiv \sigma_{-\alpha/2}(R)$. The Jacobian transforms as

$$J \to \left[1 + 4\epsilon \frac{\sigma^2_{-\alpha/2}}{\sigma^2} \left(\frac{d \ln \sigma^2_{-\alpha/2}}{d \ln \sigma^2} - 1 \right) \right] J \, , \tag{7.54}$$

---

[20]Note that under the rescaling with $\alpha = 0$, $\mathcal{A}_s$ and σ_8 change by a factor $(1 + 4\epsilon)$ and $(1 + 2\epsilon)$, respectively.

where we have used $d/d\ln M = 2J\, d/d\ln\sigma^2$. Note that for local quadratic PNG, where $\alpha = 0$, the local Jacobian is not affected by long-wavelength modes, as expected. Using Eqs. (7.51)–(7.52), we can then derive the leading non-Gaussian bias through Eq. (7.50) [98, 100]:

$$b_\psi = \left[b_\phi + 4 \left(\frac{d\ln\sigma^2_{-\alpha/2}}{d\ln\sigma^2} - 1 \right) \right] \frac{\sigma^2_{-\alpha/2}}{\sigma^2}\,. \tag{7.55}$$

Note that we use a convention different than [98] for b_ψ, whence our factor of 4. Here, $b_\phi = b_\psi(\alpha = 0)$ is the PBS bias parameter quantifying the effect of local PNG for halos following Eq. (7.52). Thus, for halos following a universal mass function, the bias parameters quantifying the response to general nonlocal PNG are directly related to those for local PNG. Refs. [98, 600] first pointed out the contribution by the Jacobian J, which is numerically important for the match to the scale-dependent bias measured from N-body simulations of nonlocal PNG, as well as cubic local-type ($g_{\mathrm{NL}}\phi^3$) PNG [600, 546]. We will not discuss g_{NL} in detail, as its signature is strongly degenerate with that of f_{NL} (see, e.g., [601]), but note that the different mass-dependence could be exploited with multi-tracer techniques (as discussed in Sec. 7.6.2) in order to isolate g_{NL} from f_{NL}.

Note that the bias parameters associated with the additional subleading contributions from the squeezed-limit expansion of the bispectrum (Sec. 7.1.5) can be calculated in the same manner. Specifically, the bias parameter $b_{\tilde\psi}$ appearing in Eq. (7.38) is defined as the response of the mean abundance $\overline{n}_h$ to a change in the amplitude of initial density perturbations

$$\delta^{(1)}(\boldsymbol{k}) \to \left[1 + 2\epsilon k^{-\alpha-2} f_0^{\{2\}}(k) \right] \delta^{(1)}(\boldsymbol{k})\,. \tag{7.56}$$

Thus, for a mass function of the form Eq. (7.52), the subleading contribution is directly related to the leading scale-dependent bias b_ψ via spectral moments involving $k^{-\alpha-2} f_0^{\{2\}}(k)$.

We now specialize Eq. (7.52) to the well-known universal form [Eq. (3.21)],

$$\overline{n}_h = \frac{\overline{\rho}_m}{M}\, \nu_c f(\nu_c)\, J\,, \quad \nu_c \equiv \frac{\delta_{\mathrm{cr}}}{\sigma}\,, \tag{7.57}$$

where $f(\nu_c)$ is in general an arbitrary function of ν_c. In addition to the relation between b_ψ and b_ϕ [Eq. (7.55)], the specific form Eq. (7.57) further allows us to connect b_ϕ to the linear Lagrangian bias b_1^L derived from the PBS:

$$b_1^L = b_1 - 1 = \frac{1}{\overline{n}_h} \frac{\partial \overline{n}_h}{\partial\ln\overline{\rho}_m} = -\frac{1}{\sigma} \frac{1}{\nu_c f(\nu_c)} \frac{d[\nu_c f(\nu_c)]}{d\nu_c}$$

$$b_\phi = \frac{1}{\overline{n}_h} \frac{\partial \overline{n}_h}{\partial\epsilon} = 2\frac{1}{\overline{n}_h} \frac{\partial \overline{n}_h}{\partial\ln\sigma} = -2\frac{\delta_{\mathrm{cr}}}{\sigma} \frac{1}{\nu_c f(\nu_c)} \frac{d[\nu_c f(\nu_c)]}{d\nu_c} = 2\delta_{\mathrm{cr}} b_1^L\,. \tag{7.58}$$

This is the original relation between the density bias parameter and the response to primordial non-Gaussianity derived in [68, 539, 538], resulting in a scale-dependent bias for local PNG of

$$\Delta b_1(k, \tau) = f_{\mathrm{NL}} b_\phi \mathcal{M}^{-1}(k, \tau) = \frac{3 f_{\mathrm{NL}} \delta_{\mathrm{cr}} b_1^L \Omega_{m0} H_0^2}{k^2 T(k) D_{\mathrm{md}}(\tau)}\,. \tag{7.59}$$

We will discuss the accuracy of this result for actual halos in Sec. 7.5 (see Figs. 32–33). The relation between non-Gaussian and Gaussian PBS bias parameters can be continued to higher order via Eq. (7.50). For example, one easily obtains [538]

$$b_{\phi\delta}^L = -b_1^L + \delta_{\mathrm{cr}} b_2^L\,, \quad \text{and} \quad b_{\phi\delta}^E = b_{\phi\delta}^L + b_\phi\,. \tag{7.60}$$

The second relation follows immediately from the relation between Lagrangian and Eulerian halo densities (Tab. 5 in Sec. 2.2).

As in the Gaussian case, the non-Gaussian PBS biases $b_{\psi\delta^N}$ are the physical, renormalized bias parameters in the large-scale limit, which enter the n-point functions of galaxies and halos (Sec. 7.1.6) as well as other halo statistics, such as moments, as described in Sec. 4.

7.4 Non-Gaussian bias from Lagrangian bias models

In the previous sections, we have seen how the effect of primordial non-Gaussianity, which couples long- and short-wavelength modes, can be included in a model-independent way through an expansion of the bispectrum in the squeezed limit. We now discuss the non-Gaussian bias in the context of Lagrangian bias models. The essential difference between this approach and the general bias expansion Eq. (7.39) is that the "microscopic" perturbative bias expansion in the Lagrangian models remains the same as in the Gaussian case. That is, no new terms are introduced in the local description of the tracer abundance. Throughout, we will neglect gravitational evolution and perform the computation in Lagrangian space. This is justified as long as we consider sufficiently large scales where linear theory describes the evolution of the cosmic density field well. As we will see below, all the Lagrangian bias models predict the correct low-k scaling (for example, $\Delta b_1(k) \propto k^{-2}$ for local PNG), but they differ in their prediction for the amplitude of the non-Gaussian scale-dependent bias. Note that, in all these Lagrangian approaches, the scale R is a physical scale associated with the scale of galaxy formation, unlike the cutoff Λ^{-1} employed in Sec. 7.1.2.

7.4.1 Thresholding

We begin with a simple but illuminating example, the simple Lagrangian LIMD ansatz discussed in Sec. 2.1, in which the comoving Lagrangian halo number density is given by

$$n_h(\boldsymbol{q}) = n_{\text{thr}}(\boldsymbol{q}) \propto \Theta_H \left[\delta_R(\boldsymbol{q}) - \delta_{\text{cr}} \right] , \tag{7.61}$$

where Θ_H is the Heaviside function and δ_R is the density field smoothed on the *physical* scale R. The normalization constant, which we omitted to write here, cancels in the computation of correlation functions. The threshold δ_{cr} can be identified with the spherical collapse threshold (Sec. 5.2.1), so that the expectation value of Eq. (7.61), after taking a derivative with respect to M, corresponds to the Press-Schechter mass function (Sec. 5.2.2). The two-point correlation function of thresholded regions in case of Gaussian initial conditions was derived in Sec. 2.1 [Eq. (2.7)]. In the *high-peak limit* $\nu_c \gg 1$, Eq. (2.7) simplifies to [12, 602, 603]

$$\xi_h^L(r) \overset{\nu_c \gg 1}{=} \exp\left(\frac{\nu_c^2}{\sigma^2(R)} \xi_{\text{L},R}(r) \right) - 1 \approx \left(\frac{\nu_c}{\sigma(R)} \right)^2 \xi_{\text{L},R}(r) \equiv (b_1^L)^2 \xi_{\text{L},R}(r) , \tag{7.62}$$

where the linear Lagrangian bias in the high-peak limit is $b_1^L \approx \nu_c / \sigma(R)$. In order to calculate the non-Gaussian bias in the same high-peak regime, Ref. [536] considered the extension of Eq. (7.62) to non-Gaussian initial conditions derived in [525, 526],

$$\xi_h^L(r) \overset{\nu_c \gg 1}{=} -1 + \exp\left\{ \sum_{N=2}^{\infty} \sum_{j=1}^{N-1} \frac{\nu_c^N \sigma^{-N}(R)}{j!(N-j)!} \xi_{\text{L},R}^{(N)} \left[\boldsymbol{q}_1, \cdots, \boldsymbol{q}_1, \boldsymbol{q}_2, \cdots, \boldsymbol{q}_2 \right] \right\} . \tag{7.63}$$

Here,

$$\xi_{\text{L},R}^{(N)}(\boldsymbol{q}_1, \cdots \boldsymbol{q}_N) \equiv \langle \delta_R^{(1)}(\boldsymbol{q}_1) \cdots \delta_R^{(1)}(\boldsymbol{q}_N) \rangle_c \tag{7.64}$$

are the connected N-point functions of the initial, linearly extrapolated density field smoothed on the scale R. Assuming that the primordial three-point function dominates all other $(N > 3)$-point functions in Eq. (7.63), the non-Gaussian correction to the two-point correlation function of thresholded regions reads, at linear order in $\xi_{\text{L},R}^{(3)}$,

$$\Delta \xi_h^L(r) = \frac{\nu_c^3}{2\sigma^3(R)} \left[\xi_{\text{L},R}^{(3)}(\boldsymbol{q}_1, \boldsymbol{q}_1, \boldsymbol{q}_2) + \xi_{\text{L},R}^{(3)}(\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_2) \right] . \tag{7.65}$$

In Fourier space, this relation can be written as follows:

$$P_h^L(k) = [b_1 + 2b_1 \Delta b_1(k)] P_{\text{L}}(k) , \tag{7.66}$$

where the scale-dependent bias is related to the primordial three-point function through [536]

$$\Delta b_1(k) = 2\nu_c^2 \mathcal{F}_R^{(3)}(k) \mathcal{M}_R^{-1}(k)$$

$$\text{where} \quad \mathcal{F}_R^{(3)}(k) \equiv \frac{1}{4\sigma^2(R)P_\phi(k)} \int_{\boldsymbol{k}_1} \mathcal{M}_R(k_1)\mathcal{M}_R(p)B_\phi(k_1, |\boldsymbol{k} - \boldsymbol{k}_1|, k) \ . \tag{7.67}$$

Before discussing this result which, again, is only valid in the high-peak limit, let us illustrate how the non-Gaussian bias can be derived from the perturbative bias expansion corresponding to thresholding. The latter takes the familiar LIMD form

$$\delta_{\text{thr}}^L(\boldsymbol{q}) = b_1^L [\delta_R](\boldsymbol{q}) + \frac{1}{2}b_2^L [\delta_R^2](\boldsymbol{q}) + \dots \ , \tag{7.68}$$

where, for Gaussian initial conditions, the renormalized bias parameters b_N^L and operators $[\delta_R^N(\boldsymbol{q})]$ are related to orthogonal polynomials as discussed in Sec. 6. For the simple thresholding models, these are Hermite polynomials as in Eq. (2.102) [99], and the b_N^L are given by [see Eq. (2.8)]

$$b_N^L = \sqrt{\frac{2}{\pi}} \left[\text{erfc}\left(\frac{\nu_c}{\sqrt{2}}\right) \right]^{-1} \frac{e^{-\nu_c^2/2}}{\sigma^N(R)} H_{N-1}(\nu_c) \overset{\nu_c \gg 1}{\cong} \frac{\nu_c^N}{\sigma^N(R)} \ . \tag{7.69}$$

The leading-order contribution induced by a primordial three-point function is

$$\Delta \xi_h^L(r) = b_1^L b_2^L \left\langle [\delta_R](\boldsymbol{q}_1) [\delta_R^2](\boldsymbol{q}_2) \right\rangle = b_1^L b_2^L \left\langle \delta_R(\boldsymbol{q}_1)\delta_R^2(\boldsymbol{q}_2) \right\rangle \ . \tag{7.70}$$

Transforming to Fourier space and taking the high-peak limit, we recover Eq. (7.67). Let us now briefly discuss this result:

- One could identify ν_c^2 in Eq. (7.67) with $b_1^L \delta_{\text{cr}}$ and, thus, recover the prediction for a universal mass function, Eq. (7.58). However, Eq. (7.70) shows that the actual prediction of thresholding is $b_\phi \propto b_2^L$, which is clearly at odds with measurements from simulations. In fact, any Lagrangian LIMD bias expansion yields $b_\phi \propto b_2^L$. This point was first made by [604, 605, 243]. Note that, here, this contribution is not absorbed by a counter-term $b_\phi \phi$ because the bias parameters are defined as coefficients of powers of the density smoothed on a *physical* scale R, rather than the artificial cutoff Λ used in Sec. 7.1.2. This is a consequence of using a microscopic Lagrangian ansatz such as Eq. (7.61), rather than an effective large-scale expansion as in Sec. 7.1.

- Eq. (7.67) generically applies to any primordial three-point function, i.e. it is not restricted to the case of local PNG. However, as will be shown shortly, for nonlocal PNG it misses an important contribution to the scale-dependent bias.

- This derivation does not rely on expanding the primordial bispectrum in the squeezed limit. This is a consequence of the precise LIMD ansatz Eq. (7.61), which does not restrict us to scales $k \ll 1/R$. One can think of this as resumming the beyond-squeezed-limit contributions discussed in Sec. 7.1.5, by making an assumption about how precisely the Lagrangian halo number density depends on the statistics of the small-scale modes. The accuracy of this approach depends on whether this dependence is well approximated by Eq. (7.61). Indeed, we will see below that the prediction Eq. (7.67) changes if the Lagrangian bias model includes higher-derivative operators (Sec. 7.4.3). Nevertheless, Eq. (7.67) does match the results of Sec. 7.1.5 at next-to-leading order in the squeezed-limit expansion when evaluated for halos following a universal mass function.

7.4.2 Excursion-set approach

Clearly, a thresholding ansatz of the type Eq. (7.61) is not sufficient for a realistic description of halo bias (see Sec. 5). One should instead also include the first-crossing constraint, which leads to a correction to the scale-dependent bias. The correction is significant for all types of PNG beyond local quadratic PNG [98, 600].

This correction term comes about because the scale-dependent coupling between long- and short-wavelength modes in the case of general PNG modifies the conversion from an interval in $\ln \nu_c$ to an interval in $\ln M$, as shown by Eq. (7.54). A general relation for the scale-dependent non-Gaussian bias can be obtained by considering the conditional mass function $\overline{n}_h(M|\delta_\ell)$, where the long mode δ_ℓ correlates with the small-scale density fluctuations owing to PNG [see Eq. (7.15) and Eq. (7.25)] [98, 600].

Namely, in the standard excursion-set approach, the halo mass function conditioned on a long mode δ_ℓ on scale R_ℓ reads

$$\overline{n}_h(M|\delta_\ell) = -2\overline{\rho}_m \frac{\partial}{\partial M} \int_{\delta_{\rm cr}}^\infty d\delta \, p\big(\delta; R \big| \delta_\ell; R_\ell\big) \,, \tag{7.71}$$

where now the conditional probability $p\big(\delta; R\big|\delta_\ell; R_\ell\big)$ for having a small-scale overdensity δ on scale R given a large scale overdensity δ_ℓ on scale $R_\ell \gg R$ is no longer Gaussian. Expanding $p\big(\delta; R\big|\delta_\ell; R_\ell\big)$ in a series of reduced cumulants, computing

$$\delta_h(\delta_\ell) = \frac{\overline{n}_h(M|\delta_\ell)}{\overline{n}_h(M)} - 1 \tag{7.72}$$

as in Sec. 5 and identifying the coefficients of the cumulant series with the Gaussian bias parameters, we can eventually read off the scale-dependent bias $\Delta b_1(k)$ as [98]

$$\Delta b_1(k) = \sum_{N=3}^\infty \frac{4}{(N-1)!} \left\{ b_{N-2}^L \, \delta_{\rm cr} + b_{N-3}^L \left[N - 3 + \frac{\partial \ln \mathcal{F}_R^{(N)}(k)}{\partial \ln \sigma(R)} \right] \right\} \mathcal{F}_R^{(N)}(k) \mathcal{M}_R^{-1}(k) \,, \tag{7.73}$$

where b_N^L are the Lagrangian halo bias parameters, with $b_0^L \equiv 1$. No assumption about the shape of the smoothing filter has been made to derive this result. Furthermore, contributions from all connected primordial N-point functions can be included by defining the general shape factor

$$\mathcal{F}_R^{(N)}(k) \equiv \frac{1}{4\sigma^2(R)P_\phi(k)} \left\{ \prod_{i=1}^{N-1} \int_{\boldsymbol{k}_i} \mathcal{M}_R(k_i) \right\} \langle \phi(\boldsymbol{k}_1) \cdots \phi(\boldsymbol{k}_{N-1}) \phi(\boldsymbol{k}) \rangle_c \,. \tag{7.74}$$

In the squeezed limit, $\mathcal{F}_R^{(N)}$ corresponds to taking one of the N momenta of the N-point function on the right-hand side to zero and integrating over the remaining $N-1$ momenta (with the total momentum constraint enforced). In the case of a primordial bispectrum ($N=3$), we obtain the modified variance given in Eq. (7.67). Let us emphasize two points here:

- In general, a primordial n-point function will contribute a modified $(n-1)$-cumulant [243]. Therefore, when including higher-order ($n > 3$) primordial n-point functions, it is important to include all terms that are of similar order as predicted by a given inflationary model. For example, when including terms of order $g_{\rm NL}$ ($N = 4$) in higher-order local PNG, one should also include terms of order $f_{\rm NL}^2$ which are generally comparable. Note that Eq. (7.70) does not include the stochastic contributions that are induced by an enhanced collapsed limit of the primordial trispectrum (Sec. 7.1.4).

- As Eq. (7.70), Eq. (7.73) exactly recovers the results of [536, 606, 542, 607] in the high-peak limit $\nu_c \gg 1$ where $b_{N-2} \approx (\nu_c/\sigma(R))^{N-2}$. The key new term in Eq. (7.73) is the contribution from $\partial \ln \mathcal{F}_R^{(N)}(k)/\partial \ln \sigma(R)$, which arises due to the mass (R-)dependence of the reduced cumulants. We have already obtained this contribution in the PBS approach of Sec. 7.3, where it corresponds to the change in the Jacobian J under a scale-dependent rescaling of the initial conditions [Eq. (7.54)]. In fact, evaluating Eq. (7.73) for $N = 3$ in the large-scale limit, corresponding to the bispectrum in the squeezed limit, one exactly recovers the result Eq. (7.55) presented in Sec. 7.3 [100].

Finally, for quadratic local PNG, the first term in the curly bracket of Eq. (7.73) (including the multiplicative factor of 2 in front) recovers $b_\phi = 2\delta_{\rm cr} b_1^L$, in agreement with the prediction Eq. (7.58) for a universal mass function (which is the case of Press-Schechter, see Sec. 5.2.2). Therefore, it appears that we have fixed the problem of Eq. (7.70), which predicts a non-Gaussian bias amplitude $\propto b_2^L$.

To understand this better, let us now derive Eq. (7.73) from the perturbative bias expansion corresponding to the thresholding ansatz with first-crossing. The corresponding number density is [e.g., 495]

$$n_{\rm PS}(\boldsymbol{q}) = -2\overline{\rho}_m \frac{\partial}{\partial M} \Theta_H\left[\delta_R(\boldsymbol{q}) - \delta_{\rm cr}\right] = 2\overline{\rho}_m \mu_R(\boldsymbol{q})\delta_D\left[\delta_R(\boldsymbol{q}) - \delta_{\rm cr}\right]\frac{dR}{dM}\,, \tag{7.75}$$

where, following [430], we have performed the derivative with respect to M in the second equality and introduced the variable $\mu_R \equiv -d\delta_R/dR$ as in Sec. 6.8. Therefore, the density contrast $\delta_{\rm PS}(\boldsymbol{q})$ is described by a bivariate perturbative expansion, whose renormalized operators $[\delta^i \mu_R^j](\boldsymbol{q})$ are, in case of Gaussian initial conditions, bivariate Hermite polynomials. Namely,

$$\begin{aligned}\delta_{\rm PS}^L(\boldsymbol{q}) = {}& b_{10}^L H_{10}(\delta_R, \mu_R) + b_{01}^L H_{01}(\delta_R, \mu_R) \\ & + \frac{1}{2}b_{20}^L H_{20}(\delta_R, \mu_R) + b_{11}^L H_{11}(\delta_R, \mu_R) + \frac{1}{2}b_{02}^L H_{02}(\delta_R, \mu_R) + \cdots.\end{aligned} \tag{7.76}$$

The corresponding bias parameters b_{ij}^L are ensemble averages of bivariate Hermite polynomials as in Sec. 6.6.1. For instance, we have [430]

$$b_{10}^L = \frac{1}{\sigma(R)}\left(\nu_c - \frac{1}{\nu_c}\right)\,, \qquad b_{01}^L = -\frac{1}{\nu_c}\left(\frac{d\sigma(R)}{dR}\right)^{-1} \tag{7.77}$$

$$b_{20}^L = \frac{\nu_c^2 - 3}{\sigma^2(R)}\,, \qquad b_{11}^L = -\frac{1}{\sigma(R)}\left(\frac{d\sigma(R)}{dR}\right)^{-1}\,, \qquad b_{02}^L = 0\,. \tag{7.78}$$

Consequently, the leading-order non-Gaussian contribution to the 2-point correlation function of the tracers is, for a primordial bispectrum,

$$\Delta\xi_h^L(r) = b_{10}^L\left[b_{20}^L\left\langle\delta_R(\boldsymbol{q}_1)\delta_R^2(\boldsymbol{q}_2)\right\rangle + 2b_{11}^L\left\langle\delta_R(\boldsymbol{q}_1)\delta_R(\boldsymbol{q}_2)\mu_R(\boldsymbol{q}_2)\right\rangle + b_{02}^L\left\langle\delta_R(\boldsymbol{q}_1)\mu_R^2(\boldsymbol{q}_2)\right\rangle\right]\,, \tag{7.79}$$

in the large-scale limit $r \to \infty$, so that we can ignore b_{01}^L, which amounts to a higher-derivative term $\propto k^2$ as discussed in Sec. 5.7. Taking the Fourier transform of Eq. (7.79), we eventually find that, in the limit $k \to 0$, the non-Gaussian bias amplitude induced by local quadratic PNG is given by [495, 430]

$$b_\phi = 2\int_{\boldsymbol{k}} c_2^L(\boldsymbol{k}, -\boldsymbol{k})P_{\rm L}(k)\,, \tag{7.80}$$

where

$$c_2^L(\boldsymbol{k}_1, \boldsymbol{k}_2) = b_{20}^L W_R(k_1)W_R(k_2) + \frac{1}{\sigma^2(R)}\frac{\partial}{\partial\ln\sigma(R)}\Big(W_R(k_1)W_R(k_2)\Big) \tag{7.81}$$

for the particular model Eq. (7.75) considered here. However, Eq. (7.80) is valid for any microscopic Lagrangian bias model. The term proportional to $\partial/\partial\ln\sigma(R)$ in Eq. (7.81) generates the new contribution, $\propto \partial\ln\mathcal{F}_R^{(N)}(k)/\partial\ln\sigma(R)$ in Eq. (7.73), for a generic primordial bispectrum. Substituting Eq. (7.81) into Eq. (7.80), we obtain

$$b_\phi = 2\left(b_{20}^L\sigma^2(R) + 2\right) = 2\delta_{\rm cr}b_{10}^L\,, \tag{7.82}$$

the prediction Eq. (7.58) for a universal mass function, which is expected since the Press-Schechter mass function is universal.

Finally, one can show [558] via the approach described in Sec. 7.1.5 that the agreement between the excursion-set result Eq. (7.73) and the general bias expansion even holds at subleading order in the squeezed limit when adopting a universal mass function $\propto f(\nu_c)$, consistent with Eq. (7.61). Fundamentally, this is a consequence of the fact that in the thresholding approach, as in general for universal mass functions, there is only a single scale R that enters all predictions of halo statistics.

7.4.3 Lagrangian density peaks

We now turn to the peak approach, which is an example of a multivariate Lagrangian bias model. We will focus on BBKS [13] peaks, rather than the excursion-set peaks (ESP) here for simplicity. BBKS peaks are described in detail in Sec. 6.2–6.6, and we will adopt the same notation here. Like in the simple excursion-set approach considered in the previous section, the effective peak perturbative bias expansion, Eq. (6.50) in Sec. 6.6.2, in terms of the variables $\nu \propto \delta$, $J_1 \propto \nabla^2\delta$, $\eta^2 \propto (\partial_i\delta)^2$, and J_2, J_3 which are invariants constructed from $\partial_i\partial_j\delta$, holds regardless of the statistical properties of the linear density field. In complete analogy with Sec. 7.4.1–7.4.2, the leading-order contribution to the non-Gaussian 2-point correlation of BBKS peaks induced by a primordial bispectrum reads

$$\Delta\xi_{\rm pk}^L(r) = 2\sigma_0^3 b_{10}^L b_{20}^L \langle \nu^2(\boldsymbol{q}_1)\nu(\boldsymbol{q}_2)\rangle + 4\sigma_0^2\sigma_2 b_{10}^L b_{11}^L \langle \nu(\boldsymbol{q}_1)J_1(\boldsymbol{q}_1)\nu(\boldsymbol{q}_2)\rangle + 2\sigma_0\sigma_2^2 b_{10}^L b_{02}^L \langle J_1^2(\boldsymbol{q}_1)\nu(\boldsymbol{q}_2)\rangle$$
$$+ 4\sigma_0\sigma_1^2\chi_1^L b_{10}^L \langle \eta^2(\boldsymbol{q}_1)\nu(\boldsymbol{q}_2)\rangle + 4\sigma_0\sigma_2^2\omega_{10}^L b_{10}^L \langle J_2(\boldsymbol{q}_1)\nu(\boldsymbol{q}_2)\rangle$$
$$+ \left(\sigma_0, b_{10}^L, \nu(\boldsymbol{q}_2)\right) \to \left(\sigma_2, b_{01}^L, J_1(\boldsymbol{q}_2)\right), \tag{7.83}$$

where the symbols are defined in Tab. 10 on p. 137, and, as in Eq. (6.50), all moments are evaluated on the scale R. The various Lagrangian, second-order peak bias parameters are given in Sec. 6.6.1 (see also Fig. 29). The 5 additional terms in the last line are obtained upon replacing each occurrence of $(\sigma_0, b_{10}^L, \nu(\boldsymbol{q}_2))$ by $(\sigma_2, b_{01}^L, J_1(\boldsymbol{q}_2))$. Restricting to local PNG, the leading-order non-Gaussian contribution to the Lagrangian peak power spectrum in the low-k limit reads [430]

$$\Delta P_{\rm pk}^L(k) = 2f_{\rm NL}\mathcal{M}_R^{-1}(k)b_\phi \left(b_{10}^L + b_{01}^L k^2\right) W_R^2(k)P_{\rm L}(k), \tag{7.84}$$

where

$$b_\phi = 2\left(\sigma_0^2 b_{20}^L + 2\sigma_1^2 b_{11}^L + \sigma_2^2 b_{02}^L + 2\sigma_1^2\chi_1^L + 2\sigma_2^2\omega_{10}^L\right). \tag{7.85}$$

Note that, with $c_2^L(\boldsymbol{k}_1, \boldsymbol{k}_2)$ given by Eq. (6.56) in Sec. 6.6.2 for BBKS peaks, b_ϕ can also be computed from Eq. (7.80). Density peaks follow a universal mass function, albeit of a generalized form, since their abundance depends on several spectral moments $\sigma_i(R)$. Thus, we would not expect them to obey Eq. (7.58). However, the general PBS prediction Eq. (7.51) should apply. For BBKS peaks, the halo mass function $\overline{n}_{\rm pk}(M)$ is given by Eq. (6.67), with $f_{\rm pk}$ being a function of both ν_c and the root-mean-square values σ_i. One can show that

$$b_\phi = \frac{\partial\ln\overline{n}_{\rm pk}}{\partial\ln\varepsilon} = 2\sum_{i=0}^{2}\frac{\partial\ln\overline{n}_{\rm pk}}{\partial\ln\sigma_i} \tag{7.86}$$

indeed yields Eq. (7.85) [430]. The physical interpretation is the same as in Sec. 7.3: in the presence of local quadratic PNG, a long-wavelength background perturbation of wavenumber k_ℓ rescales the amplitude of the power spectrum $P_{\rm L}(k_S)$ of the linear density field in a scale-independent manner,

$$P_{\rm L}(k_S) \ \to \ (1+4\epsilon)\,P_{\rm L}(k_S). \tag{7.87}$$

Thus, for local quadratic PNG, all the spectral moments are rescaled proportionally, $\sigma_i \to (1+2\epsilon)\sigma_i$, so that the parameters γ_1 and $R_\star$ remain unchanged. On setting $\epsilon \equiv f_{\rm NL}\mathcal{M}^{-1}(k)$, we recover the full non-Gaussian k-dependent correction to the linear halo bias. For a generic PNG, ϵ should be replaced by $a_0(k_\ell/k_S)^\alpha\mathcal{M}^{-1}(k)$, where α is the scaling of the leading contribution to the primordial bispectrum in the squeezed limit [see Eq. (7.37) and Eq. (7.25)]. In this case, the different moments that $\overline{n}_{\rm pk}$ depends on scale differently in Eq. (7.86), and the amplitude of the non-Gaussian bias of peaks differs from that obtained for a universal mass function, Eq. (7.55).

Note that the equivalence of b_ϕ computed from Eq. (7.86) on the one hand, and Eq. (7.85) on the other, only holds for a deterministic barrier [430]. For a fuzzy moving barrier with a phenomenological description of the scatter (cf. Sec. 5.9) as in current excursion-set peak implementations (cf. Sec. 6.8), b_ϕ does not consistently recover $\partial\ln\overline{n}/\partial\ln\varepsilon$ [608]. As shown in Sec. 7.5, this is at odds with measurements from N-body simulations.

This problem can be solved with a "microscopic" model of the scatter in the collapse barrier B (see Sec. 5.9), such as the one proposed in [326] for instance. In this case, the renormalized Lagrangian bias functions c_n^L, Eq. (6.57), acquire an explicit dependence on the shear K_{ij}. It can then be shown that, since the collapse barrier B now is independent of σ_8, the non-Gaussian bias amplitude b_ϕ, Eq. (7.85), recovers the peak-background split expectation $\partial \ln \overline{n}/\partial \ln \epsilon$ [513]. In practice, if all the scatter in the barrier B arises from the tidal shear, then b_ϕ includes two additional contributions proportional to $b_{K_2}^L$ and $b_{Q_2}^L$, where K_2 is defined in Eq. (5.101), and $Q_2 \propto \mathrm{tr}\left(K^{ij}\zeta_{ij}\right)$. Therefore, the consistency relation Eq. (7.51) can be satisfied without violating the constraints on $b_{K_2}^L$ (see Sec. 4.5), provided that $b_{Q_2}^L$ is non-zero.

To summarize the results of Sec. 7.4.1–7.4.3, in the "microscopic" Lagrangian bias framework, the fundamental reason why the standard Lagrangian LIMD model $\delta_h(\boldsymbol{x}) = b_1^L[\delta](\boldsymbol{q}) + b_2^L[\delta^2](\boldsymbol{q})/2 + \cdots$ fails at reproducing the correct amplitude of the scale-dependent bias induced by PNG is the fact that, unlike the discrete density peaks and the thresholded regions considered here, it does not include variables other than the density, especially the filter derivative $\mu = -d\delta_R/dR$ which ensures the first-crossing condition. In contrast to the approach described in Sec. 7.1, the perturbative Lagrangian bias expansions considered here remain identical regardless of the statistical properties of the initial conditions, since the proto-halo number density is constructed directly from the Lagrangian matter density field using a model that is assumed valid on all scales. The coefficient b_ϕ of the scale-dependent contribution from PNG is given by a sum of generalized second-order bias parameters, which must ensure that the PBS scaling Eq. (7.51) is recovered. This condition is not trivially enforced for a microscopic Lagrangian bias model such as ESP where the scatter in collapse threshold is only modeled in a phenomenological way, so that the model does not provide a complete description of the local relation between proto-halo density and the linear density field. As an alternative to ESP peaks in which the halo mass function is predicted from first principles, one could generalize the Press-Schechter number density Eq. (7.75) to [495]

$$n_{\Sigma\mathrm{PS}}(\boldsymbol{q}) = -2\overline{\rho}_m \frac{\partial}{\partial M}\Sigma\big[\delta_R(\boldsymbol{q}) - \delta_{\mathrm{cr}}\big] \,. \tag{7.88}$$

The free function Σ can be chosen such that the predicted halo mass function reproduces the simulated one [494]. Here again, one obtains a bivariate bias expansion b_{ij}^L, reflecting the dependence on both δ_R and μ_R. In this model, the quantity $(\partial_i\delta)^2$ does not appear in the bias expansion, due to the absence of the peak constraint. For this reason, the renormalized bias functions $c_n^L(\boldsymbol{k}_1, \ldots, \boldsymbol{k}_n)$ predicted by this model depend only on the wavenumbers $k_i = |\boldsymbol{k}_i|$. Therefore, a measurement of the scale- and configuration-dependence of the halo bispectrum would help distinguish between the predictions of different Lagrangian bias schemes.

7.5 Non-Gaussian halo bias in simulations

A number of studies have tested the theoretical prediction for the scale-dependent bias induced by local quadratic PNG against the result of large N-body simulations, including [68, 609, 610, 611, 538, 612, 546, 315] (see Fig. 32). For the purpose of an accurate comparison between the prediction Eqs. (7.43)–(7.44) and simulations, one should also include two subtle corrections [609, 538]. They arise because both the mean halo number density $\overline{n}_h(M, \tau)$ and the matter power spectrum $P_L(k, \tau)$ are modified by PNG. Firstly, the change in the mean number density (mass function) of halos in the presence of PNG induces a (scale-independent) shift $b_1 \rightarrow b_1(f_{\mathrm{NL}})$ of the linear bias, which is negative for massive halos if $f_{\mathrm{NL}} > 0$, as massive halos become more common for $f_{\mathrm{NL}} > 0$ compared to the $f_{\mathrm{NL}} = 0$ case, and hence less biased. Secondly, one should also take into account the loop contribution $\propto f_{\mathrm{NL}}$ to the matter power spectrum [613] if measurements extend into the mildly nonlinear regime.

The most recent results are from [599], who measured the scale-dependent bias $\Delta b_1(k) = f_{\mathrm{NL}} b_\phi \mathcal{M}^{-1}(k)$ for local PNG with $f_{\mathrm{NL}} = 250$ for different halo finders, taking into account the corrections discussed above. Fig. 32 shows their measurement for SO and FoF halos (see Appendix D). The solid lines are the prediction of the general PBS from Eq. (7.51), which was implemented as in [315] by performing simulations with different initial power spectrum amplitudes $\mathcal{A}_s$, and taking the derivative of the measured halo mass function with respect to $\ln \mathcal{A}_s$. The result is in excellent agreement with the data for all mass bins and halo finders.
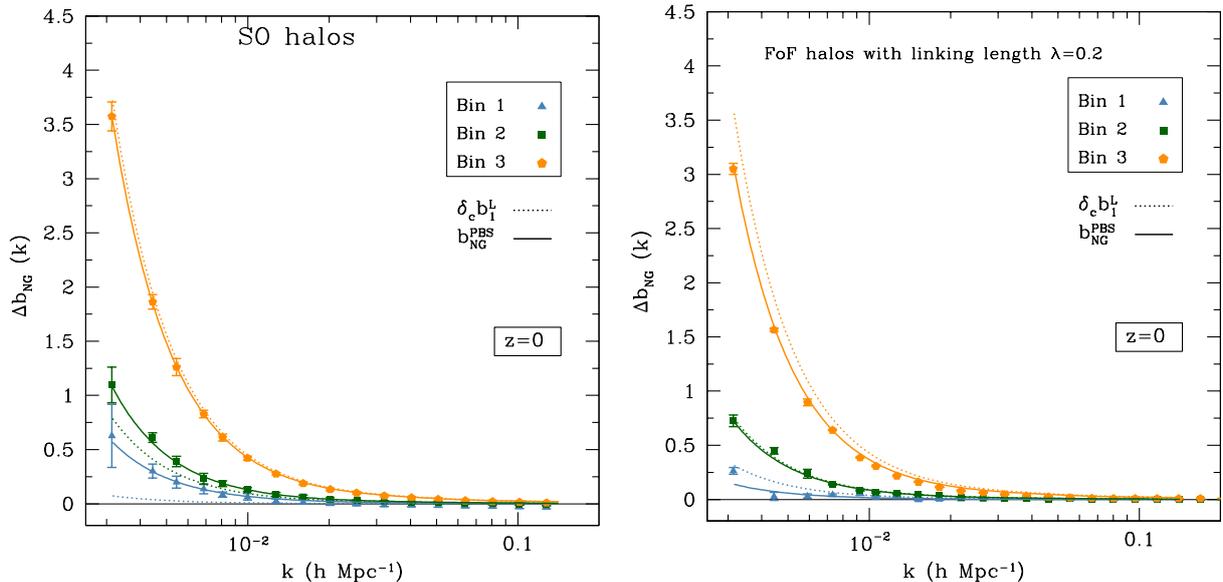
Figure 32: Scale-dependent bias $\Delta b_1(k) \simeq P_{hm}(k)/P_{mm}(k) - b_1$ measured for dark matter halos in N-body simulations with local-type PNG with $f_{\mathrm{NL}} = \pm 250$ (points). The *left panel* shows the result for halos identified using the spherical overdensity method ($\Delta_{\mathrm{SO}} = 200$), while the *right panel* presents measurements for friends-of-friends halos identified with a linking length $\lambda = 0.2$. The solid lines show the general peak-background split prediction Eq. (7.51), while the dotted lines represent the PBS evaluated for universal mass functions, Eq. (7.58), using the linear bias b_1 measured for the same mass bins in simulations with $f_{\mathrm{NL}} = 0$. The mass bins, in units of $10^{13}\,h^{-1}\,M_\odot$, are given by Bin 1: $[0.9, 1.4]$; Bin 2: $[1.4, 2.7]$; Bin 3: $[2.7, \infty)$. From [599].

By contrast, the dotted curves, which assume $b_\phi = 2\delta_{\mathrm{cr}} b_1^L$, with $\delta_{\mathrm{cr}} = 1.687$, as obtained for a universal mass function [Eq. (7.58)], are not in good agreement with the simulation results. This is more apparent in Fig. 33, which displays the relative deviation of b_ϕ, measured using Eq. (7.51), from the universal mass function prediction. Clearly, the universal mass function result overpredicts b_ϕ for rare halos with $b_1^L \gtrsim 1$ for both halo finders. For SO halos, $2\delta_{\mathrm{cr}} b_1^L$ leads to an underestimate at lower masses, in the range $0 \lesssim b_1^L \lesssim 0.5$. These results confirm the previous findings of [609, 614] who studied SO halos, and [611, 610, 546, 315] who used FoF halos. Ref. [546] express the derivative of $\overline{n}_h$ with respect to σ_8 as a derivative of the mass function with respect to halo mass. While this is motivated by the excursion-set formalism, the exact PBS prediction for b_ϕ strictly corresponds to a derivative with respect to the primordial amplitude of fluctuations, as we have seen.

At high mass, the discrepancy can be interpreted as due to the particular choice of $\delta_{\mathrm{cr}} = 1.687$, a value motivated by the spherical collapse approximation (see Sec. 5.2.1). For instance, there is some line of evidence that FoF halos with a linking length $\lambda = 0.2$ correspond to smoothed Lagrangian overdensities of less than 1.687 [e.g., 615], which would explain why $2\delta_{\mathrm{cr}} b_1^L$ with $\delta_{\mathrm{cr}} = 1.687$ overestimates b_ϕ at high mass. After all, as explained in Sec. 5.2.1, there is some freedom in the choice of δ_{cr} because actual dark matter halos never precisely correspond to the collapse of an isolated spherical perturbation. Therefore, in order to account for the dependence on the halo finding algorithm, it is sensible to make the replacement

$$\delta_{\mathrm{cr}} \to \sqrt{q}\,\delta_{\mathrm{cr}}\,, \tag{7.89}$$

where the parameter q encodes information about the departure between the halos considered and the spherical collapse approximation. Refs. [609, 610, 611, 616] have used different halo finding prescriptions and, therefore, found different values of q, as reviewed in [545]. Fig. 33 shows that, for FoF halos with linking length $\lambda = 0.2$, $\sqrt{q} \equiv r_{\mathrm{FoF}} \approx 0.85$ whereas, for the SO halos, $\sqrt{q} \equiv r_{\mathrm{SO}} \approx 0.9$.

There are two further important points to take away from Figs. 32–33 :

- The departure from $2\delta_{\mathrm{cr}} b_1^L$ observed for SO halos at low mass cannot be absorbed by a change in
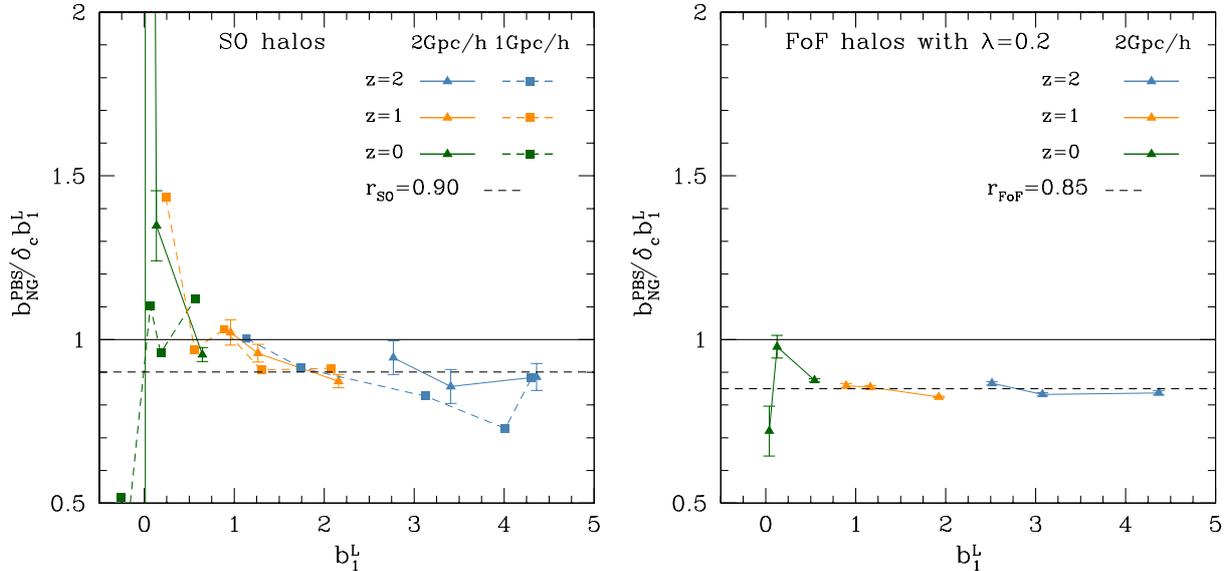
Figure 33: Measured amplitude of the scale-dependent bias of halos $b_{NG} \equiv b_\phi/2$ [through Eq. (7.51)], divided by the universal mass function prediction $b_1^L \delta_{cr}$ [Eq. (7.58)], shown as function of b_1^L. The results are from different output redshifts and different simulation box sizes, as labeled. The *left panel* displays the measurements for halos identified using the spherical overdensity method ($\Delta_{SO} = 200$), while the *right panel* shows results for friends-of-friends halos with linking length $\lambda = 0.2$. From [599].

the overdensity criterion used in the definition of SO halo masses (here, $\Delta_{SO} = 200$ with respect to matter). This is because such a change would affect the results even more strongly at high mass, where the mass function is steep. Therefore, the departure from universality observed here is unrelated to the effect discussed in [313], which is induced by their particular choice of $\delta_{cr}(z)$ as pointed out by [617, 618]. Another possible explanation is the failure of the spherical collapse approximation at low mass, which we have assumed to compute $2\delta_{cr}b_1^L$. In this case, one would replace the critical threshold δ_{cr} for spherical collapse via Eq. (7.89) by, for example, the corresponding value in the ellipsoidal collapse [619]. Clearly, this can only explain part of the deviation, since we see significant evidence that b_ϕ changes sign at a different mass than that corresponding to $b_1^L = 0$, which cannot be explained by a change of δ_{cr}.

- The non-Gaussian bias prediction of current excursion-set peak (ESP) implementations is inconsistent with the simulation data. In this approach (see Sec. 6.8 and Sec. 7.4.3 for details), the amplitude of the non-Gaussian bias is a weighted sum of all the second-order bias parameters [430]. This generally holds for any "microscopic" Lagrangian bias models [495], in contrast to models which perform a large-scale bias expansion (Sec. 7.1; [100]). However, while, when adopting a deterministic barrier, the ESP approach predicts a value of b_ϕ that matches the PBS prediction Eq. (7.51) [430], in agreement with the data, the stochastic barrier of [396] yields a value of b_ϕ that is greater than that obtained from Eq. (7.51) [608], which is clearly ruled out by the measurements in Fig. 33. To remedy this issue, the scatter around the mean barrier must be described at a "microscopic" level [513], through the tidal shear K_{ij} for instance, rather than through a white noise term as done in [303, 608].

Finally, Eq. (7.58) assumes that the clustering of halos is entirely specified by the halo mass M. This may, however, not be true for some observed tracers such as quasars whose activity may be triggered by mergers of halos. Ref. [539] used the excursion-set formalism described in Sec. 5 to estimate the bias correction Δb_{merger} induced by mergers, and obtained

$$\Delta b_{merger} = \delta_{cr}^{-1} . \tag{7.90}$$

187

Therefore, for a universal mass function, $\delta_{\mathrm{cr}} b_1^L$ should be replaced by $\delta_{\mathrm{cr}} b_1^L - 1$. The validity of this prediction was assessed by [620] using N-body simulations. On splitting the halos by the formation time identified using merger trees, they found a significant dependence of b_ϕ on the formation time in agreement with Eq. (7.90). Such a shift (albeit of smaller magnitude) would explain the different zero-crossing of b_ϕ and b_1^L found for SO halos (right panel of Fig. 33).

Other numerical studies have considered the effect of PNG beyond local PNG, in particular the cubic coupling $g_{\mathrm{NL}}\phi^3$ [606, 621]; the orthogonal and equilateral bispectrum templates [616, 546], and multi-field inflation [543, 622]. These analyses suggest that the non-Gaussian bias of dark matter halos is consistent with the PBS prediction for universal mass functions, Eq. (7.55) with Eq. (7.58), or equivalently Eq. (7.73). For these shapes however, it is crucial to take into account the second term on the right-hand side of Eq. (7.55) and Eq. (7.73). Here, the same caveat regarding $b_\phi = \delta_{\mathrm{cr}} b_1^L$ discussed above applies to the first term in Eq. (7.73).

Overall, these efforts to calibrate b_ϕ and b_ψ are important ingredients for the goal of constraining primordial non-Gaussianity using the scale-dependent bias of galaxies, since any systematic uncertainty on b_ϕ (e.g., of order 10–20%, Fig. 33) translates into a similar uncertainty in the constraint on f_{NL}.

7.6 Observational prospects

Although the cosmic microwave background (CMB) constraints from the Planck satellite already have put stringent constraints on various PNG shapes, in particular $f_{\mathrm{NL}} = 0.8 \pm 5$ for local quadratic PNG [580], there is still room left for interesting phenomenology at the level of $|f_{\mathrm{NL}}| \simeq 1$ that is unconstrained by current CMB limits. By making use of the scale-dependent features induced by PNG in the galaxy power spectrum and bispectrum (Sec. 7.1.6), future surveys of the large-scale structure (LSS) of the Universe, such as DESI, Euclid, LSST, and others, are one of our best hopes for improving the current CMB limits on PNG. Most of the signal-to-noise from the scale-dependent bias in the power spectrum comes from the largest scales accessible to a given survey. While the theoretical model for galaxy clustering on those scales is very robust, as linear theory is sufficient (although including projection effects is important, as described in Sec. 9.3.1), observational systematics can provide difficult obstacles. This is because the intrinsic signal is small on those scales, while the apparent galaxy clustering induced by systematic effects (for example, fluctuations of the survey depth across the sky, stellar contamination, and photometric calibration uncertainties) is largest on large scales. These effects can mimic the signature of a non-zero f_{NL} and must be carefully controlled for [623, 624, 625]. On the other hand, the signatures of PNG in the galaxy three-point function (bispectrum) are present on smaller scales, and the latter can thus provide important cross-checks.

Current limits on f_{NL} from LSS surveys, which are all based on power spectrum measurements only, are at the level of CMB pre-Planck constraints. For instance, we have (at 68% C.L.)

$$-16 < f_{\mathrm{NL}} < +26\,, \quad \text{from galaxies [626], and} \tag{7.91}$$

$$-39 < f_{\mathrm{NL}} < +23\,, \quad \text{from quasars [627]}\,. \tag{7.92}$$

Forecasts indicate that the statistical errors on f_{NL} should decrease by $1-2$ orders of magnitude with the next generation of large redshift surveys [628, 629, 630, 631, 632, 633, 634, 635]. We will discuss quantitative forecasts for constraints on local f_{NL} achievable with future surveys in the remainder of this section, focusing on the combination of power spectrum and bispectrum for a single tracer, and multi-tracer techniques. We will adopt an idealized setting as in Sec. 4.1.3, neglecting redshift-space distortions and assuming a trivial survey window function.

7.6.1 Galaxy power spectrum and bispectrum for a single tracer

In this section, we forecast constraints on local f_{NL} from future surveys using the galaxy power spectrum and bispectrum. For this, we use the Fisher information matrix as described in detail in Sec. 4.1.3 [see Eqs. (4.13)–(4.15)]. We use the same assumptions and survey parameters as in Sec. 4.1.3, and calculate the projected uncertainties under the null hypothesis $f_{\mathrm{NL}} = 0$; hence, the uncertainties of power spectrum and bispectrum measurements are still given by Eq. (4.17) and Eq. (4.19), respectively. We then expand the

parameter vector considered in Sec. 4.1.3 to include f_{NL},

$$\vec{\theta} = \left\{ b_1,\ b_2,\ b_{K^2},\ \ln\mathcal{A},\ P_\varepsilon^{\{0\}},\ P_{\varepsilon\varepsilon_\delta}^{\{0\}},\ B_\varepsilon^{\{0\}},\ f_{\mathrm{NL}}\ \right\}, \tag{7.93}$$

using the same fiducial values as given in Tab. 6. We include the effects of f_{NL} at second order in the general bias expansion, as summarized in the relations of Sec. 7.1.6, restricting to terms that are linear in f_{NL}. This is sufficient for a fiducial value of $f_{\mathrm{NL}} = 0$, since the Fisher matrix only involves first derivatives with respect to parameters, and the contribution of any higher-order term to $\partial P_{gg}/\partial f_{\mathrm{NL}}$ and $\partial B_{ggg}/\partial f_{\mathrm{NL}}$ vanishes upon evaluation at $f_{\mathrm{NL}} = 0$. Then, the power spectrum is given by

$$P_{gg}(k) = P_{gg}^{(G)}(k) + 2b_1\Delta b_1(k)P_{\mathrm{L}}(k), \tag{7.94}$$

where $P_{gg}^{(G)}(k)$ is the leading galaxy power spectrum with Gaussian initial conditions [Eq. (4.2)]. Further, using the notation of Sec. 4.1.1, the galaxy bispectrum is given by

$$
\begin{aligned}
B_{ggg}(k_1, k_2, k_3) = {}& B_{ggg}^{(G)}(k_1, k_2, k_3) \\
& + 2b_1^3 f_{\mathrm{NL}} \mathcal{M}(k_1)\mathcal{M}(k_2)\mathcal{M}(k_3) \left[P_\phi(k_1)P_\phi(k_2) + 2\text{ perm.} \right] \\
& + \Bigg\{ b_1^2 P_{\mathrm{L}}(k_1) P_{\mathrm{L}}(k_2) \\
& \qquad \times \left[\frac{(2\delta_{\mathrm{cr}} - 1)(b_1 - 1) + \delta_{\mathrm{cr}} b_2^L}{2\delta_{\mathrm{cr}}(b_1 - 1)} \left(\Delta b_1(k_1) + \Delta b_1(k_2) \right) + \mu_{12} \left\{ \frac{k_1}{k_2}\Delta b_1(k_1) + \frac{k_2}{k_1}\Delta b_1(k_2) \right\} \right] \\
& \qquad + b_1 \left[\Delta b_1(k_1) + \Delta b_1(k_2) \right] P_{\mathrm{L}}(k_1)P_{\mathrm{L}}(k_2) \left[2F_2(\boldsymbol{k}_1, \boldsymbol{k}_2)b_1 + b_2 + 2b_{K^2}\left(\mu_{12}^2 - \frac{1}{3} \right) \right] \\
& \qquad + 2P_{\varepsilon\varepsilon_\delta}^{\{0\}} \left[\Delta b_1(k_1)P_{\mathrm{L}}(k_1) + \Delta b_1(k_2)P_{\mathrm{L}}(k_2) \right] \\
& + 2\text{ perm.} \Bigg\},
\end{aligned} \tag{7.95}
$$

where $B_{ggg}^{(G)}(k_1, k_2, k_3)$ is the galaxy bispectrum with Gaussian initial conditions [Eq. (4.4)]. In both expressions, we take Δb_1 from Eq. (7.59),

$$\Delta b_1(k) = 2f_{\mathrm{NL}}\delta_{\mathrm{cr}}(b_1 - 1)\mathcal{M}^{-1}(k). \tag{7.96}$$

We have further inserted the universal mass function prediction for $b_{\phi\delta}^E$ [Eq. (7.60)], using the second-order Lagrangian bias given by Eq. (2.34), $b_2^L = b_2 - 8/21(b_1 - 1)$, and set $P_{\varepsilon\varepsilon_\phi}^{\{0\}} \approx (b_\phi/b_1)P_{\varepsilon\varepsilon_\delta}^{\{0\}}$ as predicted when stochasticity is described by Poisson shot noise. For the purposes of these idealized forecasts, these approximations are sufficient.

The galaxy bispectrum in the presence of local-type PNG has been considered in several previous references [243, 612, 538, 541, 547, 559]. Unlike the majority of these, we have restricted to linear order in f_{NL}, which, as explained above, is sufficient for a Fisher forecast with fiducial value $f_{\mathrm{NL}} = 0$. Further, given the constraints set by the Planck satellite, higher-order terms in f_{NL} are unlikely to be relevant in the future. We have also neglected the effect of PNG on the statistics of the matter density field beyond the leading-order bispectrum. This is consistent at the order of perturbations we are working in; the one-loop correction to the matter power spectrum from PNG is negligible for the wavenumbers and values of f_{NL} considered, while that to the bispectrum begins to be relevant at $k_{\max} \sim 0.2\,h\,\mathrm{Mpc}^{-1}$ [613]. On the other hand, the above-mentioned references did not include the complete set of bias parameters and stochastic amplitudes that appear in the general perturbative bias expansion. Ref. [550] forecast the constraints obtained on f_{NL} of the equilateral ($\alpha = 2$) and quasi-single-field ($0 \le \alpha \le 3/2$) types from the power spectrum alone, including the NLO correction as well as the relevant higher-derivative terms. Further, Ref. [636] forecast the detectability of higher-spin fields in the galaxy power spectrum through the anisotropic squeezed-limit
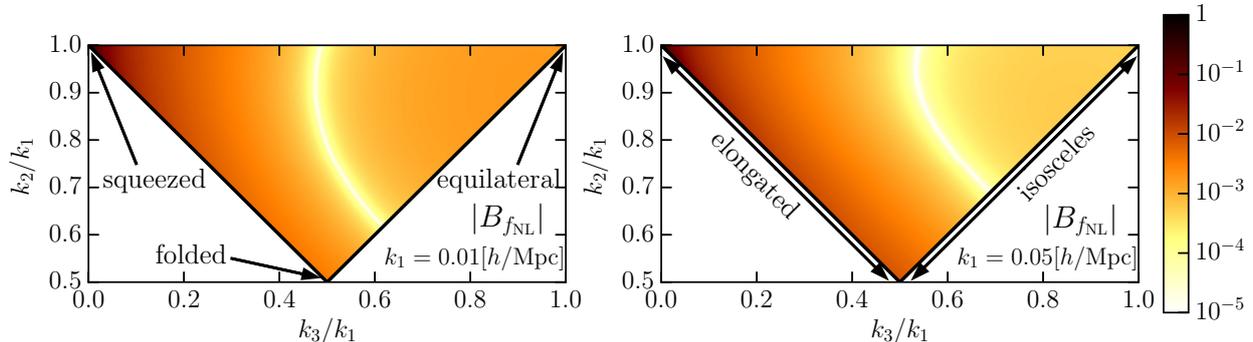
Figure 34: Shape dependence of the contribution from local-type PNG to the galaxy bispectrum, i.e. all terms except for $B_{ggg}^{(G)}$ in Eq. (7.95). We calculate the shape dependence at $z = 2$ with $f_{\text{NL}} = 5$ as fiducial value. The remaining parameters are as listed in Tab. 12. In order to highlight the shape dependence, we have divided the bispectrum by the maximum value attained in each panel. The contribution from local-type PNG is sharply peaked in the squeezed limit; this is very different from the shape dependence of the contributions present for Gaussian initial conditions which are shown in Fig. 10.

primordial bispectrum. The latter can also be constrained using galaxy shape correlations, as forecasted in [552, 554].

Fig. 34 shows the configuration dependence of the contributions to the galaxy bispectrum from PNG, in the same form as Fig. 10 in Sec. 4.1. These contributions peak for squeezed triangle configurations, showing a distinct difference to the Gaussian contributions shown in Fig. 10. For this reason, we expect the bispectrum to help significantly in improving constraints on f_{NL} from galaxy clustering.

The result of the Fisher matrix calculation is summarized in Tab. 12 for the seven galaxy surveys that we consider here: HETDEX [50], eBOSS [49], DESI [53], PFS [57], Euclid [54] and WFIRST [58]. We use the same fiducial values for the bias and stochastic parameters as in Tab. 6, and show results for $k_{\max} = 0.1 \, h \, \text{Mpc}^{-1}$ and $k_{\max} = 0.2 \, h \, \text{Mpc}^{-1}$. For all surveys, combining the galaxy power spectrum and bispectrum gives a factor of several better constraints on f_{NL} than the power spectrum alone. Moreover, when using the galaxy bispectrum, future galaxy surveys are expected to improve on the CMB constraints on f_{NL}. Comparing the different choices for $k_{\max}$, we see that, while the power-spectrum-only constraints show only incremental changes, the combined constraints improve by about a factor of two for the higher value of $k_{\max}$. This is because the signatures of PNG in the galaxy power spectrum are only prominent on large scales (see Fig. 32), while those in the galaxy bispectrum are present for triangles in the squeezed configuration on *all* scales (Fig. 34). Further, the forecasted error on f_{NL} depends sensitively on the fiducial values of the bias parameters. This is because both the signal from PNG ($\Delta b_1 \propto (b_1 - 1)$) and the signal-to-noise ratio per k-mode ($\bar{n} P_g(k) \propto b_1^2$) change favorably for highly biased tracers (when the number density is fixed). For example, we find that the projected uncertainties shown in Tab. 12 are improved by a factor of 2 to 3 when assuming $b_1 = 2$ instead of $b_1 = 1.5$. In this case, both DESI and Euclid reach $\sigma_{f_{\text{NL}}} < 1$ for $k_{\max} = 0.2 \, h/\text{Mpc}$. Note that previous forecasts in the literature [637, 638, 639, 640, 559] fixed the stochastic parameters, if included at all, as well as the primordial power spectrum amplitude to their respective fiducial values, leading to more optimistic forecasts than those shown in Tab. 12.

7.6.2 Multi-tracer methods

In the previous section, we have derived the optimal constraints from the galaxy power spectrum and bispectrum assuming a single tracer. Interestingly, a significant reduction of the statistical error on the scale-dependent bias, and thus the constraint on f_{NL}, from the galaxy power spectrum alone can in principle be achieved with multi-tracer techniques and optimal weights [641, 642, 643, 345, 644, 342]. These make use of the fact that the scale-dependent signature in the *relative* clustering of different tracers does not suffer from cosmic variance. Recently, Ref. [554] applied the same technique to galaxy shape correlations, which constrain anisotropic PNG as discussed in Sec. 7.1.3. Moreover, the variance due to shot noise of tracers

Table 12: Projected constraints on the amplitude of local-type PNG, f_{NL}, from the power spectrum and bispectrum of a single galaxy sample. The power-spectrum-only constraints are obtained after marginalizing over b_1 and $P_\varepsilon^{\{0\}}$, while the combined constraints are obtained after marginalizing over 7 parameters: { b_1, b_2, b_{K^2}, $P_\varepsilon^{\{0\}}$, $P_{\varepsilon\varepsilon_\delta}^{\{0\}}$, $B_\varepsilon^{\{0\}}$, $\ln\mathcal{A}$ }. For all cases, we use the following fiducial values: $f_{\mathrm{NL}} = 0$; $b_1 = 1.5$; $b_2(b_1)$ from the fitting formula in Tab. 7 ($b_2 \simeq -0.69$); and $b_{K^2}(b_1)$ by assuming Lagrangian LIMD bias ($b_{K^2} = -2/7(b_1 - 1) \simeq -0.14$). Further, we adopt $P_{\varepsilon\varepsilon_\delta}^{\{0\}} = b_1/(2\overline{n}_g)$ and $b_2^L = b_2 - 8/21(b_1 - 1) \simeq -0.88$ as fiducial values. Note that $\sigma_{f_{\mathrm{NL}}}$ depends sensitively on the fiducial value of b_1 (see text).

survey	redshift	Volume	$10^4\,\overline{n}_g$	$\sigma_{f_{\mathrm{NL}}}$, $k_{\max} = 0.1\,h/\mathrm{Mpc}$		$\sigma_{f_{\mathrm{NL}}}$, $k_{\max} = 0.2\,h/\mathrm{Mpc}$	
	$\bar{z}$	$[h^{-3}\mathrm{Gpc}^3]$	$[h^{-3}\mathrm{Mpc}^3]$	$P(k)$	$P(k) + B(k_1, k_2, k_3)$	$P(k)$	$P(k) + B(k_1, k_2, k_3)$
eBOSS (LRG)	0.8	6.1	4.4	19	15	18	7.5
eBOSS (QSO)	1.4	39	1.5	7.6	6.5	7.2	3.9
HETDEX	2.7	2.7	3.6	24	22	22	14
PFS	1.5	8.7	4.6	13	11	12	6.4
DESI	1.1	40	3.3	6.1	5.1	5.7	2.9
WFIRST	1.9	13	12	7.8	6.9	7.3	4.2
Euclid	1.4	63	5.2	4.0	3.4	3.8	2.0

can be reduced by applying suitable weights.

The core of this method works as follows. First, the observed galaxy distribution is split into a number of subsamples with different bias properties. These subsamples, then, can be weighted differently to take advantage of the shot-noise suppression for massive halos, which is seen in the halo clustering measured from N-body simulations (see Sec. 4.5.3). In the limit of infinitely fine binning, and vanishing minimum halo mass, a weight $\omega(M) = M$ proportional to the host halo mass minimizes the stochasticity of halos relative to the dark matter distribution and, thus, reduces the shot noise contribution [177]. Note that it is essential, though not always sufficient, that both the shot noise and cosmic variance suppression be considered at the same time, otherwise the net improvements on $\sigma_{f_{\mathrm{NL}}}$, the 68% confidence level uncertainty on f_{NL}, are small [342]. Given predictions for b_ϕ, taken from the PBS applied to universal mass functions [Eq. (7.58)] and the covariance matrix of the shot noise among the different subsamples [estimated through Eq. (4.52)], one can construct the optimal weighting function as well as estimate the projected uncertainty on f_{NL} [614], again assuming a Gaussian halo density field which is accurate on the large scales relevant for f_{NL} constraints.

Results are shown in Fig. 35 for halos at redshift $z = 1$. When the dark matter density field is available (red lines and filled symbols), weighting the halos by their mass (red dashed lines and filled circles) is always superior to the conventional uniform-weighting case (red solid lines and filled squares), especially when considering lower-mass halos. In particular, the uncertainty $\sigma_{f_{\mathrm{NL}}}$ substantially decreases with decreasing $M_{\min}$ in the weighted case, while for uniform weighting it exhibits a spike at $M_{\min} \simeq 5 \times 10^{10}\,h^{-1}\,M_\odot$. This is because the effective linear bias b_1 of the weighted halo sample is always greater than 1 for mass-weighting, but reaches 1 at the above mentioned value of $M_{\min}$ for uniform weighting. Since this implies $b_\phi \propto b_1 - 1 = 0$ assuming the PBS prediction, $\sigma_{f_{\mathrm{NL}}}$ becomes infinite. Simulation results are over-plotted as symbols for a few values of $M_{\min}$. The simulations match well with the analytical predictions based on a parametrization of the halo stochasticity covariance (see Sec. 4.5.3). The simulations yield a minimum error of $\sigma_{f_{\mathrm{NL}}} \simeq 0.8$ at $M_{\min} \simeq 10^{12}\,h^{-1}\,M_\odot$ in the optimally weighted case with the dark matter available.

The results without the dark matter are shown as blue lines and open symbols. Let us focus on a single bin first. $\sigma_{f_{\mathrm{NL}}}$ exhibits a minimum at $M_{\min} \sim 10^{13}\,h^{-1}\,M_\odot$ with $\sigma_{f_{\mathrm{NL}}} \sim 5$ for both uniform and weighted halos. Thus, in the single-tracer case, it is optimal for f_{NL} constraints to only consider the highest-mass halos. However, if the halo sample is divided into multiple bins, $\sigma_{f_{\mathrm{NL}}}$ does continue to improve when including lower-mass halos.

Comparing to the results for Euclid in Tab. 12, which has a similar volume and mean redshift to what is assumed in Fig. 35, and taking into account that the latter shows the theoretical optimum without marginalizing over other parameters, we see that the combination of power spectrum and bispectrum is
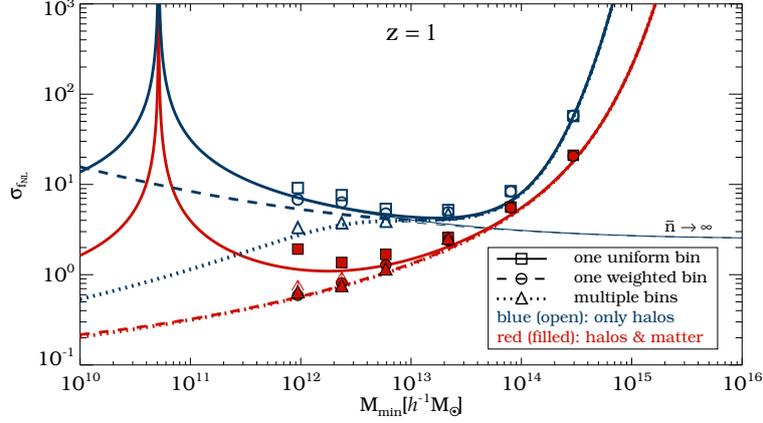
Figure 35: 68% confidence-level uncertainty on f_{NL} for a survey centered at $z = 1$ as a function of the minimum halo mass resolved by the survey. An effective survey volume $V_{eff} \simeq 50 h^{-3} Gpc^3$ (roughly comparable to Euclid) is assumed, taking into account all modes from $k_{min} = 0.0039$ to $k_{max} = 0.032 \, h \, Mpc^{-1}$. Note that the value of k_{max} is not important numerically. Solid and dashed lines show results from uniform- and mass-weighted halos of a single mass bin. The N-body measurements are overlaid as squares and circles for different minimum-mass cuts. Forecasts which assume knowledge of both halos and dark matter are plotted in red (filled symbols), whereas those based on the halo distribution only are displayed in blue (open symbols). The dotted lines (triangles) show the result of splitting the halo catalog into multiple mass bins. For the single mass bin, the high-sampling limit ($\bar{n}_h \to \infty$) is over-plotted for the uniform- (thin solid line) and the weighted case (thin dashed line). The arrows represent the effect of adding a log-normal scatter of $\sigma_{\ln M} = 0.5$ to all halo masses. They are omitted in all cases where the scatter has negligible impact. From [614].

expected to outperform the power spectrum with the multi-tracer method (assuming the matter density is not available, as assumed for Tab. 12). However, multi-tracer methods can also be applied to the bispectrum, combining the advantages of both methods and potentially leading to a further improvement in constraints on PNG.

8 Beyond cold dark matter, cosmological constant, and General Relativity

So far in this review, we have focused on a Universe filled with pressureless cold matter, and with a cosmological constant. However, the real Universe also contains a background of free-streaming neutrinos, which we now know to have nonzero masses. Further, baryons are subject to non-gravitational forces. The impact of these stress-energy components on the galaxy bias expansion is reviewed in Sec. 8.1 and Sec. 8.2, respectively. In addition, the cosmological constant could in fact be dynamical, for example caused by a slowly rolling scalar field (quintessence), or caused by modifications to General Relativity. We consider the impact of these classes of new physics in Sec. 8.3.

8.1 Massive neutrinos

Direct measurements of neutrino oscillations based on solar, atmospheric and reactor neutrinos [645] have shown that the masses of at least two of the three mass eigenstates (m_{ν_1}, m_{ν_2}, m_{ν_3}) of neutrinos must be non-zero. The experimentally constrained mass gaps are $\Delta m_{21}^2 = m_{\nu_2}^2 - m_{\nu_1}^2 \simeq (8.58\,\mathrm{meV})^2$ and $|\Delta m_{32}^2| = |m_{\nu_3}^2 - m_{\nu_2}^2| \simeq (49.6\,\mathrm{meV})^2$ [645]. Neutrinos constitute a small but non-negligible fraction of the cosmological energy budget, and their mass plays an important role in the growth of large-scale structure. The cosmic neutrino background, thermally produced during Big Bang nucleosynthesis, has a temperature of approximately $T_\nu(z) \simeq 1.95(1 + z)\,\mathrm{K} \simeq 0.168(1 + z)\,\mathrm{meV}$. Thus, at high redshifts, when $T_\nu(z) \gg m_i$, neutrinos are fully relativistic and do not contribute to the growth of structure in the CDM and baryon components. Once neutrinos transition to a non-relativistic form of matter at a redshift $z_{\mathrm{nr},i}$ of roughly

$$1 + z_{\mathrm{nr},i} \approx 1890 \left(\frac{m_{\nu_i}}{1\,\mathrm{eV}} \right) , \qquad (8.1)$$

perturbations in the neutrino density begin to grow appreciably on scales larger than their free-streaming scale, effectively forming another (though not cold) dark-matter component. The neutrino free-streaming scale, the analog of a Jeans scale for collisionless matter, is of order the distance traveled by neutrinos per Hubble time, and correspondingly shrinks as the neutrinos become non-relativistic. More precisely, it is given by [646]

$$k_{\mathrm{fs}}(m_\nu, z) = \lambda_{\mathrm{fs}}^{-1}(m_\nu, z) = \frac{0.068}{(1 + z)^2} \left(\frac{m_\nu}{0.1\,\mathrm{eV}} \right) \left[\Omega_{m0}(1 + z)^3 + \Omega_{\Lambda 0} \right]^{1/2} h\,\mathrm{Mpc}^{-1} . \qquad (8.2)$$

Within the free-streaming scale λ_{fs}, neutrino density fluctuations are damped due to their large velocity dispersion. On the other hand, on scales larger than λ_{fs}, neutrinos effectively behave as a pressureless fluid coupled gravitationally to CDM and baryons. Depending on their wavelength, modes can initially be inside the free-streaming scale and exit this scale later as λ_{fs} shrinks.

At redshifts $z \ll \min_i\{z_{\mathrm{nr},i}\}$, the fraction of the total, non-relativistic matter energy density in the form of neutrinos is

$$f_\nu \equiv \frac{\Omega_{\nu 0}}{\Omega_{c+b,0} + \Omega_{\nu 0}} \simeq \frac{\sum_i m_{\nu.i}}{93.1 \Omega_{m0} h^2\,\mathrm{eV}} . \qquad (8.3)$$

where the subscripts m and $c + b$ correspond to the total matter and the sum of CDM and baryons, respectively. We can conveniently express perturbations to the total matter density as a weighted sum of CDM+baryon and neutrino fluctuations,

$$\delta_m = (1 - f_\nu)\delta_{c+b} + f_\nu \delta_\nu , \qquad (8.4)$$

where we have assumed that CDM and baryons perfectly trace each other. This is appropriate since we are interested in large scales (but see Sec. 8.2 for the effect of relative density and velocity perturbations

between CDM and baryons). In the following, we assume that the masses of all three active neutrino species are comparable. This can easily be generalized. On scales larger than the free-streaming scale of the neutrino species, $k \ll k_{\text{fs}}(m_\nu)$, neutrino density perturbations δ_ν follow the adiabatic prediction and are proportional to δ_{c+b}, with a redshift-dependent factor (assuming there are no primordial neutrino isocurvature perturbations). On scales much smaller than the free-streaming scale, $k \gg k_{\text{fs}}(m_\nu)$, neutrino perturbations are erased, $\delta_\nu \approx 0$ so that $\delta_m \approx (1 - f_\nu)\delta_{c+b}$. In addition, the power spectrum $P_{c+b}(k)$ of the CDM+baryon component itself is also modified by massive neutrinos due to their contribution to the Poisson equation and the change in the expansion rate. As a result, the growth of CDM+baryon perturbations in matter domination is slowed down according to $\delta_{c+b} \propto a^{1-\frac{3}{5}f_\nu}$ (see [647] for a detailed review of the effect of massive neutrinos in cosmology). The following approximation [648, 649],

$$P_{\text{L}}(k) \xrightarrow{k \gg k_{\text{fs}}} (1 - 8f_\nu)P_{\text{L}}^{(m_\nu = 0)}(k) \,, \tag{8.5}$$

is a good fit to the numerical data. Numerical simulations indicate that, at redshift $z = 0$, nonlinear effects enhance the suppression, to $\Delta P_m(k)/P_m(k) \simeq -10f_\nu$ at $k \sim 1\,h\,\text{Mpc}^{-1}$, before it diminishes at smaller scales [650, 651, 652]. Thus, neutrinos leave a characteristic, scale-dependent feature in the power spectrum of total matter as well as the non-relativistic part $(c+b)$ [653, 654, 655]. Current CMB- and expansion-history-based constraints on the sum of neutrino masses are $\sum m_\nu \lesssim 0.6$ eV [97], which lead to a neutrino fraction $f_\nu \lesssim 0.05$, while the minimum value allowed by neutrino oscillation measurements is $\sum m_\nu \geq 0.06$ eV [645], corresponding to $f_\nu \gtrsim 0.005$. For a $\sum m_\nu$ in this range, the neutrino free-streaming scale is sufficiently large that the neutrino perturbations remain in the linear regime down to redshift $z \lesssim 10$, as has been shown numerically [656, 657, 658].

The effect of massive neutrinos on the clustering of halos has been studied both using semi-analytic methods (e.g. [659, 654, 655, 660, 661]) and full N-body simulations (e.g. [662, 663, 664, 665, 656, 657, 666, 667, 668, 669]). Like in the pure CDM case, analytic approaches provide useful insights into the effect of massive neutrinos on large-scale structure. Still, detailed N-body simulations are needed to fully capture their impact in the nonlinear regime.

A good first ansatz for semi-analytic predictions is to assume that halos form out of the CDM and baryon components only. That is, when the peak height is defined as $\nu_c \equiv \delta_{\text{cr}}/\sigma_{c+b}$, where σ_{c+b} is the variance of CDM+baryon fluctuations computed with the appropriate *scale-dependent* linear growth factor, then the halo mass function in this approximation is close to universal as in the standard ΛCDM case [657]. Correspondingly, we can perform the bias expansion of the halo density field in terms of δ_{c+b}, such that the halo power spectrum reads

$$P_{hh}^{\text{LO}}(k) = [b_1^{c+b}(\nu_c)]^2 P_{c+b,\text{L}}(k) + P_\varepsilon^{\{0\}} \,, \tag{8.6}$$

where we have emphasized that the linear bias b_1^{c+b} is the bias with respect to the baryon+CDM components. Eq. (8.6) is fairly accurate on linear scales and, moreover, $b_1^{c+b}(\nu_c)$ is a nearly universal function of the peak significance $\delta_{\text{cr}}/\sigma_{c+b}$ [655, 657].

However, Eq. (8.6) is not completely correct, since a scale-dependent growth generically leads to a scale-dependent bias [670, 671] (see also Sec. 8.3). In particular, if halo positions are defined in Lagrangian space and follow the density locally there, as is the case in the peak and excursion-set pictures, then their relation to the Eulerian matter density field will be scale dependent. This effect was studied in detail for the massive neutrino case in [660], who showed that b_1^{c+b} in Eq. (8.6) in fact becomes

$$b_1^{c+b}(\nu_c; k) = 1 + \frac{1}{\overline{n}_h} \frac{\partial \overline{n}_h}{\partial \delta_{\text{cr}}} \frac{\partial \delta_{\text{cr}}}{\partial \delta_{c+b,\ell}(k)} \,, \tag{8.7}$$

where $\delta_{c+b,\ell}(k)$ is a long-wavelength single Fourier mode perturbation in the $c+b$ fluid. $\partial\delta_{\text{cr}}/\partial\delta_{c+b,\ell}(k)$ can be calculated by following a multi-fluid spherical collapse calculation [661], where neutrino perturbations are treated linearly while a spherical tophat shell of the $c+b$ fluid is followed to collapse. Far inside the free-streaming scale $k \gg k_{\text{fs}}$, $\partial\delta_{\text{cr}}/\partial\delta_{c+b,\ell}(k) = -1$ as derived for the pure cold matter case in Sec. 3, while it is modified on larger scales. This leads to a scale-dependent feature in the halo bias Eq. (8.7) at the few percent level, which contributes to the scale-dependent effects of neutrinos in galaxy clustering. Specifically,

194

the scale dependence in $P_{hh}^{\rm LO}$ is predicted to be reduced by 20–40% compared to that in $P_{c+b,\rm L}(k)$ [660]. The scale-dependent effects induced by neutrinos are controlled by a different scale, the free-streaming scale $\lambda_{\rm fs} = 1/k_{\rm fs}$ [cf. Eq. (8.2)], than the higher-derivative bias contributions, which involve the scale R_*. The two sources of scale dependence can therefore be isolated provided that the range of wavenumbers probed is sufficiently large [138]. While the scale dependence in galaxy bias induced by neutrinos can also be formally captured by the higher-derivative terms described in Sec. 2.6, this expansion is only valid on scales $k \ll k_{\rm fs}$, and is thus very limiting given the large value of $1/k_{\rm fs}$. More explicit ways to include the effects of massive neutrinos in the general perturbative bias expansion should thus be developed (e.g., along the lines of [672]). Observationally, the scale dependence arising from massive neutrinos could be extracted with the multi-tracer techniques described in Sec. 7.6.2 [673]. For this, at least two tracers with different amplitudes of scale-dependent bias are necessary. It is important to stress that, if present, similar scale-dependent effects are expected for the higher-order biases, e.g. b_2^{c+b}, as well.

The presence of a scale-dependent bias with respect to δ_{c+b} is difficult to confirm unequivocally in N-body simulations including massive neutrinos due to its small amplitude and the degeneracy with the scale dependence of halo clustering induced by nonlinear and higher-derivative contributions. Refs. [657, 666, 668] reported no clear evidence for a scale dependence of $P_{h(c+b)}(k)/P_{c+b}(k)$ beyond what is expected from nonlinear contributions in the CDM+baryon sector. On the other hand, Ref. [669] used a generalization of the separate-universe technique described in Sec. 3.2 and Sec. 4.4 to isolate the scale-dependent bias. They found a scale dependence consistent with Eq. (8.7), with an overall amplitude of roughly $6f_\nu$ acting in the opposite direction as the matter power suppression due to neutrinos. Very recently, Ref. [674] reported a detection from halo clustering measured in a large suite of N-body simulations with neutrinos. In order to increase the signal while keeping the free-streaming scale in the linear regime, the number of neutrino species was multiplied by a factor ~ 10 in the simulations. The amplitude was found to be largely consistent with the separate-universe results of [669].

8.2 Imprints of primordial baryon acoustic oscillations

Above we have seen that neutrinos leave a scale-dependent imprint on structure formation below the free-streaming scale, despite the fact that they constitute at most a few percent of the cosmic energy budget today. We now turn to the two most important matter components: cold dark matter (CDM, c) and baryons (b, i.e. all standard model particles that have been non-relativistic since the end of radiation domination; this excludes standard model neutrinos within the allowed mass range). So far, we have assumed that on large scales, where non-gravitational forces can be neglected, baryons and CDM comove and, therefore, can be treated as a single fluid. The density and velocity differences due to pressure forces and feedback are, in the general bias expansion, taken into account through the higher-derivative biases such as $b_{\nabla^2\delta}\nabla^2\delta$ (Sec. 2.6). This assumption is true for the adiabatic, growing-mode density perturbations which dominate structure formation at low redshifts. However, before the epoch of recombination, where protons and electrons first combined to form neutral hydrogen, baryons were tightly coupled to photons in a plasma and thus behaved very differently from CDM. When setting the initial conditions around recombination (strictly speaking, baryon-photon decoupling), this introduces decaying modes which can have a lasting imprint on galaxy statistics. Note that in most cases, these effects are not taken into account when setting up the initial conditions for N-body+hydro simulations. That is, typically the initial conditions only include the adiabatic growing mode for both baryons and CDM. Studies considering baryons and CDM separately in perturbation theory and simulations can be found in [675, 132, 676, 677]. We now study these decaying modes, focusing on their contributions to the general galaxy bias expansion.

Consider the coupled evolution of the baryon and CDM fluids under gravity, i.e. after baryon-photon decoupling. At linear order, we have the following set of evolution equations:

$$\frac{\partial}{\partial\tau}\delta_s = -\theta_s\,, \quad s \in \{b, c\}$$

$$\frac{\partial}{\partial\tau}\theta_s + \mathcal{H}\theta_s = -\frac{3}{2}\Omega_m(a)\mathcal{H}^2\delta_m\,, \tag{8.8}$$

where $\delta_s \equiv \delta\rho_s/\overline{\rho}_s$, $\theta_s = \partial_j v_s^j$, while $\delta_m \equiv \delta_{c+b} = f_b\delta_b + (1-f_b)\delta_c$ is the total matter density perturbation and $f_b = \Omega_b/\Omega_m$ is the baryon fraction (we neglect the small contribution from massive neutrinos throughout this section). Note that the CDM density perturbation δ_c used in this section is not to be confused with the critical density for collapse δ_{cr}.

It is useful to combine these equations and to rewrite them in terms of δ_m, and the relative density perturbation $\delta_r \equiv \delta_b - \delta_c$ and relative velocity divergence $\theta_r \equiv \boldsymbol{\nabla} \cdot \boldsymbol{v}_r \equiv \boldsymbol{\nabla} \cdot (\boldsymbol{v}_b - \boldsymbol{v}_c)$ [678, 679, 680]:

$$\frac{\partial^2}{\partial\tau^2}\delta_m + \mathcal{H}\frac{\partial}{\partial\tau}\delta_m - \frac{3}{2}\Omega_m(a)\mathcal{H}^2\delta_m = 0$$
$$\frac{\partial^2}{\partial\tau^2}\delta_r + \mathcal{H}\frac{\partial}{\partial\tau}\delta_r = 0$$
$$\frac{\partial}{\partial\tau}\theta_r + \mathcal{H}\theta_r = 0\,. \tag{8.9}$$

We can now immediately obtain the general solution of these three ODE as

$$\delta_m(\boldsymbol{x},\tau) = A_+(\boldsymbol{q})D(\tau) + A_-(\boldsymbol{q})H(\tau)$$
$$\delta_r(\boldsymbol{x},\tau) = R_+(\boldsymbol{q}) - R_-(\boldsymbol{q})D_r(\tau) \qquad \text{(linear)}$$
$$\theta_r(\boldsymbol{x},\tau) = \frac{H_0}{a(\tau)}R_-(\boldsymbol{q})\,, \tag{8.10}$$

where $A_\pm(\boldsymbol{q})$, $R_\pm(\boldsymbol{q})$ are the initial conditions (hence evaluated at the Lagrangian position $\boldsymbol{q}$, with $\boldsymbol{x} = \boldsymbol{q}$ at the order considered here), and

$$D_r(\tau) \equiv H_0 \int_0^\tau \frac{d\tau'}{a(\tau')} \tag{8.11}$$

is the growth factor of the R_- mode.[21] The total matter density contrast δ_m contains the growing and decaying modes $\propto A_\pm$ of adiabatic perturbations. We have employed the former extensively throughout all of Secs. 2–7. A third mode, $R_+(\boldsymbol{q}) \equiv \delta_{bc}(\boldsymbol{q})$, is a constant compensated density perturbation $\delta\rho_c = -\delta\rho_b$, corresponding to $\delta_m = 0$ while $\delta_r \neq 0$ [681, 682, 683]. This mode can be seen as modulating the local baryon-CDM ratio (see below). The fourth mode $\propto R_-$ corresponds to an initial relative velocity between the two fluids, which scales as $a^{-1}(\tau)$ (just as expected for an unsourced peculiar velocity) and is commonly denoted as $\boldsymbol{v}_{bc}$ in the literature. In particular, we can relate the latter to R_- through

$$\boldsymbol{v}_{bc}(\boldsymbol{q},\tau_0) \equiv H_0\boldsymbol{R}_-(\boldsymbol{q}) \equiv H_0\frac{\boldsymbol{\nabla}_q}{\nabla_q^2}R_-(\boldsymbol{q})\,. \tag{8.12}$$

Correspondingly, we denote $\theta_{bc} \equiv \boldsymbol{\nabla} \cdot \boldsymbol{v}_{bc}$, with $\theta_{bc}(\boldsymbol{q},\tau_0) = H_0 R_-(\boldsymbol{q})$. While the notation $\delta_{bc}, \boldsymbol{v}_{bc}$ is more common in the literature, the mode amplitudes R_+, R_- (or $\boldsymbol{R}_-$) are more convenient for deriving the contributions to the bias expansion. However, both notations are entirely equivalent. We refer to these two modes as *baryon-CDM perturbations* in the following.

Note the distinction in notation between the (in general) *nonlinearly evolved relative density and velocity perturbations* δ_r, $\boldsymbol{v}_r$, and the *linearly extrapolated initial amplitudes of the independent modes* of the system, R_+, R_-, or, equivalently, $\delta_{bc}, \boldsymbol{v}_{bc}$. This is analogous to the distinction between $\delta^{(1)}$ and $\delta \equiv \delta_m$ in the adiabatic case. During nonlinear gravitational evolution, the decaying modes couple with the growing mode, leading to nonlinear evolution of the baryon-CDM perturbations (in particular the relative velocity), and making this distinction important. Moreover, δ_r receives contributions from both modes R_+ and R_- [Eq. (8.10)]. As we will see however, the contributions to the bias expansion can be written completely in terms of the initial amplitudes of the independent modes R_+, R_-, so that we do not need to consider δ_r, $\boldsymbol{v}_r$ explicitly. Tab. 13 provides a summary of the notation used in this section.

---

[21]Note this was defined with a different sign in [678].

δ_s	Density perturbation of species $s \in b, c$
$\delta_m = f_b \delta_b + (1 - f_b) \delta_c$	Total matter density perturbation
$\delta_r = \delta_b - \delta_c$ [Eq. (8.10)]	Relative density perturbation
$\boldsymbol{v}_r = \boldsymbol{v}_b - \boldsymbol{v}_c$	Relative velocity
$R_+(\boldsymbol{x}) \equiv \delta_{bc}(\boldsymbol{x})$	*Initial amplitude* of constant relative density perturbation
$R_-(\boldsymbol{x}) \equiv H_0^{-1} \partial_i v_{bc}^i(\boldsymbol{x}, \tau_0)$	*Initial amplitude* of decaying relative velocity, extrapolated to τ_0

Table 13: Notation for baryon-CDM perturbations used in this section. Actual relative density and velocity perturbations are denoted with a subscript r, while their initial amplitudes (extrapolated to $z = 0$ using linear theory), which appear in the bias expansion, are denoted with a subscript bc, or equivalently as $R_\pm$. Note that the CDM density perturbation δ_c is to be distinguished from the spherical collapse threshold $\delta_{\rm cr}$.

Before deriving the contributions to the bias expansion, let us discuss the significance of the modes R_+, R_-. Ref. [684] pointed out that pre-recombination plasma waves, the baryon acoustic oscillations, source R_- and so lead to a supersonic streaming velocity at the epoch of baryon-photon decoupling $\tau_{\rm dec}$. This can leave an imprint in low-redshift structures which assembled out of low-mass halos at high redshifts [685, 686, 687, 688, 680]; note that in many of these references, the initial amplitude of the relative velocity is normalized to its variance, $\boldsymbol{v}_{bc} \to \boldsymbol{v}_{bc}/\sqrt{\langle |\boldsymbol{v}_{bc}^2| \rangle}$, which we do not do here. Similarly, the R_+-mode is also sourced during recombination [681]. Both R_- and R_+ have significant large-scale correlations and in particular retain large BAO features. However, they are small numerically: the root-mean-square relative density perturbation is $\sqrt{\langle R_+^2 \rangle} \lesssim 0.03$, depending on smoothing scale, while the root-mean-square relative velocity perturbation is approximately $0.03\,{\rm km/s}$ at the present epoch; their fractional impact on the large-scale galaxy power spectrum is expected to be at the sub-percent level (Fig. 36).

Since galaxy formation depends sensitively on both baryons and CDM, it is crucial to include the additional modes R_+, R_- when making predictions for galaxy clustering. We now briefly derive which operators these modes add to the bias expansion in the general case, under the same assumptions that were made in Sec. 2.5. We will ignore the adiabatic decaying mode (A_-) throughout, as it is numerically much smaller [159]. At linear order, the result is then already clear from Eq. (8.10), as we have to allow for the remaining three modes of the baryon-CDM system to appear in the galaxy overdensity δ_g:

$$\delta_g^{(1)}(\boldsymbol{x}, \tau) = b_1 \delta_m^{(1)}(\boldsymbol{x}, \tau) + b_{R+}(\tau) R_+(\boldsymbol{q}[\boldsymbol{x}, \tau], \tau) + b_{R-}(\tau) R_-(\boldsymbol{q}[\boldsymbol{x}, \tau], \tau)$$
$$= b_1 \delta_m^{(1)}(\boldsymbol{x}, \tau) + b_\delta^{bc} \delta_{bc}(\boldsymbol{q}) + b_\theta^{bc} \theta_{bc}(\boldsymbol{q}, \tau) , \tag{8.13}$$

where in the second line we have used Eq. (8.12). Note that R_+, R_- (or θ_{bc}, δ_{bc}) have to be evaluated at the Lagrangian position $\boldsymbol{q}[\boldsymbol{x}, \tau]$ corresponding to $(\boldsymbol{x}, \tau)$ [680]. This is analogous to the case of primordial non-Gaussianity, where the additional operators are also evaluated at the Lagrangian position (Sec. 7.1.2); in both cases, additional modes are present in the initial conditions. Strictly speaking, in case of the relative baryon-CDM perturbations, we should use the position of the fluid trajectory at the time of recombination. However, the distinction between these two is very small, and of similar order as other nonlinear contributions at recombination that are neglected throughout.

At second order, we expect all combinations of R_+, R_- (δ_{bc}, θ_{bc}) with themselves as well as δ_m to appear. However, there is a further contribution with a different structure, given by

$$\delta_g^{(2)}(\boldsymbol{x}, \tau) \supset b_{R_-^2}(\tau) \left[|R_-|^2(\boldsymbol{q}, \tau) - \langle |R_-|^2 \rangle \right] = b_{v^2}^{bc}(\tau) \left[|\boldsymbol{v}_{bc}|^2(\boldsymbol{q}, \tau) - \langle |\boldsymbol{v}_{bc}|^2 \rangle \right] . \tag{8.14}$$

This is the leading contribution to the galaxy density of a *uniform relative velocity* between baryons and CDM; since the velocity is a vector, it cannot contribute to the scalar galaxy density at linear order.

In order to derive all further nonlinear baryon-CDM contributions to galaxy clustering, consider the full nonlinear evolution equations of the two-fluid system. Generalizing the results of Sec. 2.4 to two fluids, this

can be written as [678]

$$\frac{D}{D\tau}\delta_s + \theta_s = -\delta_s\theta_s - g_s v_r^i \partial_i \delta_s \tag{8.15}$$

$$\left(\frac{D}{D\tau} + \mathcal{H}\right)\theta_s + \frac{3}{2}\Omega_m\mathcal{H}^2\delta_m = -(\partial^i v_s^k)^2 - g_s v_r^i \partial_i \theta_s \,,$$

where $s \in b, c$ and

$$\frac{D}{D\tau} \equiv \frac{\partial}{\partial\tau} + v_m^i \frac{\partial}{\partial x^i} \tag{8.16}$$

is the convective time derivative with respect to the total fluid velocity $\boldsymbol{v}_m = f_b\boldsymbol{v}_b + (1-f_b)\boldsymbol{v}_c$, and we have introduced the constants $g_b = 1 - f_b$ and $g_c = -f_b$. Note that the relative velocity perturbation is not sourced even at nonlinear order; that is, subtracting the Euler equation for θ_c from that for θ_b only yields source terms that are proportional to $\boldsymbol{v}_r$ (see also Sec. 2.7). Moreover, the typical linear displacement between the baryon and CDM fluids induced by $\boldsymbol{v}_r$ is very small:[22] $\boldsymbol{s}_r = \int d\tau\, \boldsymbol{v}_r \lesssim 0.04\,h^{-1}\,\mathrm{Mpc}$. This is much smaller than the nonlinear scale at redshifts where we observe galaxies, as well as the characteristic scale R_* expected for realistic galaxies. We can then effectively assume that baryons and CDM still travel on the same fluid trajectories $\boldsymbol{x}_{\mathrm{fl}}(\tau)$, since the leading effects of the different fluid trajectories are given by higher-derivative terms of order $\boldsymbol{s}_r \cdot \boldsymbol{\nabla}\delta_s$ which are negligibly small compared to the higher-derivative terms which are present for realistic galaxies, as well as the effective fluid contributions for matter.

One can then easily verify that it is sufficient to include [267, 678] $R_+(\boldsymbol{q})$ and $\boldsymbol{R}_-$ defined in Eq. (8.12) in the bias expansion, as well as higher spatial derivatives of these quantities and their combination with the observables derived for the adiabatic single-fluid case (Sec. 2.5). Note that, unlike the contributions from the growing mode derived in detail in Sec. 2, we do not have to include time derivatives of R_+, $\boldsymbol{R}_-$, since these are fixed amplitudes defined in the initial conditions. This is analogous to the additional fields ϕ, ψ appearing in the case of PNG (Sec. 7.1). Another way of seeing this is to allow for a dependence of the galaxy density on $\delta_r = \delta_b - \delta_c$, $\boldsymbol{v}_r = \boldsymbol{v}_b - \boldsymbol{v}_c$ along the fluid trajectory. Up to any order in perturbation theory, and neglecting the displacements between the two fluids as described above, this relative density and velocity along the fluid trajectory can be written, using the equations of motion Eq. (8.15), as a local function along the fluid trajectory of R_+, $\boldsymbol{R}_-$ and the $\Pi_{ij}^{[n]}$ ($n \geq 1$) appearing in the galaxy bias expansion of Sec. 2.5.3. Thus, by allowing for R_+, $\boldsymbol{R}_-$ to appear in all combinations with the $\Pi^{[n]}$, we allow for a general dependence of the galaxy density on the baryon-CDM relative density and velocity perturbations. Eq. (8.12) also shows how spatial derivatives of R_+, $\boldsymbol{R}_-$ should be counted. Specifically, $\partial^j R_-^i$ is of the same order as R_+. Indeed, both relative density perturbations (R_+) as well as relative velocity ($\boldsymbol{R}_-$) and relative velocity shear ($\partial^i R_-^j$) between different fluids are leading local observables. On the other hand, operators involving two or more derivatives of $\boldsymbol{R}_-$, as well as those with at least one derivative on R_+, will be suppressed by the spatial scale R_* associated with the formation of the galaxies considered (e.g., the Lagrangian radius of the host halos). Again, this holds because this scale is much larger than the absolute displacement $\boldsymbol{s}_r$ due to the decaying relative velocity.

Explicitly, we make use of the single-fluid basis of operators presented in Sec. 2.5.3. The contributions of baryon-CDM perturbations can then be succinctly summarized in matrix notation by defining

$$(\nabla R_-)_{ij}(\boldsymbol{q}) \equiv \partial_q^i R_-^j(\boldsymbol{q}) = \frac{\partial_q^i \partial_q^j}{\nabla_q^2} R_-(\boldsymbol{q})\,. \tag{8.17}$$

Up to third order in perturbation theory, the relative density and velocity contributions to the general

---

[22]The value given here is at $z = 0$, and hence the maximal value. The bulk of the displacement occurs at high redshifts $z \gtrsim 30$.

Eulerian bias expansion then are [678]:

$$
\begin{aligned}
1^{\text{st}} \quad & R_+(\boldsymbol{q})\,,\ \text{tr}[(\nabla R_-)](\boldsymbol{q}) \\[4pt]
2^{\text{nd}} \quad & R_+\,\text{tr}[\Pi^{[1]}]\,,\ \boldsymbol{R}_-^2\,,\ \text{tr}[(\nabla R_-)\Pi^{[1]}]\,,\ \text{tr}[(\nabla R_-)]\,\text{tr}[\Pi^{[1]}] \\[4pt]
3^{\text{rd}} \quad & R_+\,\text{tr}[\Pi^{[1]}\Pi^{[1]}]\,,\ R_+(\text{tr}[\Pi^{[1]}])^2\,,\ \boldsymbol{R}_-^2\,\text{tr}[\Pi^{[1]}]\,, \\
& \boldsymbol{R}_-\Pi^{[1]}\boldsymbol{R}_-\,,\ \text{tr}[(\nabla R_-)\Pi^{[1]}\Pi^{[1]}]\,,\ \text{tr}[(\nabla R_-)\Pi^{[1]}]\,\text{tr}[\Pi^{[1]}]\,, \\
& \text{tr}[(\nabla R_-)]\,\text{tr}[\Pi^{[1]}\Pi^{[1]}]\,,\ \text{tr}[(\nabla R_-)](\text{tr}[\Pi^{[1]}])^2\,,\ \text{tr}[(\nabla R_-)\Pi^{[2]}]\,,
\end{aligned}
\tag{8.18}
$$

where R_+, R_- are all evaluated at the Lagrangian position $\boldsymbol{q}[\boldsymbol{x},\tau]$ while $\Pi_{ij}^{[n]}$ are evaluated at $(\boldsymbol{x},\tau)$. The first-order terms are precisely those included in Eq. (8.13), while the second line includes Eq. (8.14), among several others. Here, we have only included terms linear in R_+, ∇R_-, while we have included terms up to second order in $\boldsymbol{R}_-$. This is justified because the contributions to galaxy statistics are percent-level corrections to the standard adiabatic contributions, so that higher-order terms are highly suppressed. When extending Eq. (8.18) to include higher-derivative operators, one should correspondingly also include terms of the type $(\boldsymbol{R}_- \cdot \boldsymbol{\nabla})\,\text{tr}[\Pi^{[1]}]$, i.e. $v_{bc}^i \partial_i \delta$ (see [677], who perform a resummation of this type of term). While we have defined derivatives on R_-^i to be with respect to $\boldsymbol{q}$, this distinction is in fact not important, since the terms obtained when transforming $\partial/\partial q^i$ to a derivative with respect to Eulerian coordinate $\boldsymbol{x}$ are already included in the list Eq. (8.18). Eq. (8.18) is sufficient to derive all contributions to the leading-order galaxy bispectrum and the one-loop galaxy power spectrum; the latter is derived in detail in [678].

In addition, as noted by [680], the fact that R_+, R_- are evaluated at the Lagrangian position introduces further terms. This is again analogous to the case for primordial non-Gaussianity, Sec. 7.1.2. Expanding the Lagrangian position $\boldsymbol{q}$ around the Eulerian position $\boldsymbol{x} = \boldsymbol{q} + \boldsymbol{s}$ induces displacement terms for each operator involving R_+, R_-. However, they are multiplied by the bias parameter of the corresponding operator in Eq. (8.18), and thus do not introduce additional free parameters. Finally, in addition to the deterministic operators listed in Eq. (8.18), there are also stochastic contributions of the same type as discussed in Sec. 2.8. That is, for each operator O in the list Eq. (8.18), one introduces a stochastic field ε_O with zero mean that is, at lowest order in derivatives, fully characterized by its one-point moments.

The baryon-CDM perturbations also affect the velocities of galaxies. Specifically, at linear order one has

$$
\boldsymbol{v}_g(\boldsymbol{x},\tau) = \boldsymbol{v}_m(\boldsymbol{x},\tau) + \beta_v^{bc}(\tau)\boldsymbol{v}_{bc}(\boldsymbol{q})\,,
\tag{8.19}
$$

where β_v^{bc} is a velocity bias parameter. This contribution captures the fact that galaxy velocities could either follow the baryons ($\beta_v^{bc} = 1 - f_b$) or CDM ($\beta_v^{bc} = -f_b$), or some linear combination of the two. Thus, one expects that β_v^{bc} is at most of order one. The contribution Eq. (8.19) is relevant for galaxy surveys through the effect of redshift-space distortions (Sec. 9.3.2), where, in case of the leading galaxy two-point function, it leads to a term comparable in magnitude to the linear-order terms in Eq. (8.13) [678].

This completes the incorporation of baryon-CDM perturbations in the general perturbative bias expansion. However, in order to assess the quantitative impact of these contributions, we need estimates for the bias parameters associated with each operator in Eq. (8.18). For this, we divide the terms into three distinct classes, corresponding to different physical mechanisms:

(i) operators involving $R_+ = \delta_{bc}$: the R_+ mode essentially modulates the local baryon-CDM ratio through $(\Omega_b/\Omega_c)_{\boldsymbol{x}} = \overline{(\Omega_b/\Omega_c)}[1 + R_+(\boldsymbol{q})]$. That is, it can be described by a generalization of the separate-universe picture of Sec. 3, where instead of varying Ω_{m0}, Ω_{K0} we keep them fixed and vary Ω_{b0}, Ω_{c0} instead. Thus, in the presence of a long-wavelength R_+ perturbation, galaxies form in an environment with slightly different composition. Refs. [681, 678] argue that the bias with respect to this mode is expected to be of order unity, since a modulation of Ω_b/Ω_c corresponds to a modulation of the baryonic mass available to form stars. Moreover, it is not unreasonable to expect an enhancement for rare massive galaxies, like in the case of the bias parameters $b_1, b_2, \cdots$ for halos, due to the similar exponential cutoff of the galaxy stellar mass function at high masses. Recently, Ref. [689] constrained $|b_{R_+}| = |b_\delta^{bc}| \lesssim 6$ (95% confidence level) from the BOSS CMASS sample.
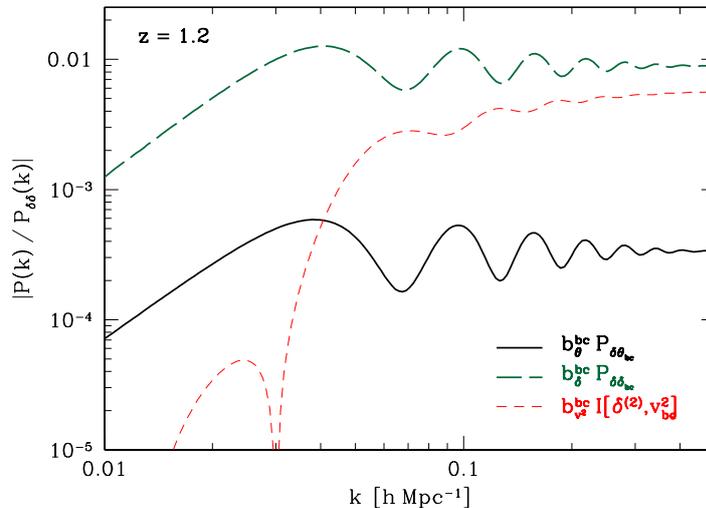
Figure 36: Leading baryon-CDM relative density and velocity contributions to the galaxy power spectrum (at $z = 1.2$), relative to the linear matter power spectrum (solid: θ_{bc}; long-dashed: δ_{bc}). Here, $b_\delta^{bc} = 1$ and $b_\theta^{bc} = 6.8/[(1+z)H_0]$ [Eq. (8.21)], setting $b_1 = 2$] was assumed. Also shown is the largest of the NLO contributions from the term $|\boldsymbol{v}_{bc}|^2$ [Eq. (8.14)], $b_{v2}^{bc}\mathcal{I}^{[\delta^{(2)},v_{bc}^2]}(k)$ [Eq. (8.22)] as short-dashed line, assuming $b_{v2}^{bc} = 0.01\langle|\boldsymbol{v}_{bc}|^2\rangle^{-1}$, close to the maximum value allowed by current constraints. From [678].

(ii) operators involving $R_-^i R_-^j \propto v_{bc}^i v_{bc}^j$: the second-order effect of a uniform relative velocity, first introduced by [684], has been investigated extensively in the recent literature [685, 686, 687, 688, 680]. Ref. [685] argued that $\boldsymbol{v}_{bc}$ increases the effective sound speed c_s of the neutral gas, so that the Jeans mass M_J increases by a factor $[1 + v_{bc}^2/c_s^2]^{3/2}$. This leads to large effects on low-mass halos prior to reionization, as investigated using small-box simulations in [690, 691, 692]. These could be transferred to galaxies of much larger mass observed at low redshifts by modulations of the metal enrichment and hence star formation efficiency, for example. To what extent this occurs quantitatively is still unclear. b_{v2}^{bc} could be as small as $\sim 10^{-5}\langle|\boldsymbol{v}_{bc}|^2\rangle^{-1}$ [680, 690]. The fiducial value adopted in previous studies for redshifts $z \lesssim 2$ is [685, 686, 687, 680],

$$b_{v2}^{bc} \sim 0.01\,\langle|\boldsymbol{v}_{bc}|^2\rangle^{-1} \approx 9 \times 10^{11}\,(1+z)^{-2}\,, \qquad (8.20)$$

where we work in units where the speed of light is $c = 1$. Note that, given the non-detection in current data, b_{v2}^{bc} cannot be much larger than this [687, 693]; in particular, [693] find an upper limit of $b_{v2}^{bc} < 0.01$ from the three-point function of the SDSS BOSS DR12 CMASS sample [30].

(iii) operators involving $\partial^i R_-^j \propto \partial^i v_{bc}^j$: these include the linear operator θ_{bc}. Physically, this corresponds to a uniform initial relative velocity divergence between baryons and CDM. A very simple estimate for the associated bias parameter can be obtained by noting that from Eq. (8.10), this induces an associated relative density perturbation of order $\delta_r \simeq a^{1/2}\theta_{bc}/H_0$. Assuming that the response of the galaxy density to this relative density perturbation is of order unity, as for the constant compensated mode δ_{bc}, one obtains $b_\theta^{bc}(\tau) \sim a^{1/2}(\tau)H_0^{-1}$. However, the physics of this mode, an initial relative velocity divergence, is quite different from that of δ_{bc}, which corresponds to a changed composition. A potentially more accurate estimate was presented in [678] using a spherical collapse calculation. By following baryon and CDM shells separately, one can approximately take into account the initial relative velocity divergence in the collapse. This in turn can be used to derive the effect on the collapse threshold δ_{cr}, which can be used to estimate the bias through the excursion-set argument (Sec. 5.6.2). Ref. [678] obtained

$$b_\theta^{bc}(z) = [(1+z)H_0]^{-1}\frac{\partial\delta_{cr}(z)}{\partial(\theta_{bc,0}/H_0)}(1 - b_1) \approx 6.8[(1+z)H_0]^{-1}(b_1 - 1)\,. \qquad (8.21)$$

200

Thus, we expect the bias parameters associated with the operators in this class to be of order several times H_0^{-1}. The streaming velocity effects on very low-mass halos discussed above can induce an additional contribution to b_θ^{bc} that is proportional to $b_{v^2}^{bc}$. Depending on the value of the latter bias parameter, this could be as large as the estimate Eq. (8.21), but is likely to be smaller [678].[23]

Fig. 36 shows the leading contributions to the galaxy power spectrum from each of these three classes of terms. We see that the $R_+ = \delta_{bc}$ contribution dominates, while the $\theta_{bc} = \partial_i R_-^i$ contribution is the smallest. The contribution from v_{bc}^2 shown here is only one of several terms at NLO [680, 678], namely

$$\mathcal{I}^{[\delta^{(2)}, v_{bc}^2]}(k) = -2 \int_{\boldsymbol{p}} F_2(\boldsymbol{p}, \boldsymbol{k} - \boldsymbol{p}) \frac{\boldsymbol{p} \cdot (\boldsymbol{k} - \boldsymbol{p})}{p^2 |\boldsymbol{k} - \boldsymbol{p}|^2} P_{\delta\theta_{bc}}(p) P_{\delta\theta_{bc}}(|\boldsymbol{k} - \boldsymbol{p}|) \,. \tag{8.22}$$

Note however that the bias parameter $b_{v^2}^{bc}$ is highly uncertain, and the value adopted for Fig. 36 is near the maximum currently allowed value. While Fig. 36 clearly shows that the baryon-CDM contributions are very small at low redshifts, their prominent BAO features, which are not exactly in phase with those in the adiabatic growing mode of matter, make them relevant for the BAO feature as standard ruler in galaxy correlations (see the recent discussion in [30]).

Finally, Ref. [159] recently pointed out another physical effect which enters the large-scale galaxy bias expansion. After reionization, the free electrons in the Universe are weakly coupled to the free-streaming CMB photons via Thomson scattering. The electrons in turn are bound to the nuclei via Coulomb forces. If the baryons are at rest in the CMB frame, this scattering has no dynamical effect. On the other hand, if the gas is moving relative to the CMB, it sees a dipole in its rest frame, which leads to a drag force which is proportional to the velocity. In fact, this is the same drag that baryons experience before recombination. This drag force supplies a source term to the baryon-CDM relative velocity given by [159]

$$\boldsymbol{v}_r' + \mathcal{H}\boldsymbol{v}_r = -\, x_e \alpha \mathcal{H} \, \boldsymbol{v}_b \,, \tag{8.23}$$

where $x_e(\tau)$ is the electron ionization fraction, $\boldsymbol{v}_b$ is the baryon fluid velocity relative to the CMB frame, and the dimensionless coefficient $\alpha(\tau)$ is given by

$$\alpha(\tau) = a(\tau) \frac{\sigma_{\mathrm{T}} u_\gamma(\tau)}{Y_e m_p c \mathcal{H}(\tau)} \propto (1+z)^4 H^{-1}(z) \,. \tag{8.24}$$

Here, σ_{T} is the Thomson cross section, $u_\gamma(T)$ is the energy density of blackbody radiation of temperature T, $Y_e \approx 1.08$ is the electron molecular weight and m_p is the proton mass (α is of order 10^{-6} today). The drag contribution leads to a relative velocity which, at redshifts $z \lesssim 10$, is of the same order of magnitude as the primordial contribution $\boldsymbol{v}_{bc}$ discussed above. However, the scale dependence is very different, as $\boldsymbol{v}_{bc}$ carries an imprint of the baryon acoustic oscillations, while the Compton drag is controlled by $\boldsymbol{v}_b \approx \boldsymbol{v}$, which is dominated by dark matter, and its transfer function is proportional to $T(k)/k$, where $T(k)$ is the transfer function of the adiabatic growing mode (Sec. 7.1.1). This means that we have to separately account for the Compton drag contribution to the baryon-CDM relative velocity in the bias expansion. As shown in [159], the leading contributions are second order, and given by

$$\delta_g^{(2)}\Big|_{\mathrm{drag}} = b_{\mathrm{drag}} v^2 + b_{\mathrm{drag.bc}} \boldsymbol{v} \cdot \boldsymbol{v}_{bc} \,, \tag{8.25}$$

where the first term is induced by Compton drag alone, while the second term corresponds to the coupling between Compton drag and primordial relative velocity. Note that, when considering gravity alone, the equivalence principle forbids the velocity itself from appearing in the bias expansion (Sec. 2.9). However, the CMB radiation provides a preferred frame, so that the local velocity with respect to the CMB is an observable; the coupling to the gas induced by Thomson scattering after reionization provides the physical process by which this observable affects galaxy formation and thus enters the bias expansion.

---

[23]Ref. [680] adopt an effective b_θ^{bc} obtained from a loop integral that is approximately 20 times larger than Eq. (8.21) for the assumed value of $b_{v^2}^{bc}$. Here, we have absorbed this loop integral into a renormalized bias parameter b_θ^{bc} whose value needs to be determined from the data, following renormalized perturbation theory (Sec. 2.10). Note that both parametrizations are equivalent as long as one allows for both $b_{v^2}^{bc}$ and b_θ^{bc} to be free parameters in the fit to galaxy statistics.

8.3   Galaxy bias with dark energy and modified gravity

Throughout this review, we have assumed a cosmological constant Λ as explanation of the current accelerated expansion of the Universe. While all current cosmological observations appear consistent with this scenario, it is worth exploring other physical paradigms. The most popular alternative to Λ, *dark energy*, is usually described as a light scalar field whose potential energy provides the source of the accelerated expansion (see [694, 695] for reviews).

Most models of galaxy clustering incorporate dark energy approximately by including the effects of the modified expansion history in the linear growth factor $D(\tau)$. This approximation neglects perturbations in the dark energy density, which scale as $1 + w$, where w is the equation of state which is observationally constrained to be close to -1. For this reason, perturbations in the dark energy typically have a very small effect on the growth of structure. More precisely, this case parallels closely that of neutrinos (Sec. 8.1), with the free-streaming scale being replaced by the sound horizon, or Jeans scale, $k_J^{-1} = c_s/\mathcal{H}$ of the dark energy, where c_s is the sound speed. For a canonical scalar field, often referred to as *quintessence*, $c_s = 1$ so that the sound horizon is given by the comoving horizon $\mathcal{H}^{-1}$. Hence, structure formation happens on scales far inside the sound horizon of the dark energy, where dark energy perturbations are negligible. In that case, the above-mentioned approximation is accurate.

In the opposite limit, $c_s = 0$, realized for example by k-essence models, dark energy can be accurately modeled as a collisionless fluid [696]. In principle, the bias expansion should now contain the dark energy and matter density separately, including the relative velocity between the two [697]. However, in the absence of initial isocurvature perturbations between dark matter and dark energy, the relative velocity vanishes, and both matter and dark energy follow the same fluid trajectories. This implies that the density perturbations in dark energy and matter are proportional to each other. In that case, no new terms are added to the bias expansion even in the presence of clustering dark energy.

If the sound horizon of the dark energy is on intermediate, observable scales, then one expects the bias parameters to become scale dependent analogously to the case of massive neutrinos. The scale dependence of the linear bias was studied numerically in [698], by considering the two limiting cases of scales that are far inside and far outside the dark energy sound horizon. This study was based on a generalization of the separate-universe approach (Sec. 3.2) to include pressure perturbations [216]. The fractional difference in the Lagrangian bias b_1^L of dark matter halos between the sub-Jeans and super-Jeans limits was found to be at the few-percent level, for a dark energy with equation of state $w = -0.5$, and independent of halo mass (Fig. 11 in [698]; note that such a strong departure from $w = -1$ is already ruled out by observations). This implies that in the presence of dark energy, halos show a scale-dependent bias around the scale $k \sim k_J$ with an amplitude very roughly given by $\Delta b_1 \sim 0.1(1 + w)\, b_1^L$.

A fundamental alternative to dark energy is to modify General Relativity (GR) on large scales in order to yield an accelerating Universe without an exotic stress-energy component. We now discuss the implications of such modifications to GR (*modified gravity*), where we restrict to modified gravity models that obey the weak equivalence principle at the level of the particle action. This means that there is a well-defined spacetime whose geodesics govern the motion of test particles in the absence of non-gravitational forces (universality of free fall). Typically, modified gravity theories introduce an additional scalar degree of freedom, so that searching for this degree of freedom is a promising avenue to test gravity. We refer to [699, 700] for comprehensive and [701] for a concise review of modified gravity in the context of cosmology.

As discussed in Sec. 2.9, the local gravitational observables corresponding to long-wavelength perturbations (i.e., at lowest order in derivatives) consist of the local Hubble rate, tidal field, and spatial curvature on constant-proper-time slices (K_F in Sec. 2.9). These are supplemented by the matter density δ and the velocity divergence θ and shear tensor $\partial^i v^j$. The derivation of this fact is purely geometrical, i.e. it does not rely on the validity of the Einstein equations. Thus, all of this still holds in modified gravity. Moreover, δ and θ are related by the continuity equation, and θ, $\partial^i v^j$ are related to the tidal field through the Euler equation, both of which are unmodified. On the other hand, for non-relativistic tracers, the local effect of the spatial curvature is suppressed by $(v/c)^2$. Thus, even though the relations between δ, θ, and K_F are modified from those in GR, the reasoning in Sec. 2.5.1 still holds, and it is sufficient to include the matter density, tidal field, and convective time derivatives thereof in the bias expansion. However, the reduction of

these terms to only a handful of terms at each order in perturbation theory, which holds in GR as described in Sec. 2.5.2–2.5.3, is no longer possible in general for modified gravity.

To illustrate this, we adopt a scalar-tensor theory of Brans-Dicke type [702] as toy example of modified gravity. The well-studied $f(R)$ [703, 704] gravity model falls into this class [705]. This theory introduces a scalar degree of freedom ϕ with potential $V(\phi)$ and a specific coupling strength to matter. The linear growth factor equation, Eq. (B.9) in GR, is then modified to

$$\frac{d^2}{d\tau^2}D(k,\tau) + \mathcal{H}\frac{d}{d\tau}D(k,\tau) - \frac{3}{2}\Omega_m(a)\mathcal{H}^2\left[1 + \alpha(\tau)\frac{k^2}{k^2 + a^2m^2(\tau)}\right]D(k,\tau) = 0\,, \qquad (8.26)$$

where $\alpha(\tau) > 0$ is a coupling constant, while $m^2(\tau) = d^2V(\phi)/d\phi^2|_{\bar{\phi}(\tau)}$ is the mass of the scalar field at its cosmological background value $\bar{\phi}(\tau) = \langle\phi(\boldsymbol{x},\tau)\rangle$. Eq. (8.26) is derived by solving the usual linearized Euler-Poisson system augmented by the Klein-Gordon equation for ϕ, where time derivatives of the latter are neglected (the so-called quasi-static approximation appropriate on subhorizon scales). Clearly, the growth factor becomes scale-dependent unless $m = 0$. For $k \ll am$, corresponding to scales larger than the Compton length of the field, gravity reduces to GR, while on scales within the Compton length ($k \gg am$) gravity is enhanced by a factor $1 + \alpha$.

Now consider the evolution of a linear LIMD bias relation. At time $\tau = \tau_*$, we write $\delta_g^* = b_1^*\delta^*$. Then, linear evolution via Eq. (2.24) immediately yields [670, 671]

$$\delta_g(\boldsymbol{k},\tau) = b_1^E(k,\tau)\delta(\boldsymbol{k},\tau)\,, \quad b_1^E(k,\tau) = 1 + (b_1^* - 1)\frac{D(k,\tau_*)}{D(k,\tau)}\,. \qquad (8.27)$$

Thus, a LIMD (scale-independent) initial bias becomes nonlocal (scale-dependent) at a later time, unless $D(k,\tau)$ is separable in k and τ. This holds in the same way for the scale-dependent growth induced by free-streaming massive neutrinos (Sec. 8.1). From Eq. (8.26) we infer that, in the scalar-tensor example, the scale dependence will appear at $k \sim am(\tau)$. Since a general bias expansion should be able to describe the special case of a conserved, initially locally biased tracer, we clearly see that, in the case of a modified gravity scenario with scale-dependent growth, additional terms need to be included in the bias expansion already at linear order.

In full generality, a scale-dependent growth factor $D(k,\tau)$ that is not separable in k and τ precludes us from constructing a rigorous bias expansion in terms of a finite set of bias parameters. Recall that this construction relied on modes evolving at the same rate on large scales, so that time derivatives could be reordered to be successively higher order in perturbation theory. This no longer holds for a general $D(k,\tau)$. However, as shown by Eq. (8.26), on scales much larger than the Compton length of the additional degree of freedom, we can perform a perturbative expansion in $k^2/(am)^2$. Effectively, higher-derivative biases then absorb the effects of the fifth force consistently.

Viable modified gravity models typically include screening mechanisms which suppress the additional degrees of freedom in dense regions to satisfy Solar System constraints on gravity (see [700] for a general discussion). These are nonlinear mechanisms and hence need to be taken into account for nonlinear bias. Screening mechanisms provide motivation to use low-mass or low-density tracers such as dwarf galaxies or voids to probe gravity. For screening of the chameleon or symmetron types, we have to include ϕ itself in the bias expansion, since the screening threshold depends on the ambient field value. Note that for $k \ll am$, $\phi(\boldsymbol{k}) \sim k^2/(am)^2\Phi(\boldsymbol{k})$ is proportional to the density, rather than the potential. Models that invoke screening of this type have a Compton wavelength that is constrained to be less than $\sim 10\,h^{-1}$ Mpc in order to satisfy Solar System tests [706]. Thus, this dependence on ϕ can be taken into account via the higher derivative terms mentioned above. For screening of the MOND or k-essence type, the relevant variable is $\partial_i\phi/\mathfrak{a}_0$, where $\mathfrak{a}_0 \sim H_0$ is the MOND acceleration scale. Thus, the lowest-order contribution from MOND-type screening to the galaxy density is $(\partial_i\phi)^2/\mathfrak{a}_0^2$, which is of order $(v/c)^2 \lesssim 10^{-4}$ and thus expected to have a very small impact numerically. Finally, for models with screening of the Vainshtein type, the screening depends on $\partial_i\partial_j\phi$. This does not lead to new terms in the bias expansion, since $\partial_i\partial_j\phi$ is already captured by including $\partial_i\partial_j\Phi$ and ∂_iv_j (as well as their time derivatives) separately.

9 Connection to observations

So far, we have described the clustering of galaxies and halos in their rest frame. This was appropriate, since the focus of this review are physical bias expansions, which naturally describe the rest-frame galaxy density. We now turn to the connection of these rest-frame statistics to observations. The purpose of this section is to describe briefly all effects that enter the *theoretical prediction for the observed statistics of galaxies*, starting from the general bias expansion in the rest-frame of galaxies. We begin with the local (statistical) connection between galaxies and halos in Sec. 9.1, commonly described via halo occupation or abundance matching approaches. Although, in the context of the general bias expansion, *there is no need to describe galaxies in terms of their relation to halos,* as the bias expansion effectively takes into account the small-scale physics of galaxy formation, the relation between galaxies and halos provides useful physical insights, and can be used to extend the models of halo statistics described in Sec. 5–6 to galaxies.

We then review astrophysical selection effects in Sec. 9.2. We show that the general bias expansion is able to describe a diverse population of galaxies with a single set of effective bias parameters. However, certain selection effects can lead to additional terms in the bias expansion, which we describe there.

Sec. 9.3 then derives how the galaxy density is mapped to observed redshifts and positions on the sky (*projection effects*); this includes the important complication of redshift-space distortions, as well as so-called relativistic effects. Finally, Sec. 9.4 deals with the issue that galaxy surveys do not map out the galaxy distribution on a fixed time slice and a flat sky, but rather on the past light cone, which is especially relevant for large-scale perturbations on scales comparable to the typical radial distance to the observed galaxies in the survey.

9.1 The connection between galaxies and halos

Galaxy formation is a complex process involving the inflow and cooling of gas, and self-regulation via feedback effects (see [92] for an overview). Clearly, these processes go beyond the collapse of pressureless matter which governs the formation of dark matter halos in N-body simulations and which forms the starting assumption of both excursion-set and peak approaches. On the other hand, realistic numerical simulations of the large-scale distribution of galaxies remain computationally extremely challenging. Since the fact that galaxies are hosted by dark matter halos has been verified by both simulations and observations, analytic approaches which rely on the halo model [707, 296, 708, 709, 710, 711] (see [343] for a review) have been developed to describe the clustering of galaxies. In standard *halo occupation distribution (HOD)* models [712, 713, 714], galaxies are assumed to follow the matter distribution within their host dark matter halo, implemented for example by sampling the positions of dark matter particles or by using the universal mean mass profile of the NFW form [366]. Furthermore, the host halo mass M is usually the sole property determining the number of galaxies [715]. We will also assume steady emission from galaxies; the relation between abundance and clustering is modified for intermittent sources such as quasars, which can be used to constrain their lifetime [716].

Adopting the halo model ansatz, contributions to the clustering of galaxies can be split into those terms arising from distinct host halos ("two-halo term", in case of the galaxy two-point function), and those involving the same host halo ("one-halo term"). In keeping with the focus of this review, we here derive the large-scale clustering properties of galaxies predicted in this approach.

In most implementations of HOD approaches, galaxies are divided into "central" and "satellite" galaxies. Every halo hosting one or more galaxies has exactly one central galaxy which resides at or close to the

center-of-mass of the halo. All other galaxies within the same halo are then denoted as satellite galaxies. Let $N_c \in \{0, 1\}$ and $N_s \in \{0, 1, 2, \cdots\}$ be the number of central and satellite galaxies in a given halo of mass M, which we consider to be random variables. Since the existence of satellite galaxies is conditioned on the presence of a central galaxy, it is convenient to define $\mathcal{N}_s$ through $N_s = N_c \mathcal{N}_s$. Assuming that N_c and N_s are independent random variables, the average number density of galaxies reads

$$\overline{n}_g = \int d\ln M\, \overline{n}_h(M) \langle N_c \rangle_M \left[1 + \langle \mathcal{N}_s \rangle_M \right], \tag{9.1}$$

where $\langle N_c \rangle_M$ and $\langle \mathcal{N}_s \rangle_M$ are the expectation values of central and satellite galaxy numbers in halos of mass M, respectively. By definition, $\langle N_c \rangle_M \leq 1$. The linear bias parameter $b_{1,g}$ is usually written as a number-weighted integral over the halo bias $b_1(M)$ [717],

$$b_{1,g} = \overline{n}_g^{-1} \int d\ln M\, \overline{n}_h(M) \langle N_c \rangle_M \left[1 + \langle \mathcal{N}_s \rangle_M \right] b_1(M). \tag{9.2}$$

According to the general peak-background split argument, which is exact (Sec. 3), $b_{1,g}$ is the linear response of the mean galaxy density to a change in the mean density of the Universe. We then see that Eq. (9.2) is derived using the assumption that $\langle N_c \rangle_M$ and $\langle \mathcal{N}_s \rangle_M$ are independent of the background cosmology. That is, while the halo number density changes due to a long-wavelength density perturbation via $b_1(M)$, the occupation statistics at a given fixed halo mass are assumed to be unchanged. This might be a good first-order assumption in practice, although it is important to keep in mind that it is an approximation. The PBS argument can similarly be applied (with the same assumptions) to derive the higher-order LIMD bias parameters $b_{N,g}$.

One can also calculate the large-scale stochasticity of galaxies, assuming that the stochasticity of halos of mass M is given by Poisson shot noise $1/\overline{n}_h(M)$. Then, one obtains (e.g., [718])

$$P_{\varepsilon,g}^{\{0\}} = \frac{1}{\overline{n}_g^2} \int d\ln M\, \overline{n}_h(M) \langle N_c \rangle_M \left[1 + 2\langle \mathcal{N}_s \rangle_M + \langle \mathcal{N}_s^2 \rangle_M \right] \tag{9.3}$$

$$= \frac{1}{\overline{n}_g} + \frac{1}{\overline{n}_g^2} \int d\ln M\, \overline{n}_h(M) \langle N_c \rangle_M \left[2\langle \mathcal{N}_s \rangle_M + \left(\langle \mathcal{N}_s \rangle_M \right)^2 \right], \tag{9.4}$$

where the second line further assumes a Poisson distribution for $\mathcal{N}_s$. The stochastic contribution to the galaxy power spectrum thus also depends on the second moment of the distribution of satellite numbers at fixed halo mass.

Various parametrizations of $\langle N_c \rangle_M$ and $\langle \mathcal{N}_s \rangle_M$ exist in the literature. $\langle N_c \rangle_M$ typically follows a step-like function, in agreement with what is found for the distribution of subhalos (bound substructures of halos) in N-body simulations [713], whereas $\langle \mathcal{N}_s \rangle_M$ can be parametrized by a power-law with logarithmic slope $\alpha \approx 1$. In the model of [719] for instance, it is assumed that halos hosting satellite galaxies in a given luminosity-limited sample also host a central galaxy from the same sample,

$$\langle N_c \rangle_M = \frac{1}{2} \left[1 + \mathrm{erf}\left(\frac{\log M - \log M_{\min}}{\sigma_{\log M}} \right) \right] \tag{9.5}$$

$$\langle \mathcal{N}_s \rangle_M = \begin{cases} \left(\frac{M - M_0}{M_1'} \right)^\alpha, & M > M_0 \\ 0, & M \leq M_0. \end{cases} \tag{9.6}$$

Here, $M_{\min}$ represents the halo mass cutoff of central galaxies and $\sigma_{\log M}$ takes into account the scatter between galaxy luminosity and halo mass. For the satellite galaxies, M_0 is the halo mass cutoff, M_1' is the characteristic mass of halos harboring one satellite, and $\alpha \approx 1$ is the power-law slope. $\mathcal{N}_s$ is usually assumed to follow a Poisson distribution. All these model parameters depend on the galaxy sample considered, for example the galaxy luminosity.

For illustration, the left panel of Fig. 37 shows the projected two-point correlation function of galaxies with different luminosity threshold along with the best-fit HOD models. In order to use an N-body simulation
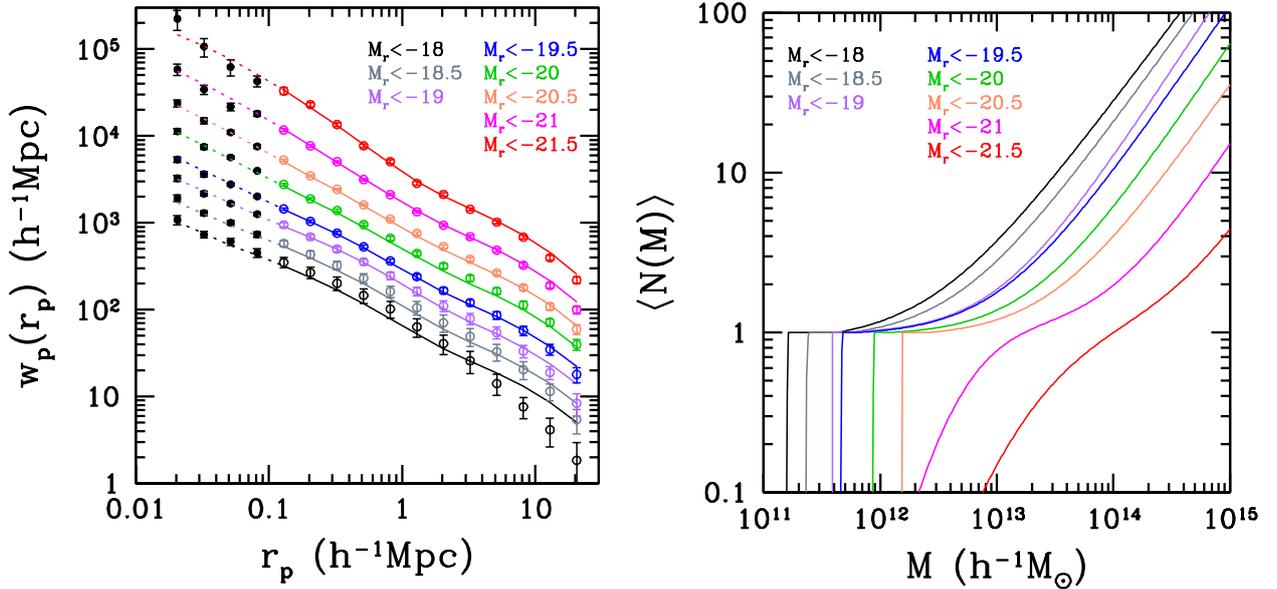
Figure 37: *Left panel*: 2-point projected galaxy correlation function $w_p(r_p)$ measured from the SDSS DR7 main sample with different luminosity thresholds as indicated in the figure. The solid curves represent the best-fit HOD models. Only the data at projected separation $r_p > 0.1\,h^{-1}$ Mpc was included in the fit. *Right panel*: Mean halo occupation function $\langle N(M) \rangle$ [Eq. (9.7)] of the best-fitting models. A value $\sigma_{\log M} = 0$ was assumed for the faint galaxy samples, for which the cutoff mass scale is not well constrained, which is visible as a sharp cut at low masses in the right panel. From [720].

to fit the HOD parametrization Eq. (9.6) to a given galaxy sample, centrals are assigned, for each halo in the simulation, with a probability $\langle N_c \rangle_M$ to the center-of-mass position and velocity of halos. Further, $\mathcal{N}_s$ is sampled from a Poisson distribution with mean $\langle \mathcal{N}_s \rangle_M$, and satellite galaxies are assigned the position and velocity of random dark matter particles within the halo. Then, the mean number density and the projected correlation function of this HOD sample is measured and compared to the data (see [720] for details). Overall, the N-body-based HOD approach describes the shape of the galaxy correlation function at different luminosity thresholds quite well over a wide range of scales. More luminous galaxies reside on average in more massive halos and, therefore, are more biased. In the right panel, the mean halo occupation function

$$\langle N(M) \rangle \equiv \langle N_c \rangle_M \big[1 + \langle \mathcal{N}_s \rangle_M \big] \tag{9.7}$$

of the best-fit models is shown as a function of halo mass. The characteristic halo mass cutoff $M_{\min}$ and M_1 increase with the luminosity threshold. Note however, that $\sigma_{\log M}$ is weakly constrained for the faint galaxy samples, which leads to the value $\sigma_{\log M} \approx 0$ adopted here. Fitting the small-scale clustering of galaxies via the HOD approach makes a prediction for the large-scale bias via Eq. (9.2). Ref. [721] pointed out that simple HOD models tend to overestimate the actual large-scale galaxy bias b_1 measured on scales $\gtrsim 60\,h^{-1}$ Mpc. Note that the possible dependence of $\langle N_c \rangle_M$ and $\langle \mathcal{N}_s \rangle_M$ on large-scale density perturbations discussed after Eq. (9.2), which is usually neglected, leads to a correction to the linear bias which could help explain this discrepancy.

Abundance matching techniques are an alternative to HOD approaches, and are based on the assumption that there is a monotonic relation between some galaxy property and the host halo (or subhalo) mass. In particular, this has been applied to the case of the stellar mass for central galaxies [713], or the stellar mass in connection with the subhalo mass for satellite galaxies [722, 723]. In both cases, the underlying assumption is that there is exactly one galaxy per dark matter halo, with the most massive galaxies residing in the most massive (sub)halos. These non-parametric methods, which by construction rely heavily on simulations, predict very well the clustering of observed galaxies (e.g. [724]). Based on this technique, Refs. [725, 726] present an empirical mapping between stellar mass and halo mass at redshifts $z = 0 - 1$, and $z = 0 - 8$,

respectively. These results offer interesting insights into the efficiency of galaxy formation at different mass scales.

HOD and abundance matching approaches provide simple, physically motivated procedures to generate mock galaxy catalogs from N-body simulations [see 727, for a recent review]. However, they clearly rely on oversimplified parametrizations of galaxy formation. Much progress on the understanding of the physics of galaxy formation has been made recently through detailed hydrodynamical simulations of galaxy formation. Recent suites of simulations [728, 729, 730, 731] reproduce fairly well the evolution of early- and late-type galaxies, quasars and their distribution in the cosmic web. They suggest that feedback from supernovae and active galactic nuclei has a significant impact on the local galaxy abundance, and even on halo masses. These effects can influence the clustering of galaxies and thus need to be accounted for in order to achieve the accuracy required by the analysis of forthcoming galaxy surveys. This highlights the crucial difference between the perturbative bias expansion on the one hand, and HOD and abundance matching approaches on the other. The former is agnostic regarding the small-scale processes that shape galaxy formation, while the latter necessarily relies on a specific (simplified) model.

9.2 Astrophysical selection effects

In practice, all galaxy samples collected in sky surveys are selected on some observable property such as luminosity and color. The previous section described how these complications can be modeled through an HOD approach. However, employing an HOD model or other empirical method for connecting galaxies and halos is not necessary for the general bias expansion of Sec. 2 to be valid, as we now show. *This is the case as long as all deterministic and stochastic terms are included at the relevant order, with in general free bias parameters* (see Sec. 2.11 for the result up to third order). We continue to work with the galaxy density in the galaxy rest frame, and defer redshift-space distortions and other projection effects to the next section.

Let $\boldsymbol{\alpha}$ denote a set of physical observables (e.g., type, luminosity, color, half-light radius, shape etc.) that is used to select galaxies. Again, we ignore any projection effects on these observables, and assume that they are true intrinsic properties in the galaxy rest frame. Given a selection function $S(\boldsymbol{\alpha})$, the observed galaxy density at the spacetime coordinate $(\boldsymbol{x}, \tau)$ is

$$n_g^{\text{sel}}(\boldsymbol{x}, \tau) = \int d\boldsymbol{\alpha}\, S(\boldsymbol{\alpha}) n_g(\boldsymbol{x}, \tau; \boldsymbol{\alpha}) = \int d\boldsymbol{\alpha}\, S(\boldsymbol{\alpha}) \overline{n}_g(\tau; \boldsymbol{\alpha}) \left[1 + \delta_g(\boldsymbol{x}, \tau; \boldsymbol{\alpha})\right], \tag{9.8}$$

where we allow for both the mean galaxy density $\overline{n}_g$ and density contrast δ_g to depend on the set of galaxy properties $\boldsymbol{\alpha}$. Let us apply the bias expansion to this "conditional" density contrast $\delta_g(\boldsymbol{x}, \tau; \boldsymbol{\alpha})$:

$$\delta_g(\boldsymbol{x}, \tau; \boldsymbol{\alpha}) = \sum_O \left[b_O(\tau, \boldsymbol{\alpha}) + \varepsilon_O(\boldsymbol{x}, \tau; \boldsymbol{\alpha})\right] O(\boldsymbol{x}, \tau) + \varepsilon(\boldsymbol{x}, \tau; \boldsymbol{\alpha}). \tag{9.9}$$

Here operators and bias parameters are understood to be renormalized. Note that this expansion consistently takes into account that the distribution of the observables $\boldsymbol{\alpha}$ at each point $(\boldsymbol{x}, \tau)$ can depend on the environment via $\{O\}$ as well (for example, a higher fraction of red galaxies in dense regions), in addition to having a stochastic component. Thus, inserting Eq. (9.9) into Eq. (9.8), we obtain

$$n_g^{\text{sel}}(\boldsymbol{x}, \tau) = \overline{n}_g(\tau) \left[1 + \sum_O \left[b_O^S(\tau) + \varepsilon_O^S(\boldsymbol{x}, \tau)\right] O(\boldsymbol{x}, \tau) + \varepsilon^S(\boldsymbol{x}, \tau)\right], \tag{9.10}$$

where

$$\overline{n}_g(\tau) = \int d\boldsymbol{\alpha}\, S(\boldsymbol{\alpha}) \overline{n}_g(\tau; \boldsymbol{\alpha})$$

$$b_O^S(\tau) = \int d\boldsymbol{\alpha}\, S(\boldsymbol{\alpha}) b_O(\tau, \boldsymbol{\alpha})$$

$$\varepsilon_O^S(\boldsymbol{x}, \tau) = \int d\boldsymbol{\alpha}\, S(\boldsymbol{\alpha}) \varepsilon_O(\boldsymbol{x}, \tau; \boldsymbol{\alpha}). \tag{9.11}$$

Thus, the seemingly complicated and nonlinear sample selection effects are consistently taken into account by effective mean bias parameters b_O^S and stochastic fields ε_O^S. While not necessary for the bias expansion to apply, there are still good reasons to split observed galaxy samples by physical properties (or proxies thereof) [732, 733, 734, 735, 736, 737]. This allows for insights into galaxy formation and evolution, by measuring the bias parameters of different galaxy populations (see Fig. 37). Moreover, the relative bias parameters can be measured without sampling variance, and allow for multi-tracer techniques to be applied (Sec. 7.6.2). Alternatively, "marked correlation" analyses, in which galaxies are weighted by some property or "mark," help in quantifying how the galaxy properties $\boldsymbol{\alpha}$ correlate with their large-scale environment [738, 359, 739, 740, 741].

There is, however, an additional selection effect which is nontrivial. In the bias expansion discussed throughout this review, we have written the galaxy density as a 3-scalar, which implies that there are no preferred directions and all tensor indices in bias terms are contracted with each other, such as $K_{ij}K^{ij}$. However, in reality, galaxies are observed through the photons emitted along the line of sight $\hat{\boldsymbol{n}}$ to the observer. Thus, there is a preferred direction involved through the observational selection function, i.e. the probability that a given galaxy is included in the survey sample. Two prominent cases where this can happen are:

- The selection function depends on the orientation of the galaxy [742, 743, 744]. This is particularly relevant for disk galaxies, since a disk is typically dimmer when viewed edge-on compared to face-on due to dust obscuration. Orientations of galaxies in turn tend to correlate with large-scale tidal fields [745, 746, 747, 748], leading to a dependence of the selection probability on the tidal field projected along the line of sight.

- Galaxies are identified through emission or absorption lines, whose observed strength depends on the line of sight due to radiative transfer effects; for example, the escape probability of a resonance line photon depends on the local velocity gradient of matter [749, 750] (see also [751]), which again is proportional to the tidal field projected along the line of sight. This applies in particular to the Lyman-α line.

At linear order, these effects add one additional term, the line-of-sight projection of the tidal field,

$$b_{K_\parallel} \hat{n}^i \hat{n}^j K_{ij}\,. \tag{9.12}$$

In Fourier space, this term becomes $b_{K_\parallel}(\mu^2 - 1/3)\delta(\boldsymbol{k})$, where $\mu = k_i \hat{n}^i/k$. The significance of this term is that it is degenerate with the leading RSD contribution $f\mu^2\delta$, and can thus hamper the use of RSD to measure the growth rate f. However, this degeneracy can be broken by using the galaxy bispectrum as discussed in Refs. [743, 752] for the two cases mentioned above. At higher orders in perturbations and derivatives, the additional terms introduced by the line-of-sight dependence of the selection function multiply rapidly; see [753] for a complete list up to third order. For this reason, keeping such selection effects small is an important consideration in the design of large galaxy surveys.

Finally, we consider the bias of density fields constructed through *intensity mapping*. In this approach, no target selection is done, and spectra are obtained for many lines of sight across a survey region. Then, the flux in each pixel of a spectrum is interpreted as a line intensity for a given transition at the corresponding redshift. This has or will be applied to the 21cm hydrogen hyperfine-structure transition [754], molecular lines such as CO [755], as well as atomic transitions, in particular Hα 6563, as employed by the proposed SPHEREx [633] mission. The resulting intensity map can be seen as a biased tracer of large-scale structure just as galaxies, and obeys an analogous bias expansion. One simply replaces $n_g^{\text{sel}}(\boldsymbol{x}, \tau)$ in Eq. (9.10) with the observed intensity $\mathcal{I}_\nu^{\text{obs}}(\boldsymbol{x}, \tau)$. Due to the fact that it measures line fluxes, intensity mapping is often affected by the radiative transfer effects mentioned above, and thus usually requires the bias Eq. (9.12) to be included in the model predictions.

9.3 Projection effects: from proper to observed galaxy density

In practice, all tracers of large-scale structure are observed via photon arrival directions (right ascension and declination) and redshifts, inferred from the shift in frequency of the observed spectral energy distri-

bution (SED) of the galaxy relative to the rest-frame frequency. That is, we do not have access to the rest-frame galaxy density (or line intensity) directly. Hence, an essential ingredient in the interpretation of large-scale structure is the mapping from rest-frame quantities to observations, which is the subject of this section. We refer to the contributions to the observed galaxy density obtained from transforming the rest-frame galaxy density to observed coordinates as *projection effects*. Ref. [756] provide a concise recent review of the subject.

Using the observed arrival direction, described by the unit vector $\hat{\boldsymbol{n}}$, and observed redshift $\tilde{z}$ of a given galaxy, the standard practice is to assign the galaxy a comoving position using the photon geodesics in a given unperturbed flat FRW spacetime described by fiducial cosmological parameters, which, when using conformal time τ, are simply parametrized as straight lines:

$$(x^0(\tau), \boldsymbol{x}(\tau)) = (\tau_0 - \tau, \bar{\chi}(\tau)\hat{\boldsymbol{n}}) \,, \tag{9.13}$$

where τ_0 is the conformal time at present (i.e., at the time of observation), and $\bar{\tau}(z)$ and $\bar{\chi}(z)$ are, respectively, the conformal-time-redshift relation and the comoving-distance-redshift relation in the adopted fiducial background cosmology. Thus, a galaxy with given $(\tilde{z}, \hat{\boldsymbol{n}})$ is assigned the comoving position

$$(\tilde{\tau}, \tilde{\boldsymbol{x}}) = (\bar{\tau}(\tilde{z}), \bar{\chi}(\tilde{z})\hat{\boldsymbol{n}}) \,. \tag{9.14}$$

In general, this position of course only corresponds to the true physical position of the galaxy in the absence of spacetime perturbations (and if the fiducial cosmology is the true one). On the other hand, in the presence of perturbations as in Eq. (1.2), this is a convenient coordinate (gauge) choice because the variables are directly related to large-scale structure observations. Hereafter, quantities denoted with a tilde are directly related to observables, while barred quantities refer to quantities evaluated for the fiducial background cosmology. Moreover, throughout we assume that the true background cosmology is used to assign apparent positions via Eq. (9.14). In practice, we have to allow for the possibility that the true expansion history is different from the fiducial assumption. These deviations can effectively be taken into account as additional coordinate rescalings, known as *Alcock-Paczyński* distortions [757], which themselves contain information on the expansion history [44, 758, 759].

The calculation is illustrated in Fig. 38. Here, we show two galaxies with inferred positions $\tilde{x}^\mu$ and $\tilde{x}'^\mu$ based on the observables $(\tilde{z}, \hat{\boldsymbol{n}})$ and $(\tilde{z}', \hat{\boldsymbol{n}}')$; the true coordinates of these galaxies, in some given global coordinate system, are $x^\mu \equiv \tilde{x}^\mu + \Delta x^\mu$ and $x'^\mu = \tilde{x}'^\mu + \Delta x'^\mu$, respectively. We define Δx^μ as the spacetime deviation between the true spacetime coordinate and the inferred coordinate of a galaxy. The deviation Δx^μ can be calculated by integrating the geodesic equation from the observer's location to the source, given the photon momentum at the observer specified by $\tilde{z}, \hat{\boldsymbol{n}}$. General-relativistic effects such as Shapiro time delay, Sachs-Wolfe effect, Integrated Sachs-Wolfe effect, lensing magnification as well as redshift-space distortions contribute to Δx^μ. Explicit expressions for Δx^μ can be found in, for example, [761, 762, 756] and [763], at linear and second order in perturbations, respectively. Here, we present the linear-order results in terms of the general perturbed FRW metric, i.e. without restricting the gauge degrees of freedom (also known as *gauge-ready form*)

$$ds^2 = a^2(\tau) \left[-(1 + 2A[\boldsymbol{x}, \tau])d\tau^2 - 2B_i[\boldsymbol{x}, \tau]d\tau dx^i + (\delta_{ij} + h_{ij}[\boldsymbol{x}, \tau]) \, dx^i dx^j \right] \,. \tag{9.15}$$

First, we evaluate the proper time t_F at which a given observed photon was emitted, for an observer comoving with the source. In case of an unperturbed Universe, this is simply given by $t_F = \bar{t}(\tilde{a})$, where $\tilde{a} = 1/(1 + \tilde{z})$ and $\bar{t}(a)$ is the time-scale factor relation in the fiducial background. For convenience, instead of using the proper time t_F, we transform to $\ln a(t_F)$ using the scale factor $a(t)$ in the fiducial background. Moreover, we phrase the departure of the actual "log-scale factor of emission" from the prediction in the unperturbed background, $\ln \tilde{a} \equiv -\ln(1 + \tilde{z})$, through

$$\mathcal{T}(\tilde{\boldsymbol{x}}, \tilde{\tau}) \equiv \ln\left(\frac{a(t_F[\tilde{\boldsymbol{x}}, \tilde{\tau}])}{\tilde{a}}\right) = \Delta \ln a(\tilde{\boldsymbol{x}}, \tilde{\tau}) + \tilde{H} \int_0^{\tilde{\tau}} A[\boldsymbol{x}, \tau']a(\tau')d\tau' \,, \tag{9.16}$$

209
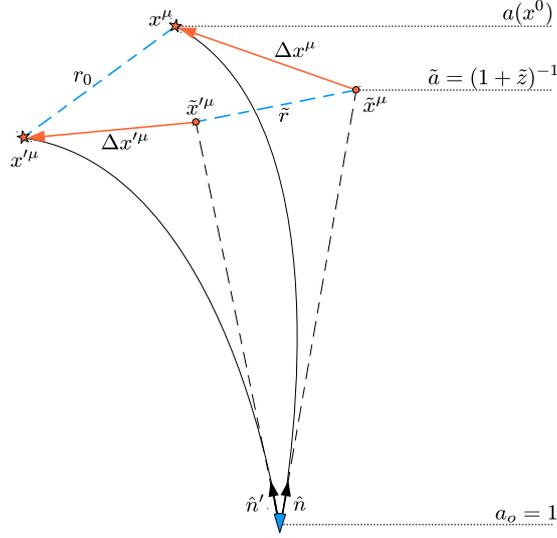
Figure 38: Illustration of the projection effects, i.e. the mapping from observed photon directions and redshifts to the galaxy rest frame. The actual geodesics of photons emitted from two galaxies at x^μ and x'^μ are depicted as solid lines. The observer at the bottom of the plot infers radial distances and angular positions of the galaxies based on the observed redshift and arrival direction of photons, assigning them the spacetime positions $\tilde{x}^\mu$ and $\tilde{x}'^\mu$ respectively. The deviation four-vectors Δx^μ and $\Delta x'^\mu$ can be calculated by following the photon geodesics, taking into account the velocity four-vectors of each galaxy and the observer. Note that the deflection of the photon geodesics is greatly exaggerated here for illustration purposes. From [760].

where $\tilde{H} \equiv H(\tilde{a})$, and

$$\Delta \ln a \equiv A_o - A + v_\parallel - v_{\parallel o} - \int_0^{\tilde{\chi}} d\chi \, \frac{\partial}{\partial \tau} \left[A - B_\parallel - \frac{1}{2} h_\parallel \right]_{\chi \hat{\boldsymbol{n}}, \, \tau_0 - \chi} - H_0 \int_0^{\tau_o} A(\boldsymbol{0}, \tau') a(\tau') d\tau' , \qquad (9.17)$$

where all quantities without arguments are evaluated at $(\tilde{\boldsymbol{x}}, \tilde{\tau})$, while a subscript o indicates a quantity evaluated at the observer: $v_{\parallel o} \equiv v_\parallel(\boldsymbol{0}, \tau_0)$. The subscript $\parallel$ denotes quantities projected onto the line of sight; for example, $B_\parallel \equiv B_i \hat{n}^i$, $h_\parallel \equiv h_{ij} \hat{n}^i \hat{n}^j$. Thus, $\mathcal{T}$ is the time shift, phrased in terms of $\ln a$, between a constant-observed-redshift surface (defined for a fixed central observer) and a constant-proper-time surface. In Eq. (9.16), $\Delta \ln a$ gives the difference $\mathcal{H} \delta \tau$ in coordinate time, while the second term maps coordinate time to proper time. Note that $\mathcal{T}$ is an observable and gauge-invariant: if the observed source is a clock, i.e. communicates its proper time since the Big Bang to the distant observer, then $\mathcal{T}$ can be measured directly. Indeed, the observed CMB temperature fluctuation on large scales, outside of the sound horizon, is exactly given by $\mathcal{T}$ [764]. $\Delta \ln a$ on the other hand is a coordinate-dependent and unobservable quantity. Only if a gauge choice is made such that $A = 0$ in Eq. (9.15) does $\Delta \ln a = \mathcal{T}$ become an observable.

In addition, the spatial displacement, decomposed into components parallel $(\Delta x_\parallel = \hat{n}^i \Delta x_i)$ and perpendicular to the line of sight $(\Delta x_\perp^i = \mathcal{P}^{ij} \Delta x_j$, where $\mathcal{P}^{ij} \equiv \delta^{ij} - \hat{n}^i \hat{n}^j$ is the projection operator on the sky) can be written as

$$\Delta x_\parallel = - \int_0^{\tau_o} A(\boldsymbol{0}, \tau') a(\tau') d\tau' + \int_0^{\tilde{\chi}} d\chi \left[A - B_\parallel - \frac{1}{2} h_\parallel \right]_{\chi \hat{\boldsymbol{n}}, \, \tau_0 - \chi} - \frac{1 + \tilde{z}}{H(\tilde{z})} \Delta \ln a . \qquad (9.18)$$

$$\Delta x_\perp^i = \left[\frac{1}{2} \mathcal{P}^{ij} (h_{jk})_o \hat{n}^k + B_{\perp o}^i - v_{\perp o}^i \right] \tilde{\chi}$$

$$- \int_0^{\tilde{\chi}} d\chi \left[\frac{\tilde{\chi}}{\chi} \left(B_\perp^i + \mathcal{P}^{ij} h_{jk} \hat{n}^k \right) + (\tilde{\chi} - \chi) \partial_\perp^i \left(A - B_\parallel - \frac{1}{2} h_\parallel \right) \right]_{\chi \hat{\boldsymbol{n}}, \, \tau_0 - \chi} . \qquad (9.19)$$

Here, the spacetime arguments follow those in Eq. (9.17). We see that $\Delta \ln a$ appears in the line-of-sight displacement. Further, the combination of metric perturbations $A - B_\parallel - h_\parallel/2$ is the gravitational lensing

potential in a general gauge, which reduces to $\Phi + \Psi$ in conformal-Newtonian gauge [Eq. (1.2)]. The dominant terms in the displacements on small scales are those which involve spatial derivatives of the metric potentials. This is $(v_\parallel - v_{\parallel o})$, which appears in $\Delta \ln a$ and is responsible for redshift-space distortions, and $\partial_\perp^i (A - B_\parallel - h_\parallel/2)$ in $\Delta x_\perp^i$ which is responsible for lensing magnification effects.

The displacement $\tilde{x}^\mu \to x^\mu = \tilde{x}^\mu + \Delta x^\mu$ defines a coordinate transformation from observed to true positions. We can now derive how the galaxy density transforms under this coordinate transformation. Specifically, the physical (rather than comoving) rest-frame galaxy density n_g is the 0-component of the galaxy four-momentum $n_g u_g^\mu = n_g u^\mu$ (since we restrict to large scales here, we neglect velocity bias, see Sec. 2.7). The number of galaxies within a volume V on the past light-cone of the observer, defined in terms of observed Cartesian coordinates $\tilde{x}$, is then given by an integral of the three-form that is dual to $n_g u^\mu$. In components, this becomes

$$N(V) = \int_V d^3\tilde{x} \ \sqrt{-g(x)} \, n_g(x) \varepsilon_{\mu\nu\rho\sigma} u^\mu(x) \frac{\partial x^\nu}{\partial \tilde{x}^1} \frac{\partial x^\rho}{\partial \tilde{x}^2} \frac{\partial x^\sigma}{\partial \tilde{x}^3} \, , \tag{9.20}$$

where $\sqrt{-g(x)}$ is the determinant of the metric, $n_g(x)$ is the physical number density of galaxies at spacetime position x, $u^\mu = dx^\mu/dt_F = a^{-1}(1 - A, v^i)$ the galaxy velocity four-vector, and $\varepsilon_{\mu\nu\rho\sigma}$ the Levi-Civita symbol. Eq. (9.20) is fully nonlinear, and we will evaluate it in two limits: first, restricting to linear order in perturbations but including all relativistic terms; second, working to nonlinear order in perturbations but restricting the projection effects to those relevant on small scales, i.e. those terms that involve two spatial derivatives on metric perturbations. As discussed in Sec. 2.9, these two limits essentially cover the parameter space where projection effects are numerically important.

9.3.1 Observed galaxy density contrast at linear order

Combining the expression Eq. (9.20) with Eqs. (9.16)–(9.19), we can now obtain the fully relativistic expression for the observed galaxy density contrast at linear order in perturbations. For this, we need an expression for the galaxy density $n_g(x)$, which is, of course, dependent on the coordinate system used to evaluate Eq. (9.20). Rather than expressing n_g in terms of the galaxy density perturbation δ_g in some arbitrary gauge, which is of course absolutely legitimate, we choose to fix coordinates to the *constant-observed-redshift* ("or") *gauge*. The reason is that the mean density of galaxies $\bar{n}_g$ is measured at fixed observed redshift. For this reason, observations of the galaxy overdensity are naturally in the "or" gauge, simplifying the calculation both practically and conceptually.

We thus write n_g in terms of the mean *comoving* number density $a^3 \bar{n}_g$ and the perturbation δ_g^{or} to the comoving number density, in the constant-observed-redshift gauge, as

$$a^3 n_g(x) = \tilde{a}^3 \bar{n}_g(\tilde{z}) \left[1 + \delta_g^{\mathrm{or}}(\boldsymbol{x}, \tilde{z}) \right] \, , \tag{9.21}$$

where $\tilde{z}$ is the observed redshift corresponding to the spacetime location x, and $\tilde{a} = 1/(1 + \tilde{z})$. Eq. (9.21) can be understood as the definition of δ_g^{or}. At linear order in perturbations, we can neglect the distinction between $\boldsymbol{x}(\tilde{x})$ and $\tilde{\boldsymbol{x}}$ in the argument of δ_g^{or}, since the latter is already first order. Then, the right-hand side of Eq. (9.20) becomes

$$N(V) = \int_V d^3\tilde{\boldsymbol{x}} \left(1 + A + \frac{h}{2} \right) \tilde{a}^3 \bar{n}_g(\tilde{z}) \left[1 + \delta_g^{\mathrm{or}}(\tilde{\boldsymbol{x}}, \tilde{z}) \right] \left((1 - A) \left| \frac{\partial x^i}{\partial \tilde{x}^j} \right| + v_\parallel \right) . \tag{9.22}$$

The observationally inferred galaxy number density $\tilde{n}_g$ is defined by evaluating Eq. (9.20) on the unperturbed (fiducial) background, yielding

$$N(V) = \int_V d^3\tilde{\boldsymbol{x}} \ \tilde{a}^3 \tilde{n}_g(\tilde{\boldsymbol{x}}, \tilde{z}) . \tag{9.23}$$

By equating Eq. (9.22) and Eq. (9.23), we find the observed galaxy density contrast as

$$\tilde{\delta}_g(\tilde{\boldsymbol{x}}, \tilde{z}) \equiv \frac{\tilde{n}_g(\tilde{\boldsymbol{x}}, \tilde{z})}{\bar{n}_g(\tilde{z})} - 1 = \delta_g^{\mathrm{or}}(\tilde{\boldsymbol{x}}, \tilde{z}) + \frac{h}{2} + \partial_\parallel \Delta x_\parallel + \frac{2\Delta x_\parallel}{\tilde{\chi}} - 2\hat{\kappa} + v_\parallel . \tag{9.24}$$

Here, $h \equiv \delta^{ij} h_{ij}$, and we have expanded the Jacobian through

$$\left| \frac{\partial x^i}{\partial \tilde{x}^j} \right| = 1 + \frac{\partial \Delta x^i}{\partial \tilde{x}^i} = 1 + \partial_\parallel \Delta x_\parallel + \frac{2 \Delta x_\parallel}{\tilde{\chi}} - 2\hat{\kappa} \,, \tag{9.25}$$

where the *coordinate* (i.e. gauge-dependent) lensing convergence $\hat{\kappa}$ is defined as the divergence on the sky of the perpendicular displacement,

$$\hat{\kappa} = -\frac{1}{2} \mathcal{P}^{ij} \partial_i \Delta x_{\perp j} \,. \tag{9.26}$$

All contributions in Eq. (9.24) apart from δ_g^{or} are the induced, apparent modulation of the galaxy abundance due to volume distortion effects.

Next, we have to relate δ_g^{or} in Eq. (9.24) to the matter density through a bias relation. The galaxy density contrast δ_g^{or} in Eq. (9.24) is defined in the constant-observed-redshift slicing. On the other hand, as discussed in detail in Sec. 2.9, the linear bias relation between galaxy density contrast and matter density contrast is valid on *constant-proper-time* ("pt") slices: $\delta_g^{\mathrm{pt}} = b_1 \delta^{\mathrm{pt}}$. This is because in the large-scale limit, galaxies only know about the local age of the Universe and the local matter density.[24] At linear order, the transformation of the galaxy density between the two gauges is completely determined by the time shift between them. This shift is precisely the observable $\mathcal{T}$ that we have derived in Eq. (9.16). Then, the relation between $\delta_g^{\mathrm{or}}(\tilde{\boldsymbol{x}}, \tilde{z})$ and the galaxy density perturbation $\delta_g^{\mathrm{pt}} = b_1 \delta^{\mathrm{pt}}$ in the constant-proper-time (or synchronous) gauge is given by a standard, linear gauge transformation,

$$\delta_g^{\mathrm{or}}(\tilde{\boldsymbol{x}}, \tilde{z}) = b_1 \, \delta^{\mathrm{pt}}(\tilde{\boldsymbol{x}}, \tilde{z}) + b_e \mathcal{T}(\tilde{\boldsymbol{x}}, \tilde{z}) \,, \qquad b_e \equiv \frac{d \ln(a^3 \overline{\mathrm{n}}_g)}{d \ln a} \,, \tag{9.27}$$

where we have introduced the dimensionless parameter b_e quantifying the evolution of the mean comoving number density of galaxies. Note that this relation only involves observable quantities, so that both b_1 and b_e are well defined and gauge-invariant. It also serves as the unambiguous starting point for extending the bias relation to higher order in perturbations, for example by adding a term $(b_2/2)(\delta^{\mathrm{pt}})^2$ to the right-hand side. For halos following a universal mass function, Ref. [96] derived $b_e = \delta_{\mathrm{cr}} f(b_1 - 1)$.

Finally, δ^{pt} is related to the matter density perturbation δ in the chosen gauge through

$$\delta^{\mathrm{pt}} = \delta + 3\tilde{H} \int_0^{\tilde{\tau}} A(\boldsymbol{x}, \tau) a(\tau) d\tau \,. \tag{9.28}$$

Combining the last two equations, we find the galaxy density contrast on the constant-observed-redshift slice in terms of the density contrast in an arbitrary gauge as

$$\delta_g^{\mathrm{or}}(\tilde{\boldsymbol{x}}, \tilde{z}) = b_1 \left[\delta + 3\tilde{H} \int_0^{\tilde{\tau}} A(\tilde{\boldsymbol{x}}, \tau) a(\tau) d\tau \right] + b_e \mathcal{T}. \tag{9.29}$$

This yields our final expression:

$$\tilde{\delta}_g(\tilde{\boldsymbol{x}}, \tilde{z}) = b_1 \left[\delta + 3\tilde{H} \int_0^{\tilde{\tau}} A(\tilde{\boldsymbol{x}}, \tau) a(\tau) d\tau \right] + b_e \mathcal{T} + \frac{1}{2} h + \partial_{\tilde{\chi}} \Delta x_\parallel + \frac{2 \Delta x_\parallel}{\tilde{\chi}} - 2\hat{\kappa} + v_\parallel \,. \tag{9.30}$$

Here, the line-of-sight derivative of the longitudinal displacement is given by

$$\begin{aligned} \partial_{\tilde{\chi}} \Delta x_\parallel = {} & A - B_\parallel - \frac{1}{2} h_\parallel - H(\tilde{z}) \left(\frac{\partial}{\partial \tilde{z}} \frac{1 + \tilde{z}}{H(\tilde{z})} \right) \Delta \ln a \\ & - \frac{1 + \tilde{z}}{H(\tilde{z})} \left(-\partial_\parallel A + \partial_\parallel v_\parallel - \frac{\partial}{\partial \tau} v_\parallel + \frac{1}{2} \frac{\partial}{\partial \tau} h_\parallel + \frac{\partial}{\partial \tau} B_\parallel \right) \,. \end{aligned} \tag{9.31}$$

---

[24]This assumes Gaussian initial conditions, and that there are no additional degrees of freedom relevant on large scales, such as dark energy perturbations, or modified gravitational forces. The generalizations are discussed in Sec. 7 and Sec. 8, respectively, and can easily be included here.

One subtlety we have neglected so far is that observational selection effects can modify the observed galaxy density, Eq. (9.30). Usually, surveys observe galaxies above a certain apparent flux, or magnitude threshold. Weak lensing magnifies/de-magnifies the flux of the source galaxies and therefore induces another contribution to the observed galaxy density (*magnification bias* [765]). For a population of galaxies at fixed redshift $\tilde{z}$ with cumulative luminosity function $\bar{\mathrm{n}}_g(> L_{\min})$, we define

$$\mathcal{Q} \equiv -\frac{d\ln\bar{\mathrm{n}}_g(> L_{\min})}{d\ln L_{\min}}\,. \tag{9.32}$$

More generally, other selection effects, such as galaxy size, can contribute to $\mathcal{Q}$ [766]. In that case, the coefficient should be defined as $\mathcal{Q} = d\ln\bar{\mathrm{n}}_g/d\mathcal{M}$, where $\mathcal{M}$ is the gauge-invariant magnification. Then, magnification bias adds a contribution $\mathcal{Q}\mathcal{M}$ to $\tilde{\delta}_g$, where

$$\mathcal{M} = -2\Delta\ln a - \frac{1}{2}(h - h_\parallel) - \frac{2}{\tilde{\chi}}\Delta x_\parallel + 2\hat{\kappa}\,. \tag{9.33}$$

Note that we neglect the effect from the evolution of intrinsic luminosity of the galaxies which can in principle contribute to the magnification (see [760] for the complete expression); this contribution is typically much smaller than $\mathcal{M}$ itself. We finally obtain the observed density contrast including magnification bias as

$$\tilde{\delta}_g(\tilde{\boldsymbol{x}}, \tilde{z}) = b_1\left[\delta + 3\tilde{H}\int_0^{\tilde{\tau}} A(\tilde{\boldsymbol{x}}, \tau)a(\tau)d\tau\right] + b_e\mathcal{T} + 2\mathcal{Q}\tilde{H}\int_0^{\tilde{\tau}} A(\tilde{\boldsymbol{x}}, \tau)a(\tau)d\tau + \frac{1}{2}(1 - \mathcal{Q})h + \frac{\mathcal{Q}}{2}h_\parallel$$
$$+ \partial_{\tilde{\chi}}\Delta x_\parallel + (1 - \mathcal{Q})\frac{2}{\tilde{\chi}}\Delta x_\parallel + 2(\mathcal{Q} - 1)\hat{\kappa} + v_\parallel\,. \tag{9.34}$$

Eq. (9.34) provides the complete result for the observed overdensity of a tracer at linear order in a general gauge. When restricted to conformal-Newtonian gauge, this agrees with [190, 95] (note the discussion around Eq. (31) of the former reference); restricting to synchronous-comoving gauge yields the results derived in [96]. Here, we have assumed a sharp source redshift, but the projection over a finite redshift bin is straightforward.

Assuming that the coefficients b, b_e, $\mathcal{Q}$ are all of order unity, the various terms in Eq. (9.34) can be ranked in terms of relative importance according to their scaling, relative to δ, with $\mathcal{H}/k$ in Fourier space. The largest terms, "order 1", are in conformal-Newtonian gauge given by

$$\tilde{\delta}_g^{\mathcal{O}(1)} = b_1\,\delta + \frac{1 + \tilde{z}}{\tilde{H}}\partial_\parallel v_\parallel + 2(\mathcal{Q} - 1)\hat{\kappa}\,. \tag{9.35}$$

Note that in conformal-Newtonian gauge, $\delta = \delta^{\mathrm{pt}} + \mathcal{O}(\mathcal{H}^2/k^2)$. Eq. (9.35) is the standard small-scale result for the apparent galaxy overdensity, including the leading redshift-space distortion [47] and magnification bias. The subleading corrections in Eq. (9.34) scale as $i\mathcal{H}/k$ ("velocity-type") and $(\mathcal{H}/k)^2$ ("potential-type"). Fig. 39 shows the angle-averaged auto-power spectrum of the galaxy density including relativistic projection effects [Eq. (9.34), solid line], and in the small-scale approximation [Eq. (9.35), using δ^{pt}, dotted line]. We have dropped terms that are integrated along the line of sight. These cannot be represented by a three-dimensional power spectrum. Moreover, on such large scales, employing the flat-sky approximation is not sufficient (we will generalize this in Sec. 9.4). Thus, Fig. 39 is only meant as an order-of-magnitude illustration of the relativistic projection effects, which clearly become numerically relevant only for $k/k_H \equiv k/\mathcal{H} \lesssim 10$.

The power spectrum of the relativistic linear galaxy density perturbation Eq. (9.34), again neglecting all contributions that involve integrals along the line of sight, contains three types of contributions beyond the small-scale limit Eq. (9.35): the auto-correlation of velocity-type terms, which scales as $(\mathcal{H}/k)^2 P_{\mathrm{L}}(k)$, compared to the leading contribution $\propto P_{\mathrm{L}}(k)$; the cross-correlation of potential-type terms with leading terms, which obeys the same scaling; and the auto-correlation of potential-type terms, which scales as $(\mathcal{H}/k)^4 P_{\mathrm{L}}(k)$. Note that the cross-correlation of velocity-type terms with either potential- or density-type terms vanishes by symmetry when considering the auto-correlation of any tracer. The first two contributions
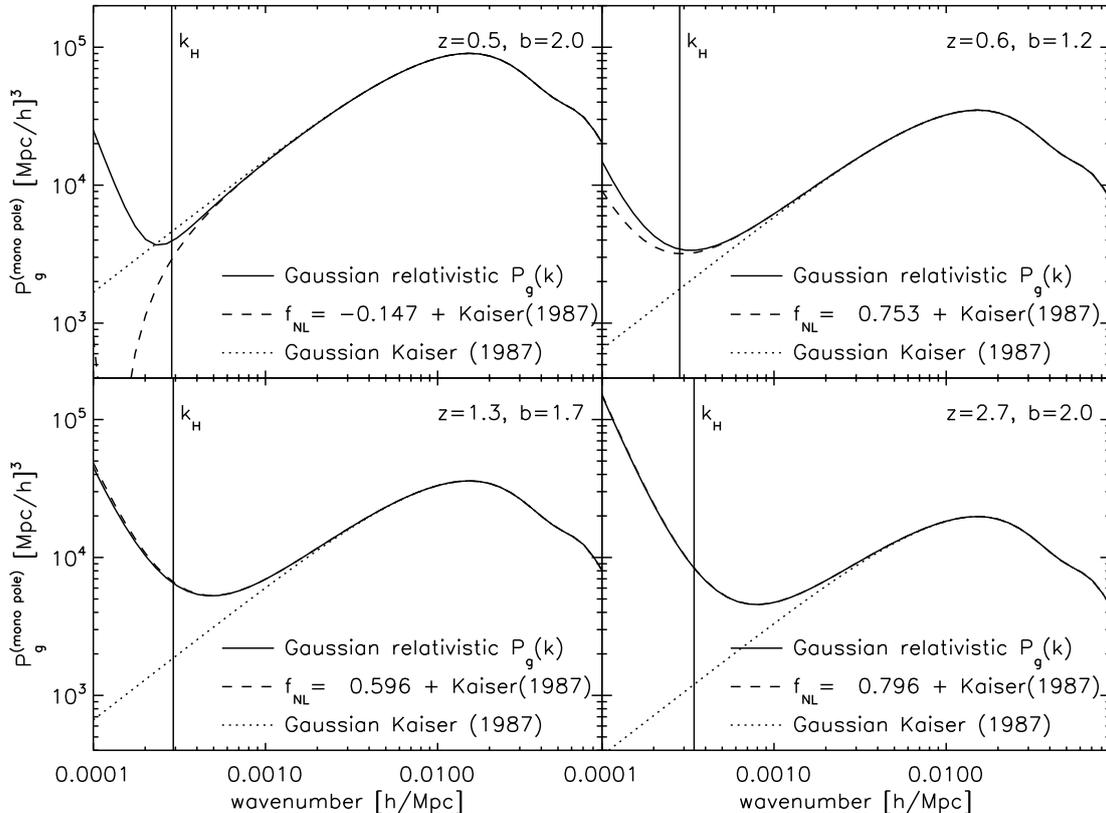
Figure 39:   Linear galaxy power spectrum including relativistic projection effects following Eq. (9.34) (solid lines), and corresponding prediction keeping only terms relevant on small scales, Eq. (9.35) (dotted). The four panels show different values of linear bias parameters (with $b \equiv b_1$) and redshift. b_e is calculated by using the prediction for the universal mass function, $b_e = \delta_{\rm cr} f(b_1 - 1)$ [96]. In each case, we perform an angle-average over $\hat{\boldsymbol{k}}$ (power spectrum monopole), and drop contributions that are integrated along the line of sight (see discussion in the text). Vertical lines show the comoving horizon wavenumber ($k_H = \mathcal{H}$) at each redshift. For comparison, we also show the galaxy power spectrum in the small-scale approximation but including the scale-dependent bias induced by local-type PNG via Eq. (7.94), with a value of $f_{\rm NL}$ adjusted in each case to match the amplitude of the relativistic projection effects (dashed lines). Depending on bias and redshift, the relativistic projection effects are of the same order as the scale-dependent bias for PNG with $f_{\rm NL} \lesssim 1$. From [96].

have the same k-dependence on large scales as the leading contribution to $P_{gg}(k)$ from the scale-dependent bias $\Delta b(k)$ induced by local-type PNG [68], Eq. (7.94). We can thus find the effective $f_{\rm NL}$ that would lead to a scale-dependent term of similar magnitude. These results are also shown in Fig. 39. We see that relativistic projection effects amount to the effect of local primordial non-Gaussianity with $f_{\rm NL} \lesssim 1$ [96, 767] for typical expected values of $b_1, b_e, \mathcal{Q}$. Given our forecasts in Sec. 7.6.1, the projection effects of potential-type are thus expected to be marginally detectable in future galaxy surveys. The velocity-type contributions $\propto \mathcal{H}/k$, which come from terms involving $v_\parallel$ and $\partial_\parallel \Psi$ in case of conformal-Newtonian gauge, are larger and more easily detectable [614, 768, 769]. As mentioned above however, they cancel out in the auto-correlation of tracers, and necessitate the use of two different tracers to measure the dipole of their cross-power spectrum. Note that projection effects will also amount to a non-zero $f_{\rm NL} \lesssim 1$ in the galaxy bispectrum, although the exact amplitude of this effect is still being debated in the literature [763, 770, 771, 772, 773].

Ref. [774] derived the contribution of tensor modes to the observed galaxy clustering statistics, which, at linear order, only enter through projection effects. Unfortunately, for currently allowed amplitudes of

primordial gravitational waves, these contributions are much smaller than even the relativistic projection effects from scalar perturbations shown in Fig. 39.

9.3.2 Nonlinear galaxy density contrast in redshift space

We now turn to the projection effects on small scales, where the "potential-type" terms discussed at the end of the Sec. 9.3.1 can be neglected. The dominant projection contributions to the galaxy density then come from the line-of-sight velocity $v_\parallel$ and its derivatives, and from the lensing convergence $\hat{\kappa}$. In the following, we will neglect the contribution from $\hat{\kappa}$, which, while of the same order in derivatives as the other leading terms, is suppressed by the integration over the line of sight. However, this effect can be taken into account in a straightforward manner [775, 776, 777].

On the other hand, on small scales linear perturbation theory no longer applies. Hence, we now evaluate Eq. (9.20) without expanding at linear order in perturbations, but instead making the approximations

$$\mathcal{T} \to 0, \ \Delta x_\parallel \to -\frac{1+\tilde{z}}{H(\tilde{z})} v_\parallel, \ \Delta x_\perp \to 0 \,. \tag{9.36}$$

Here, we have also dropped the observer's velocity $v_{\parallel o}$, since it only contributes to the dipole of the galaxy density and can thus also be neglected when focusing on small scales. Further, we will continue to denote the galaxy velocity as $\boldsymbol{v}$ for clarity, although this can now differ from the matter velocity due to velocity bias; the matter velocity will never appear explicitly in the following, however. With these assumptions, Eq. (9.20) becomes

$$N(V) = \int_V d^3\tilde{\boldsymbol{x}} \, a^3 \overline{\mathrm{n}}_g(z)(1 + \delta_g(\boldsymbol{x}[\tilde{\boldsymbol{x}}])) \left| \frac{\partial x^i}{\partial \tilde{x}^j} \right| \equiv \int_V d^3\tilde{\boldsymbol{x}} \, \tilde{a}^3 \overline{\mathrm{n}}_g(\tilde{z})(1 + \tilde{\delta}_g(\tilde{\boldsymbol{x}}, \tilde{z})) \,, \tag{9.37}$$

where, from Eq. (9.36)

$$\tilde{x}^i = x^i + u_\parallel(\boldsymbol{x})\hat{n}^i \tag{9.38}$$

is the (purely spatial) coordinate shift between real space (x^i) and redshift space ($\tilde{x}^i$) coordinates, using the scaled galaxy velocity $\boldsymbol{u}(\boldsymbol{x}) \equiv \boldsymbol{v}(\boldsymbol{x})/\mathcal{H}$, and $u_\parallel(\boldsymbol{x}) \equiv \hat{\boldsymbol{n}} \cdot \boldsymbol{u}(\boldsymbol{x})$. In general, the inverse function $\boldsymbol{x}(\tilde{\boldsymbol{x}})$ of Eq. (9.38) is multi-valued. Therefore, a simply connected volume V in observed coordinates does not in general correspond to a simply connected region in real space. This multi-to-one mapping of the redshift-space distortion can happen for galaxies residing in massive halos which exhibit large peculiar (virial) velocities. In the following, we will ignore this non-perturbative effect, since our focus is on the perturbative description of galaxy clustering. Then, we can invert Eq. (9.38) and write the observed galaxy density perturbation as

$$1 + \tilde{\delta}_g(\tilde{\boldsymbol{x}}) = [1 + \delta_g(\boldsymbol{x})] \left| \frac{\partial x^i}{\partial \tilde{x}^j} \right| = [1 + \delta_g(\boldsymbol{x})] \left| \delta_i^j + \hat{n}^j \frac{\partial}{\partial x^i} u_\parallel(\boldsymbol{x}) \right|^{-1} \,. \tag{9.39}$$

Here and in the following, the galaxy density and velocity will always be evaluated at the observed redshift $\tilde{z}$, i.e. at coordinate time $\bar{\tau}(\tilde{z})$. Hence, we drop the time argument in what follows for clarity. Sylvester's theorem allows us to evaluate the determinant via

$$\left| \delta_i^j + \hat{n}^j \frac{\partial}{\partial x^i} u_\parallel(\boldsymbol{x}) \right| = 1 + \hat{n}^j \partial_j u_\parallel(\boldsymbol{x}) = 1 + \partial_\parallel u_\parallel(\boldsymbol{x}) \,, \tag{9.40}$$

and we obtain

$$\tilde{\delta}_g(\tilde{\boldsymbol{x}}) = \frac{1 + \delta_g(\boldsymbol{x})}{1 + \partial_\parallel u_\parallel(\boldsymbol{x})} - 1 \,. \tag{9.41}$$

This result is the *fully nonlinear* expression of the redshift-space galaxy density contrast in configuration space, as long as Eq. (9.38) describes a one-to-one mapping from real to redshift space. Finally, we evaluate the right-hand side of the above equation using the quantities at the observed (redshift-space) coordinate. For some scalar function $f(\boldsymbol{x})$ (e.g., $\delta_g(\boldsymbol{x})$), expanding $\boldsymbol{x}$ around $\tilde{\boldsymbol{x}}$ via Eq. (9.38) yields

$$f(\boldsymbol{x}) = f(\tilde{\boldsymbol{x}}) - \left[1 - \partial_\parallel u_\parallel\right] u_\parallel \partial_\parallel f(\tilde{\boldsymbol{x}}) + \frac{1}{2} u_\parallel^2 \partial_\parallel^2 f(\tilde{\boldsymbol{x}}) + \mathcal{O}(u_\parallel^3) \,, \tag{9.42}$$

215

where all fields on the right-hand side are evaluated at $\tilde{\boldsymbol{x}}$, and the $1 - \partial_\parallel u_\parallel$ factor comes from expanding the argument of $u_\parallel(\boldsymbol{x})$. Note that, even though the velocity appears without a spatial derivative here,[25] it always enters in combination with a spatial derivative acting on another perturbative quantity. Hence, terms of order $u_\parallel \partial_\parallel$ in the expansion Eq. (9.42) are of the same order in derivatives as $\partial_\parallel u_\parallel$.

We can now expand the right-hand side of Eq. (9.41) to obtain the perturbative expression for the observed *galaxy density in redshift space*, i.e. including the leading projection effects on small scales. Up to third order in perturbations, counting each power of δ as well as $u_\parallel$, Eq. (9.41) becomes

$$\tilde{\delta}_g(\tilde{\boldsymbol{x}}) = \delta_g(\boldsymbol{x}) + [1 + \delta_g(\boldsymbol{x})]\left\{-\partial_\parallel u_\parallel(\boldsymbol{x}) + \left[\partial_\parallel u_\parallel(\boldsymbol{x})\right]^2\right\} - \left[\partial_\parallel u_\parallel(\boldsymbol{x})\right]^3 + \mathcal{O}(\delta^4)\,. \tag{9.43}$$

Applying the coordinate transformation relation in Eq. (9.42), we obtain the third-order galaxy density perturbation in redshift space as

$$\tilde{\delta}_g = \delta_g - \partial_\parallel\left[u_\parallel\left(1 + \delta_g\right)\right] + \frac{1}{2}\partial_\parallel^2\left[u_\parallel^2\left(1 + \delta_g\right)\right] - \frac{1}{6}\partial_\parallel^3\left(u_\parallel^3\right) + \mathcal{O}(\delta^4)\,, \tag{9.44}$$

where all quantities are evaluated at the redshift-space coordinate, $\tilde{\boldsymbol{x}}$. This expression agrees with the Fourier-space expression derived in [778, 250],

$$\tilde{\delta}_g(\boldsymbol{k}) = \delta_g(\boldsymbol{k}) + \int d^3\boldsymbol{x}\, e^{-i\boldsymbol{k}\cdot\boldsymbol{x}}\left(e^{-ik_\parallel u_\parallel(\boldsymbol{x})} - 1\right)[1 + \delta_g(\boldsymbol{x})]\,. \tag{9.45}$$

This in turn is formally equivalent in configuration space to a series expansion valid at all orders,

$$\tilde{\delta}_g = \delta_g + \sum_{n=1}^{\infty}\frac{(-1)^n}{n!}\partial_\parallel^n\left[u_\parallel^n(1 + \delta_g)\right]\,. \tag{9.46}$$

The contributions in Eq. (9.44) that involve $u_\parallel$ are known as *redshift-space distortions (RSD)*. We now derive the effect of RSD on observed galaxy statistics, in particular the leading-order power spectrum. This calculation simplifies considerably in the flat-sky limit, i.e. when the line of sight $\hat{\boldsymbol{n}}$ is taken to be a constant unit vector. We will describe approaches to galaxy clustering on the full sky in the next section. First, at linear order in perturbation theory, we have

$$u_\parallel(\boldsymbol{k}, \tau) = i\frac{\boldsymbol{k}\cdot\hat{\boldsymbol{n}}}{k^2}f\,\delta^{(1)}(\boldsymbol{k}, \tau)\,, \tag{9.47}$$

where $f = d\ln D/d\ln a$ is the linear growth rate. The leading-order prediction for the galaxy power spectrum in redshift space in this limit ("Kaiser formula") is given by

$$P_{g,s}^{\mathrm{LO}}(k, \mu) = (b_1 + f\mu^2)^2 P_{\mathrm{L}}(k) + P_\varepsilon^{\{0\}}\,, \tag{9.48}$$

where $\mu \equiv \hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{n}}$ is the cosine of the wavevector with the line of sight. This expression was first derived by [47]; the corresponding derivation in configuration space can be found in [779]. Note that we use the term "configuration space" for the complement to Fourier space, to distinguish from "real space," which is commonly used to denote the rest-frame galaxy density. These results correspond to the auto-power spectrum and correlation function, respectively, of Eq. (9.35) in the flat-sky limit and when neglecting magnification bias. Note that at this order, RSD do not add any free parameters to the galaxy power spectrum. Instead, the anisotropy (μ-dependence) of the observed galaxy power spectrum can be used to constrain the growth rate f. Thus, a measurement of RSD on large scales yields constraints on the growth of structure without the complications of bias [780, 781].

---

[25]Strictly speaking, $u_\parallel$ denotes the relative velocity between source and observer [cf. Eq. (9.17)], which is boost-invariant and observable.

The idea of using the observed anisotropic distribution of galaxies in redshift space as an indicator of velocity structure was first proposed by [45] and further extended by [46], who showed that the mean anisotropy of the galaxy two-point function is a probe of the matter density parameter Ω_m, via the growth rate f. This technique has been applied to real data, providing constraints on modified gravity and non-canonical dark energy [782, 783, 784, 785, 786, 787]. The multi-tracer method discussed in Sec. 7.6.2 can also be applied to RSD, improving constraints on the growth rate, provided that the stochasticity $P_\varepsilon^{\{0\}}$ of the galaxy samples is well understood [643, 342, 788, 614, 789]. This has recently been applied to the GAMA survey sample [790]. Ref. [791] provides a pedagogical review on redshift-space distortions at linear order.

It is important to stress that Eq. (9.48) can only be considered a clean probe of velocities if the real-space galaxy density δ_g itself does not depend on $\partial_\parallel u_\parallel$. However, since, at linear order, $\partial_\parallel u_\parallel \propto \partial_\parallel^2 \Phi \propto K_{ij}\hat{n}^i\hat{n}^j$, this is only true if the galaxy density does not depend on the tidal field projected along the line of sight. As we have seen in Sec. 9.2, such a dependence can be induced by selection or radiative-transfer effects. If those are present, leading-order RSD no longer provide a direct probe of velocities. In case of the Lyman-α forest, for example, such radiative-transfer effects are important [see, e.g., 792, 793, 794, 795, 796, 797, 798, 799, 800, 801, 802], as the optical depth of Lyman-α photons depends strongly on the velocity gradient along the line of sight.

Going beyond linear order, other references considered the galaxy three-point function [803] and bispectrum [804, 805] in redshift space. Ref. [806] presented the calculation of RSD in the two-point function using the Zel'dovich approximation. The formulation of RSD in Lagrangian perturbation theory can be found in [807, 808]. For the galaxy power spectrum, the expressions derived in either [778] ("Scoccimarro" model) or [809] ("TNS" model) are widely applied for the interpretation of galaxy survey data beyond linear order. It is worth noting that the coupling between nonlinear bias and nonlinear RSD, as well as higher-derivative, velocity-bias and stochastic-bias contributions, are not completely accounted for in the expressions given in these references. Recently, Ref. [180] (see also [810, 753]) presented complete expressions for the NLO galaxy power spectrum in redshift space (that is, the generalization of the results presented in Sec. 4.1.4 to redshift space) including the velocity bias contributions which become relevant at this order. In the context of the general bias expansion, the NLO redshift-space contributions to the galaxy power spectrum add three additional parameters, which are all related to velocity bias. Two parameters describe deterministic velocity bias (see Sec. 2.7), while the third describes the amplitude of random, small-scale galaxy velocities (see Sec. 2.8). It is worth emphasizing that displacement-type terms such as $u_\parallel \partial_\parallel \delta_g$, which become relevant beyond the leading-order galaxy two-point function, are protected from selection and radiative-transfer effects as well as velocity bias (on large scales) through the equivalence principle. Thus, even in the presence of all these observational complications, they in principle allow for robust constraints on the growth of structure from observed galaxy statistics.

While the rigorous perturbative approach (EFT of matter coupled with the general bias expansion) is an approach whose advantages have been explained in detail in Sec. 2 and Sec. 4, it is worth pointing out that alternative approaches to the modeling of small-scale RSD have been developed [811, 812, 813] in the context of the halo occupation distributions (HOD, see Sec. 9.1).

Finally, redshift-space distortions also complicate significantly the interpretation of nonlinear transformations of the galaxy density field, such as void catalogs [814, 815, 816, 817, 818] and "clipped" [819, 820] or log-transformed galaxy density fields [821, 822, 823, 824, 825, 826].

9.4 Galaxy statistics on the sky

After having described how the rest-frame galaxy density is transformed to observed coordinates $(\tilde{z}, \hat{\boldsymbol{n}})$, we are now in a position to describe the statistics of galaxies in these observed coordinates beyond the flat-sky limit adopted in the previous section. We now consider the fully general case, allowing for wide angles [see [827, 828, 829, 830, 831, 832] who use Eq. (9.35), and [833, 834] for treatments including the full expression Eq. (9.34)], and for deep surveys. This is of particular importance when one is interested in measuring long-wavelength fluctuations which extend across a significant redshift range and/or portion of the sky, for example to probe the signature of primordial non-Gaussianity. Here, we focus on the galaxy two-point function (see [530, 835, 771] for extensions to the three-point function allowing for wide angles),

and neglect subtleties involved in its optimal estimation (such as discussed in [836]). Moreover, we will not consider the effect of the survey window, which further complicates the treatment. Since we will always deal with observed, redshift-space coordinates in this section, we will drop the tilde, i.e. $\tilde{\boldsymbol{x}} \to \boldsymbol{x}$.

The standard approach of analyzing galaxy statistics on the sky is to decompose the observed galaxy overdensity in terms of spherical harmonics Y_{lm}, and radial window functions $W_l(p, \chi)$, labeled by a parameter p:

$$\tilde{\delta}_g(p; \hat{\boldsymbol{n}}) = \sum_{lm} \delta_{lm}(p) Y_{lm}(\hat{\boldsymbol{n}}), \qquad \text{where} \quad \delta_{lm}(p) = \int_0^\infty \chi^2 d\chi \int d^2\hat{\boldsymbol{n}} \, Y_{lm}^*(\hat{\boldsymbol{n}}) W_l(p, \chi) \delta_g(\boldsymbol{x}, \tau[\chi]) \,, \qquad (9.49)$$

and χ is the comoving radial coordinate, $W_l(p, \chi)$ is an l-dependent window function parametrized by a parameter p (specified below), and

$$\tilde{\delta}_g(\boldsymbol{x}, \tau) = \frac{n_g^{\text{obs}}(\boldsymbol{x}, \tau)}{\bar{n}_g^{\text{obs}}(\tau)} - 1 \qquad (9.50)$$

is the *observed* fractional galaxy overdensity. Throughout, we let $\chi = \bar{\chi}(\tilde{z})$ be a proxy for the observed redshift $\tilde{z}$, and similarly $\tau \equiv \bar{\tau}(\tilde{z})$ (see Sec. 9.3). Correspondingly, $\boldsymbol{x} \equiv \chi\hat{\boldsymbol{n}}$ parametrizes the observationally inferred (redshift-space) spatial location. Similarly, $\bar{n}_g^{\text{obs}}$ is defined on a constant-observed-redshift slice.

The parameter p of the window function $W_l(p, \chi)$ parametrizes the radial selection function. One option is to choose $W_l(p, \chi) = \chi^{-2}\delta_D(\chi - \bar{\chi}[p])$, which yields, at the two-point level, the angular two-point correlation function in narrow redshift bins $C_l(z, z')$ [123, 190, 837, 774]. Alternatively, choosing $W_l(p, \chi) = j_l(p\chi)$, we recover the spherical Fourier-Bessel approach developed by [838, 839, 840]. The spherical Fourier-Bessel basis has been applied to the analysis of the IRAS Redshift Survey [841], and to the reconstruction of the velocity and the gravitational potential fields [839]. More recently, Ref. [842] computed the spherical power spectrum of matter density fluctuations on the BAO scale, Ref. [843] cross-correlated their spherical Fourier-Bessel analysis of RSD with the CMB temperature anisotropies, while Ref. [833] computed the spherical power spectrum including all terms in Eq. (9.34). The advantages and disadvantages of the spherical Fourier-Bessel decomposition with respect to angular correlations in redshift bins $C_l(z_i, z_j)$ are reviewed in [844]. Ref. [845] applied this method to the latest data release of the BOSS CMASS sample from the Sloan Digital Sky Survey. Finally, Ref. [846] proposed the so-called logarithmic spherical waves $e^{i\omega \ln r}$ which can be used to express "Pseudo-Karhunen-Loève" (signal-to-noise) eigenmodes.

Let us then write the redshift-space galaxy density as

$$\tilde{\delta}_g(\boldsymbol{x}, \tau) = \sum_O b_O(\tau) O_s(\boldsymbol{x}, \tau) \,, \qquad (9.51)$$

where the O_s are renormalized redshift-space operators. For now, we neglect stochasticity, but return to this below. Since the effect of tensor perturbations on the galaxy density contrast is negligibly small [774], we can restrict to scalar perturbations here. Then, any linear redshift-space operator O_s can be derived from a real-space scalar O_r [cf. Eq. (9.35) and Eq. (9.46)] as,

$$O_s(\boldsymbol{x}, \tau) = \hat{\boldsymbol{n}}^{i_1} \cdots \hat{\boldsymbol{n}}^{i_n} \partial_{i_1} \cdots \partial_{i_n} O_r(\boldsymbol{x}, \tau) \,, \qquad (9.52)$$

where n can be any number, although the leading terms only involve $n = 1$ and $n = 2$. Equivalently, in Fourier space

$$O_s(\boldsymbol{k}, \tau) = (i\hat{\boldsymbol{n}} \cdot \boldsymbol{k})^n O_r(\boldsymbol{k}, \tau) \,. \qquad (9.53)$$

More generally, this applies to all leading-order contributions to redshift-space N-point functions. At loop level, $\hat{\boldsymbol{n}}$ can also be contracted with the loop momentum, which makes the derivations in the following more complicated. However, loop terms only become important on small scales, where the flat-sky limit applies and the treatment becomes much simpler. We will thus restrict to leading-order correlations, and consider in particular the two-point function. Then, the operators O_s are linear order in perturbation.

Using homogeneity and isotropy of the two-point function of the real-space scalar O_r, the galaxy two-

point function in redshift-space can be written as

$$
\begin{aligned}
\langle \tilde{\delta}_g(\boldsymbol{x}, \tau) \tilde{\delta}_g(\boldsymbol{x}', \tau') \rangle &= \sum_{O_s, O_s'} b_{O_s}(\tau) b_{O_s'}(\tau') \int_{\boldsymbol{k}} (i\boldsymbol{k} \cdot \hat{\boldsymbol{n}})^n (-i\boldsymbol{k} \cdot \hat{\boldsymbol{n}}')^{n'} P_{O_r O_r'}(k; \tau, \tau') \exp\left[i\boldsymbol{k} \cdot (\boldsymbol{x} - \boldsymbol{x}')\right] \\
&= \sum_{O_s, O_s'} b_{O_s}(\tau) b_{O_s'}(\tau') \int_{\boldsymbol{k}} \left(\frac{\partial}{\partial \chi}\right)^n e^{ik\chi(\hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{n}})} \left(\frac{\partial}{\partial \chi'}\right)^{n'} e^{-ik\chi'(\hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{n}}')} P_{O_r O_r'}(k; \tau, \tau') \\
&= \sum_{O_s, O_s'} b_{O_s}(\tau) b_{O_s'}(\tau') \frac{2}{\pi} \sum_{l,m} Y_{lm}(\hat{\boldsymbol{n}}) Y_{lm}^*(\hat{\boldsymbol{n}}') \\
&\quad \times \int k^2 dk\, k^{n+n'} \left(\frac{\partial}{\partial x}\right)^n j_l(x) \left(\frac{\partial}{\partial x'}\right)^{n'} j_{l'}(x') P_{O_r O_r'}(k; \tau, \tau'), \qquad (9.54)
\end{aligned}
$$

where $x = k\chi$, $x' = k\chi'$. Inserting this into Eq. (9.49), we obtain the two-point correlation of the $\{l, m, p\}$-decomposition of the galaxy overdensity as

$$
\langle \delta_{lm}(p) \delta_{l'm'}^*(p') \rangle = \delta_{ll'} \delta_{mm'} \sum_{O_s, O_s'} \frac{2}{\pi} \int k^2 dk\, F_{O_s}^l(k; p) F_{O_s'}^l(k; p') P_{O_r O_r'}(k), \qquad (9.55)
$$

where

$$
F_{O_s}^l(k; p) \equiv k^{n_O} \int \chi^2 d\chi\, W_l(p, \chi) \left[\left(\frac{\partial}{\partial x}\right)^{n_O} j_l(x)\right]_{x=k\chi} b_{O_s}(\tau[\chi]) D_{O_r}(\tau[\chi]) \quad \text{and}
$$

$$
P_{O_r O_r'}(k; \tau, \tau') = D_{O_r}(\tau) D_{O_r'}(\tau') P_{O_r O_r'}(k). \qquad (9.56)
$$

Let us consider the three operators that appear at linear order in $\tilde{\delta}_g$. In order to simplify notation, we relate them directly to the density field δ [i.e., $P_{O_r O_r'} \to P_{\mathrm{L}}$ in Eq. (9.55)]. We obtain:

$$
\text{(i)}\ O_s = \delta: \quad F_\delta^l(k; p) = \int \chi^2 d\chi\, W_l(p, \chi) j_l(k\chi) b_1(\tau[\chi]) D(\tau[\chi])
$$

$$
\text{(ii)}\ O_s = v_\parallel = \hat{\boldsymbol{n}}^i v_i: \quad F_{v_\parallel}^l(k; p) = -k^{-1} \int \chi^2 d\chi\, W_l(p, \chi) j_l'(k\chi) (\mathcal{H} f D)(\tau[\chi])
$$

$$
\text{(iii)}\ O_s = \partial_\parallel v_\parallel = \hat{\boldsymbol{n}}^j \hat{\boldsymbol{n}}^i \partial_j v_i: \quad F_{\partial_\parallel v_\parallel}^l(k; p) = - \int \chi^2 d\chi\, W_l(p, \chi) j_l''(k\chi) b_{\partial_\parallel v_\parallel}(\tau[\chi]) (\mathcal{H} f D)(\tau[\chi]). \qquad (9.57)
$$

The above relations can easily be generalized to include contributions that are integrated along the line of sight (e.g., [190, 774]). For terms (i) and (iii), we have allowed for bias parameters, where $b_{\partial_\parallel v_\parallel} \propto b_{K_\parallel}$ is only induced by the specific observational selection effects discussed in Sec. 9.2. On the other hand, the term (ii) is necessarily a pure projection effect, and hence unbiased, since the galaxy density cannot depend on the local matter velocity by way of the equivalence principle. Including the leading-order stochastic contribution is straightforward; it is given by

$$
\langle \delta_{lm}(p) \delta_{l'm'}^*(p') \rangle \Big|_{\text{stoch}} = \delta_{ll'} \delta_{mm'} \int \chi^2 d\chi\, W_l(p, \chi) W_l(p', \chi) P_\varepsilon^{\{0\}}(\tau[\chi]). \qquad (9.58)
$$

The Poisson shot-noise approximation corresponds to inserting $P_\varepsilon^{\{0\}}(\tau) = 1/\overline{n}_g(\tau)$.

Finally, one should note that actual surveys involve a window function describing the survey footprint on the sky. This complicates the expression Eq. (9.55), since the orthogonality of $\delta_{lm}(p)$ with respect to lm no longer holds [847].

Let us now consider a survey with limited footprint on the sky, such that the angle between different lines of sight within the survey footprint can be approximated as infinitesimal (i.e., wide-angle effects are negligible). Further, we assume that the extent of the survey in the line-of-sight direction is small; this can of

course be achieved by dividing the survey into redshift bins. Then, if we define $\delta_{lm}(p)$ using $W_l(p, \chi) = j_l(p\chi)$, we can construct the flat-sky version of the angular multipoles (App. C of [848]),

$$\delta^{\mathrm{fs}}(\boldsymbol{l}, p) \equiv \sqrt{\frac{4\pi}{2l+1}} \sum_{m=-l}^{l} i^{-m} \delta_{lm}(p) e^{im\varphi_{\boldsymbol{l}}} , \tag{9.59}$$

where (l, m) is replaced with $\boldsymbol{l} = (l_x, l_y)$, and $\varphi_{\boldsymbol{l}} = \arctan(l_y/l_x)$. The flat-sky multipoles $\delta^{\mathrm{fs}}(\boldsymbol{l}, p)$ in turn can be mapped onto the three-dimensional Euclidean Fourier transform $\delta_g(\boldsymbol{k})$ used extensively throughout this review, via

$$\delta_g(\boldsymbol{k}) \equiv \delta^{\mathrm{fs}}(\bar{\chi}\boldsymbol{k}_\perp, k_\parallel) \quad \text{where} \quad k_\parallel \equiv \hat{\boldsymbol{n}} \cdot \boldsymbol{k}; \ \boldsymbol{k}_\perp \equiv \boldsymbol{k} - k_\parallel \hat{\boldsymbol{n}} . \tag{9.60}$$

Here, $\bar{\chi}$ is the mean comoving distance of the redshift bin. Note that $\delta_g(\boldsymbol{k})$ has dimension $(\mathrm{length})^3$. In this flat-sky-shallow-survey limit, the two-point function $\langle \delta_g(\boldsymbol{k}) \delta_g(\boldsymbol{k}') \rangle$ then directly yields an estimate of the three-dimensional galaxy power spectrum in redshift space at an effective redshift $\bar{z}$, convolved with the three-dimensional window function of the survey. This is the approach commonly adopted in the analysis of current data sets such as 2dF, SDSS, 6dFGRS, Wigglez, and BOSS [25, 849, 850, 851, 852]. Future large-scale galaxy surveys such as SPHEREx [633], DESI [53], and Euclid [54], whose footprints exceed 10,000 square degrees, must include the wide-angle effect in order to correctly interpret the measured galaxy clustering on large scales. This is particularly important for detecting primordial non-Gaussianity (Sec. 7) and relativistic projection effects (Sec. 9.3).

10 Summary and outlook

Bias describes the relation between the observed number density of galaxies and the underlying matter density and spacetime perturbations, and is an indispensable ingredient in our model of the observed Universe. Key progress has been made on this problem in the past few years.

We now have a general framework for galaxy bias on large scales in the context of perturbation theory (Sec. 2). These bias parameters have well-defined physical interpretations, which become most clear by way of the generalized peak-background split argument (PBS, Sec. 3, Sec. 6.6.1, and Sec. 7.3). In fact, one can rigorously define the PBS argument to yield *exact* predictions for the bias parameters, provided one can accurately simulate the formation of the tracers of interest for different cosmologies (with different curvature in the case of the LIMD bias). This is certainly possible for dark matter halos. However, the perturbative bias expansion is much more general and applies, with caveats described below, to any tracer of the large-scale structure, including clusters of galaxies, voids, the Lyman-α forest, and 21cm line emission from neutral hydrogen.

Assuming that values are given for the relevant bias parameters, the perturbation-theory framework makes definite predictions for all observables related to galaxy clustering (Sec. 4): n-point functions (power spectrum, bispectrum), statistics of counts-in-cells, and cross-correlations with the matter and among different tracers. Cross-correlating different tracers can yield precise measurements of *relative* bias parameters which cancel cosmic variance at leading order. Crucially, at a given order in perturbation theory, one set of bias parameters describes all observables involving a given tracer. These predictions are, however, valid only on sufficiently large scales (see below), and in practice some or all of these bias parameters have to be determined from the data.

The excursion set (Sec. 5) and the peak approach (Sec. 6) have already furnished many insights into the scale dependence, nonlinearity and stochasticity of bias, as well as the validity of the peak-background split and its extension to variables other than the density. Furthermore, they provide quantitative predictions for the bias parameters, which can be tested against N-body simulations. However, the models considered so far rely on a number of assumptions, such as the spherical collapse approximation, which we expect to become increasingly less accurate towards lower halo masses $M \lesssim M_\star$.

A key application of bias is the incorporation of the effect of non-Gaussian initial conditions, which encode a rich array of signatures of early Universe physics, and lead to additional scale dependencies that in many cases cannot be mimicked by nonlinear gravitational evolution and baryonic effects (Sec. 7).

Even though the bulk of this review focuses on structure forming out of collisionless matter in a universe described by General Relativity, which is an excellent first-order approximation, the perturbative bias approach can be extended to tracers of a multi-component fluid made of CDM, baryons and massive neutrinos (Sec. 8), as is the case in the standard ΛCDM cosmology. Further, it can be extended to incorporate the effects of a dynamical dark energy, or modifications to General Relativity. The perturbative approach also allows for a consistent mapping from the local rest-frame galaxy density to the observed galaxy redshifts and positions on the sky, which we refer to as *projection effects* (Sec. 9). This includes relativistic effects which become relevant on the large scales targeted by forthcoming surveys. Moreover, a physical bias expansion of the galaxy velocity field (Sec. 2.7; Sec. 6.9.1) is a crucial ingredient for the prediction of projection effects beyond linear order.

While theoretically well-defined and rigorous, the inevitable downside of the perturbative bias expansion is that it breaks down on small scales. In addition to the perturbative description of the matter density itself, which breaks down at the nonlinear scale $R_{\rm NL}$ at which the density contrast is of order one, galaxy bias adds another scale, the nonlocality scale R_* (Sec. 2.5–2.6). By definition, the formation of galaxies in a given observed sample depends on the detailed distribution of matter within a region of this scale. Thus, we cannot hope to describe galaxy clustering perturbatively on scales that are of order R_* or smaller. Which of these two scales, R_* or $R_{\rm NL}$, is more limiting, depends on the galaxy sample at hand, as well as its redshift. While for halos we expect that R_* is of order the Lagrangian radius $R(M)$, little is known about this scale for galaxies. The halo occupation distribution approach (Sec. 9.1), in which the 1-halo term is calibrated with N-body simulations to describe the distribution of galaxies at small scales, could circumvent this limitation for the two-point function, but only if the transition region between 1- and 2-halo terms can

be described accurately and robustly.

Beyond the state of the field reviewed here, various important questions remain open, including:

- What is the scale R_* for galaxies? How much cosmological information can we extract from the galaxy power spectrum, bispectrum, and possibly higher n-point functions, in the context of the general perturbative bias expansion?

- Are there hierarchies and/or relations between the bias parameters of dark matter halos? For example, are the biases in Lagrangian space that involve tidal fields, and time derivatives thereof, systematically smaller than the Lagrangian LIMD bias parameters b_N^L? Further, to what level do we need to understand galaxy formation in order to make use of any such relations for actual galaxy samples? Does halo assembly bias play a significant role for galaxy bias?

- Lagrangian bias models such as the peaks and excursion-set formalisms can in principle provide a bias relation in Lagrangian space that is valid on all scales, and not restricted to $r \gg R_*$. Is there a way to derive a connection to Eulerian statistics for these models that does not break down on small scales?

- It would be desirable to extend the validity of these Lagrangian bias models to lower-mass halos, such as those targeted by current and upcoming galaxy redshift surveys. In particular, how can we incorporate deviations from spherical collapse and other effects due to the nonlinear evolution of small-scale density perturbations into these models?

Sophisticated models of galaxy bias will play a major role in the interpretation of upcoming large-scale structure surveys. We now finally have a rigorous framework of galaxy bias, which allows for the development of a robust theoretical description of galaxy clustering that matches the level of statistical and systematic uncertainties expected from these surveys. By advancing our understanding of these open issues, we will thus be able to makes the most of the data's potential to further our knowledge on galaxy formation, the history of the Universe, and fundamental physics.

Appendix A   Statistical field theory

In standard cosmological models, the initial conditions of the large-scale structure of the Universe is generated from quantum mechanical vacuum fluctuations. Due to its quantum mechanical origin, we cannot predict, in a deterministic sense, the precise initial conditions of the Universe we observe. Our theoretical treatment of the large-scale structure is, therefore, based on describing random fields. Throughout the review we draw on results from basic statistical field theory which we briefly review here. More detailed introductions to this topic can be found in [72, 482, 92].

Here and throughout the review, we assume that, on a spatial slice of fixed proper time, all the cosmological random fields are statistically homogeneous and isotropic, in accordance with the *cosmological principle*. The unperturbed FRW metric [Eq. (1.2) with $\Phi = 0 = \Psi$] is a manifestation of this, since spatial slices ($\tau = \text{const}$) are homogeneous and isotropic (maximally symmetric). Furthermore, we shall assume the *fair sample hypothesis*, which states that samples extracted from regions of the Universe that are sufficiently distant from each other are independent realizations of the same physical process. Observationally, the fair sample hypothesis has been found to be satisfied on scales above $\sim 80\, h^{-1}\, \text{Mpc}$ [853, 854]. Therefore, ensemble averages can be traded with spatial averages provided that the volume is large enough: this is the *ergodic hypothesis*.

A fundamental issue in the analysis of cosmic structures is to find the most appropriate observables to retrieve information on the distribution of fluctuations (the matter density field, for example), their initial conditions and subsequent evolution. Here, we shall concentrate on poly-spectra and correlation functions, which include moments and encompass all observables considered in this review.

Appendix A.1   Random fields in 3D Euclidean space

In accordance with the cosmological principle, we consider random fields defined on a given spatial slice through spacetime; in practice, one should think of this as defined through a fixed proper time. We choose Euclidean coordinates $\boldsymbol{x}$ for this slice, as written in Eq. (1.2). Further, we will focus on real scalar fields. This applies to most examples encountered in large-scale structure (temperature, density, pressure, gravitational potential and so on). We will briefly consider the generalization to vector and tensor fields at the end of this section.

A random scalar field in 3D Euclidean space is a set of random variables $\rho(\boldsymbol{x})$, together with a collection of distribution functions $p_n(\rho(\boldsymbol{x}_1) \ldots, \rho(\boldsymbol{x}_n))$, with $n \geq 1$. In order to provide a complete statistical description of the random field $\rho(\boldsymbol{x})$ it is in general necessary (but not sufficient [855]) to specify all of its correlation functions (i.e. moments).

The n-point correlation function is a specific expectation value given by

$$\langle \rho(\boldsymbol{x}_1) \ldots \rho(\boldsymbol{x}_n) \rangle = \int d\rho_1 \ldots d\rho_n \, P_n(\rho_1, \ldots, \rho_n) \rho_1 \ldots \rho_n \, , \tag{A.1}$$

where angle brackets denote ensemble averages and $\rho_i \equiv \rho(\boldsymbol{x}_i)$. In particular, the 1-point correlation is the mean $\langle \rho \rangle$ of the random field. Higher-order correlation functions follow a particular hierarchy, which we demonstrate in the case of the 2-point function

$$\langle \rho(\boldsymbol{x}_1)\rho(\boldsymbol{x}_2) \rangle \, . \tag{A.2}$$

In the event that we move one point (say $\boldsymbol{x}_1$) far away from the other, the value of the field at the points become independent, so that the expectation value tends towards

$$\langle \rho(\boldsymbol{x}_1)\rho(\boldsymbol{x}_2) \rangle \to \langle \rho(\boldsymbol{x}_1) \rangle \langle \rho(\boldsymbol{x}_2) \rangle = \langle \rho \rangle^2 \, , \tag{A.3}$$

where we have used the cosmological principle, which implies translation invariance and hence the same expectation value of $\rho(\boldsymbol{x})$ at any point on the slice. Therefore, we can write

$$\langle \rho(\boldsymbol{x}_1)\rho(\boldsymbol{x}_2) \rangle = \langle \rho \rangle^2 \left[1 + \xi^{(2)}(\boldsymbol{x}_1, \boldsymbol{x}_2) \right] \, . \tag{A.4}$$

This defines the *reduced* or *connected* 2-point correlation function $\xi_2(\boldsymbol{x}_1, \boldsymbol{x}_2)$. The *n*-point *connected correlation function* $\xi^{(n)}$ is recursively defined in such a way that $\langle \rho(\boldsymbol{x}_1) \ldots \rho(\boldsymbol{x}_n) \rangle$ is a sum of terms, where each term is associated with a partition of the set of n points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$. For the first four orders for instance, we have

$$\langle \rho(\boldsymbol{x}_1)\rho(\boldsymbol{x}_2) \rangle = \langle \rho \rangle^2 \left[1 + \xi^{(2)}(\boldsymbol{x}_1, \boldsymbol{x}_2) \right]$$

$$\langle \rho(\boldsymbol{x}_1)\rho(\boldsymbol{x}_2)\rho(\boldsymbol{x}_3) \rangle = \langle \rho \rangle^3 \left[1 + \xi^{(2)}(\boldsymbol{x}_1, \boldsymbol{x}_2) + 2 \text{ perm.} + \xi^{(3)}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3) \right]$$

$$\langle \rho(\boldsymbol{x}_1)\rho(\boldsymbol{x}_2)\rho(\boldsymbol{x}_3)\rho(\boldsymbol{x}_4) \rangle = \langle \rho \rangle^4 \Big[1 + \xi^{(2)}(\boldsymbol{x}_1, \boldsymbol{x}_2) + 5 \text{ perm.}$$
$$+ \xi^{(3)}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3) + 3 \text{ perm.} + \xi^{(2)}(\boldsymbol{x}_1, \boldsymbol{x}_2)\xi^{(2)}(\boldsymbol{x}_3, \boldsymbol{x}_4) + 2 \text{ perm.}$$
$$+ \xi^{(4)}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4) \Big] . \tag{A.5}$$

The "*n* perm." indicates that the term immediately preceeding is repeated with all possible cyclic permutations of the indices. In the case of a random field with zero mean, $\langle \delta(\boldsymbol{x}) \rangle = 0$ (which in this review denotes the matter density contrast), the hierarchy of correlation functions reads

$$\langle \delta(\boldsymbol{x}_1)\delta(\boldsymbol{x}_2) \rangle = \xi^{(2)}(\boldsymbol{x}_1, \boldsymbol{x}_2)$$

$$\langle \delta(\boldsymbol{x}_1)\delta(\boldsymbol{x}_2)\delta(\boldsymbol{x}_3) \rangle = \xi^{(3)}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3)$$

$$\langle \delta(\boldsymbol{x}_1)\delta(\boldsymbol{x}_2)\delta(\boldsymbol{x}_3)\delta(\boldsymbol{x}_4) \rangle = \xi^{(2)}(\boldsymbol{x}_1, \boldsymbol{x}_2)\xi^{(2)}(\boldsymbol{x}_3, \boldsymbol{x}_4) + 2 \text{ perm.} + \xi^{(4)}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4) . \tag{A.6}$$

We will also use the notation

$$\langle \delta(\boldsymbol{x}_1)\delta(\boldsymbol{x}_2) \cdots \delta(\boldsymbol{x}_n) \rangle_c \equiv \xi^{(n)}(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n) . \tag{A.7}$$

for the connected correlators.

The cosmological principle dictates that all expectation values and hence all correlation functions are invariant under global translations $\boldsymbol{x}_i \to \boldsymbol{x}_i + \Delta\boldsymbol{x}$ on a fixed time slice, which is also known as *statistical homogeneity*. For the reduced 2-point correlation, this implies

$$\xi^{(2)}(\boldsymbol{x}_1, \boldsymbol{x}_2) \equiv \xi^{(2)}(\boldsymbol{x}_1 - \boldsymbol{x}_2) , \tag{A.8}$$

so that it depends only on the separation between the two points. Similarly, in the absence of preferred directions, expectation values are invariant under a global rotation $\boldsymbol{R}$ of the coordinate system $\boldsymbol{x} \to \boldsymbol{R} \cdot \boldsymbol{x}$, which is known as *statistical isotropy*. This implies that the 2-point correlation function depends only on the magnitude $|\boldsymbol{x}_1 - \boldsymbol{x}_2|$ of the separation vector,

$$\xi^{(2)}(\boldsymbol{x}_1, \boldsymbol{x}_2) \equiv \xi^{(2)}(|\boldsymbol{x}_1 - \boldsymbol{x}_2|) . \tag{A.9}$$

This holds also if one cross-correlates different fields, and for higher *n*-point correlation functions, which can only depend on $x_{ij} \equiv |\boldsymbol{x}_i - \boldsymbol{x}_j|$. Note that projection effects such as redshift-space distortions induce a dependence on $(\boldsymbol{x}_i - \boldsymbol{x}_j) \cdot \hat{\boldsymbol{n}}$ in the observed *n*-point functions, where $\hat{\boldsymbol{n}}$ is the line of sight (Sec. 9.3). While the cosmological principle requires statistical homogeneity, the ergodic hypothesis is valid only if the *n*-point correlation functions decay sufficiently rapidly to zero in the limit of large separations. This is indeed the case in the standard ΛCDM cosmology.

Note that these definitions can be extended to point processes, i.e. for distributions rather than continuous fields (in which case the correlators $\langle \rho(\boldsymbol{x}_1) \ldots \rho(\boldsymbol{x}_n) \rangle$ are also called joint intensities), such as the peaks discussed in Sec. 6, but there are some subtleties (such as shot noise corrections) owing to discreteness.

Finally, we turn to vector and tensor fields. Homogeneity and isotropy imply that two-point scalar-vector cross- and vector auto-correlation functions can be written as

$$\langle \delta(\boldsymbol{0})v^i(\boldsymbol{r}) \rangle = \hat{r}^i \xi_{\delta v}(r)$$

$$\langle v^i(\boldsymbol{0})v^j(\boldsymbol{r}) \rangle = \delta^{ij}\xi_{v,1}(r) + \left(\hat{r}^i\hat{r}^j - \frac{1}{3}\delta^{ij} \right) \xi_{v,2}(r) , \tag{A.10}$$

where $\xi_{\delta v}(r)$, $\xi_{v,1}(r)$, and $\xi_{v,2}(r)$ are scalar functions. Note that the scalar-vector correlation $\xi_{\delta v}(r)$ is generated only for longitudinal vector fields, while the vector two-point auto-correlation functions $\xi_{v,1}(r)$ and $\xi_{v,2}(r)$ receive contributions from both longitudinal and transverse vectors. $\xi_{v,1}(r)$ and $\xi_{v,2}(r)$ are regular at $r = 0$. The auto- and cross-correlations of symmetric tensor fields K_{ij} can be similarly constructed out of δ_{kl} and $(\hat{r}_m \hat{r}_n - \delta_{mn}/3)$. While these relations in general become lengthy, the auto-correlation function of a trace-free tensor field at zero lag simplifies to

$$\langle K_{ij}(\boldsymbol{x}) K_{lm}(\boldsymbol{x}) \rangle = \frac{1}{15}\sigma^2 \left(\delta_{il}\delta_{jm} + \delta_{im}\delta_{jl} - \frac{2}{3}\delta_{ij}\delta_{lm} \right), \quad \text{where} \quad \sigma^2 = \frac{3}{2}\langle K_{ij} K^{ij} \rangle. \tag{A.11}$$

Note that, if we identify K_{ij} with the tidal field [Eq. (2.21) on p. 24], then σ^2 corresponds to the variance of the density field.

Appendix A.2   Fourier representation

In a statistically homogeneous Universe, it is convenient to represent random fields by their Fourier components. In the following, we assume a field δ with zero mean. Adopting the Fourier convention

$$\delta(\boldsymbol{k}) = \int d^3\boldsymbol{x}\, \delta(\boldsymbol{x})\, e^{-i\boldsymbol{k}\cdot\boldsymbol{x}}, \qquad \delta(\boldsymbol{x}) = \int \frac{d^3\boldsymbol{k}}{(2\pi)^3}\delta(\boldsymbol{k})\, e^{i\boldsymbol{k}\cdot\boldsymbol{x}} \equiv \int_{\boldsymbol{k}}\delta(\boldsymbol{k})\, e^{i\boldsymbol{k}\cdot\boldsymbol{x}}, \tag{A.12}$$

the power spectrum $P(k)$ of the field is the expectation value

$$\langle \delta(\boldsymbol{k})\delta^*(\boldsymbol{k}') \rangle \equiv \langle \delta(\boldsymbol{k})\delta^*(\boldsymbol{k}') \rangle' \, (2\pi)^3 \delta_D(\boldsymbol{k} - \boldsymbol{k}') = P(\boldsymbol{k})\, (2\pi)^3 \delta_D(\boldsymbol{k} - \boldsymbol{k}'). \tag{A.13}$$

Here, the superscript $*$ stands for the complex conjugate, which makes $P(k)$ positive definite. The Dirac delta is a consequence of translational invariance (homogeneity). Otherwise, the ensemble average $\langle \delta(\boldsymbol{k})\delta(\boldsymbol{k}') \rangle$ would acquire a phase factor when $\boldsymbol{x} \to \boldsymbol{x} + \Delta\boldsymbol{x}$. If the field $\delta(\boldsymbol{x})$ is real, $\delta^*(\boldsymbol{k}) = \delta(-\boldsymbol{k})$, and we obtain

$$\langle \delta(\boldsymbol{k})\delta(-\boldsymbol{k}') \rangle = P(\boldsymbol{k})(2\pi)^3 \delta_D(\boldsymbol{k} + \boldsymbol{k}'). \tag{A.14}$$

Requesting further rotational invariance implies that the power spectrum depends only on $k = |\boldsymbol{k}|$, i.e.

$$\langle \delta(\boldsymbol{k})\delta^*(\boldsymbol{k}') \rangle = P(k)(2\pi)^3 \delta_D(\boldsymbol{k} - \boldsymbol{k}'). \tag{A.15}$$

Note that, with our Fourier convention, the Dirac delta distribution is given by

$$\delta_D(\boldsymbol{k}) = \frac{1}{(2\pi)^3} \int d^3\boldsymbol{r}\, e^{-i\boldsymbol{k}\cdot\boldsymbol{r}}. \tag{A.16}$$

The connected 2-point correlation function is the Fourier transform of the power spectrum. For a three-dimensional homogeneous and isotropic random field with zero mean, we have

$$\xi^{(2)}(r) = \int_{\boldsymbol{k}} P(\boldsymbol{k})\, e^{i\boldsymbol{k}\cdot\boldsymbol{r}} = \frac{1}{2\pi^2} \int_0^\infty dk\, k^2 P(k) j_0(kr). \tag{A.17}$$

Here, $j_0(x) = \sin(x)/x$ is a spherical Bessel function. For sake of completeness, note that, in one and two dimensions, we have

$$\xi(r) = \frac{1}{\pi} \int_0^\infty dk\, P(k) \cos(kr) \qquad \text{(1D)} \tag{A.18}$$

$$\xi(r) = \frac{1}{2\pi} \int_0^\infty dk\, kP(k) J_0(kr) \qquad \text{(2D)} \tag{A.19}$$

respectively, where $J_0(x)$ is a regular Bessel function. The fact that $P(k)$ is positive definite however does not imply that $\xi^{(2)}(r)$ is also positive definite. Indeed, we have

$$\int d^3\boldsymbol{r}\, \xi^{(2)}(r) = \lim_{k \to 0} P(k), \tag{A.20}$$

which should vanish following the ergodic hypothesis for the zero-mean field δ. This implies that $\xi^{(2)}(r)$ has to be negative for some values of r. The variance of an isotropic random field is given by

$$\sigma^2 = \xi(0) = \frac{1}{2\pi^2} \int_0^\infty dk\, k^2 P(k) = \int_0^\infty \frac{dk}{k} \Delta^2(k) \,, \qquad (A.21)$$

where we have introduced the dimensionless power spectrum $\Delta^2(k)$ which quantifies the variance of the density field per unit $\ln k$ and, unlike $P(k)$, is independent of the Fourier convention used. In our convention, $\Delta^2(k) \equiv k^3 P(k)/(2\pi^2)$.

These results can be extended to higher-order correlation functions. For instance, the bispectrum $B(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3)$ is the expectation value

$$\langle \delta(\boldsymbol{k}_1)\delta(\boldsymbol{k}_2)\delta(\boldsymbol{k}_3) \rangle = B(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3)(2\pi)^3 \delta_D(\boldsymbol{k}_1 + \boldsymbol{k}_2 + \boldsymbol{k}_3) \,. \qquad (A.22)$$

The Dirac delta ensures that the wavevectors $\boldsymbol{k}_i$ correspond to the three sides of a triangle, so that the three-point function is invariant under translations. Isotropy further implies that B be a function of the wavenumbers $k_i = |\boldsymbol{k}_i|$ solely. In analogy with the power spectrum, the 3-point connected correlation function is the Fourier transform of the bispectrum,

$$\xi^{(3)}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3) = (2\pi)^3 \int_{\boldsymbol{k}_1} \int_{\boldsymbol{k}_2} \int_{\boldsymbol{k}_3} \delta_D(\boldsymbol{k}_1 + \boldsymbol{k}_2 + \boldsymbol{k}_3) B(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3) e^{i[\boldsymbol{k}_1 \cdot \boldsymbol{x}_1 + \boldsymbol{k}_2 \cdot \boldsymbol{x}_2 + \boldsymbol{k}_3 \cdot \boldsymbol{x}_3]} \,. \qquad (A.23)$$

The trispectrum and higher-order poly-spectra are defined analogously.

Finally, the Fourier-space two-point correlation functions of vector and tensor fields, such as

$$\langle \delta(\boldsymbol{k}) v^i(\boldsymbol{k}') \rangle', \ \langle v^i(\boldsymbol{k}) v^j(\boldsymbol{k}') \rangle', \ \langle \delta(\boldsymbol{k}) K_{ij}(\boldsymbol{k}') \rangle', \ \langle K_{ij}(\boldsymbol{k}) K_{lm}(\boldsymbol{k}') \rangle', \qquad (A.24)$$

can be decomposed into products of the vector $\hat{k}^i$, the tensors δ_{ij} and $(\hat{k}_m \hat{k}_n - \delta_{mn}/3)$ [cf. Eq. (2.22)], and the totally trace-free projection tensor

$$\mathcal{P}_{ijlm}(\hat{\boldsymbol{k}}) = \hat{k}_i \hat{k}_j \hat{k}_l \hat{k}_m - \frac{1}{7}\left(\delta_{ij}\hat{k}_l \hat{k}_m + 5 \text{ perm.}\right) + \frac{1}{35}\left(\delta_{ij}\delta_{lm} + 2 \text{ perm.}\right) \,. \qquad (A.25)$$

We often deal with a smoothed version of the density field δ_R, obtained by convolving with a spherically symmetric filtering kernel W_R, where R denotes the filter scale. Throughout, the kernels are normalized through $\int d^3\boldsymbol{x}\, W_R(|\boldsymbol{x}|) = 1$, which implies $\lim_{k \to 0} W_R(k) = 1$. The three most popular filtering kernels $W_R(x)$ and their Fourier transform $W_R(k)$ are, respectively,

- sharp-k filter:

$$W_R(x) = \frac{3}{4\pi R^3}\left[3\frac{j_1(x/R)}{x/R}\right], \qquad\qquad W_R(k) = \Theta_H(1 - kR) \qquad (A.26)$$

- Gaussian filter:

$$W_R(x) = \frac{1}{[2\pi R^2]^{3/2}} e^{-\frac{1}{2}x^2/R^2}, \qquad\qquad W_R(k) = e^{-\frac{1}{2}R^2 k^2} \qquad (A.27)$$

- Tophat filter:

$$W_R(x) = \frac{3}{4\pi R^3}\Theta_H\left(1 - \frac{x}{R}\right), \qquad\qquad W_R(k) = 3\frac{j_1(kR)}{kR} \,. \qquad (A.28)$$

Here, $\Theta(x)$ is the Heaviside step function (Tab. 2) and

$$j_1(x) = \frac{\sin x - x \cos x}{x^2} \tag{A.29}$$

is a spherical Bessel function. Bearing in mind that convolution in real space is a multiplication in Fourier space (and vice versa), the variance of the random field on scale R thus is

$$\sigma^2(R) = \langle \delta_R^2 \rangle = \int_0^\infty \frac{dk}{k} \, \Delta^2(k) \, |W_R(k)|^2 \,. \tag{A.30}$$

Appendix A.3   Gaussian random fields

Gaussian random fields are essential in cosmology for mainly two reasons. Firstly, the inflationary paradigm predicts that the primordial fluctuations which gave rise to the large-scale structure of the Universe closely followed Gaussian statistics; moreover, this is confirmed by CMB observations, which find that the CMB temperature is Gaussian to roughly one part in a thousand [580]. Secondly, the central limit theorem, which states that the superposition of a large number of (uncorrelated) random processes asymptotes to a Gaussian distribution, can often be applied in large-scale structure; for example, in the case of projected density fields [856].

The definition of a Gaussian random field, with zero mean, is that its distribution functions are given by Gaussian multi-variates,

$$P_n(\boldsymbol{y}) = \frac{\sqrt{\det \mathrm{C}^{-1}}}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2} \boldsymbol{y}^\top \mathrm{C}^{-1} \boldsymbol{y} \right) \,, \tag{A.31}$$

where $\boldsymbol{y}$ is the random vector $(\delta(\boldsymbol{x}_1), \ldots, \delta(\boldsymbol{x}_n))$ and C^{-1} denotes the inverse of the covariance matrix, with entries given by

$$\mathrm{C}_{ij} = \langle \delta(\boldsymbol{x}_i) \delta(\boldsymbol{x}_j) \rangle \,. \tag{A.32}$$

Note that the covariance matrix is symmetric and non-negative definite (which also implies $|\mathrm{C}| \geq 0$).

A key property of Gaussian random fields is that, in the Fourier representation, the phases of the Fourier modes follow independent white-noise distributions. Equivalently, the real and imaginary part of the coefficients follow independent Gaussian distributions. Let us assume that the fluctuation field $\delta(\boldsymbol{x})$ is periodic over a volume $V = L^3$, so that the integral $\int_{\boldsymbol{k}}$ in Eq. (A.12) is replaced by $(1/V)\sum_{\boldsymbol{k}}$, where the sum runs over integer multiples $(n_x, n_y, n_z)k_F$ and $k_F = 2\pi/L$ is the fundamental wavenumber. Then, $\delta(\boldsymbol{k})$ is dimensionless. Furthermore, let us denote the real and imaginary parts of $\delta(\boldsymbol{k})$ by $\mathrm{Re}\,\delta_{\boldsymbol{k}}$ and $\mathrm{Im}\,\delta_{\boldsymbol{k}}$. Their joint probability distribution is the bivariate Gaussian

$$p\big(\mathrm{Re}\,\delta_{\boldsymbol{k}}, \mathrm{Im}\,\delta_{\boldsymbol{k}}\big) d\,\mathrm{Re}\,\delta_{\boldsymbol{k}} d\,\mathrm{Im}\,\delta_{\boldsymbol{k}} = \frac{1}{\pi\sigma_k^2} \exp\left[-\frac{(\mathrm{Re}\,\delta_{\boldsymbol{k}})^2 + (\mathrm{Im}\,\delta_{\boldsymbol{k}})^2}{\sigma_k^2} \right] d\,\mathrm{Re}\,\delta_{\boldsymbol{k}} d\,\mathrm{Im}\,\delta_{\boldsymbol{k}} \,, \tag{A.33}$$

where $\sigma_k^2 = (2\pi)^3 P(k)/k_F^3$ is the variance of $\mathrm{Re}\,\delta_{\boldsymbol{k}}$ and $\mathrm{Im}\,\delta_{\boldsymbol{k}}$ (which depends only on k due to isotropy). In terms of the amplitude $|\delta_{\boldsymbol{k}}|$ and phase $\phi_{\boldsymbol{k}}$, we have

$$p(|\delta_{\boldsymbol{k}}|, \phi_{\boldsymbol{k}}) d|\delta_{\boldsymbol{k}}| d\phi_{\boldsymbol{k}} = \frac{1}{\pi\sigma_k^2} \exp\left[-\frac{|\delta_{\boldsymbol{k}}|^2}{\sigma_k^2} \right] |\delta_{\boldsymbol{k}}| d|\delta_{\boldsymbol{k}}| d\phi_{\boldsymbol{k}} \,, \tag{A.34}$$

i.e. $|\delta_{\boldsymbol{k}}|$ follows a Rayleigh distribution. Note that for real scalar fields $\delta(\boldsymbol{x})$, we have the additional constraints $\mathrm{Re}\,\delta_{-\boldsymbol{k}} = \mathrm{Re}\,\delta_{\boldsymbol{k}}$, $\mathrm{Im}\,\delta_{-\boldsymbol{k}} = -\mathrm{Im}\,\delta_{\boldsymbol{k}}$, in which case $\delta_{\boldsymbol{k}}$ must be generated over half of the plane (e.g. $k_z \geq 0$) only.

For a homogeneous Gaussian random field, all the statistical information is contained in the connected 2-point correlation $\xi^{(2)}(x)$ or, equivalently, in the power spectrum $P(k)$. All the higher-order connected correlation functions are identically zero, so that their measurement from observations strictly tests Gaussianity. The n-point correlation functions $\langle \delta(\boldsymbol{x}_1) \ldots \delta(\boldsymbol{x}_n) \rangle$ then simplify to a sum of products of irreducible 2-point functions $\xi^{(2)}$. This property is known as *Wick's theorem*. For instance, the 4-point correlator in

Eq. (A.6) simplifies to

$$\langle \delta(\boldsymbol{x}_1)\delta(\boldsymbol{x}_2)\delta(\boldsymbol{x}_3)\delta(\boldsymbol{x}_4)\rangle = \xi^{(2)}(\boldsymbol{x}_1,\boldsymbol{x}_2)\xi^{(2)}(\boldsymbol{x}_3,\boldsymbol{x}_4) + 2 \text{ perm. .} \tag{A.35}$$

While the *initial conditions* are currently found to be consistent with perfect Gaussianity, in particular using measurements of the CMB, the observed, late-time distribution of galaxies is highly non-Gaussian due to gravitational instability and galaxy bias. Therefore, unlike the primary CMB anisotropies whose distribution is almost perfectly Gaussian and hence completely described by the power spectrum, higher-order correlation functions are necessary to specify the galaxy distribution. Note that the density contrast (or CMB temperature, for that matter) cannot be *exactly* Gaussian because the field has to satisfy the physical constraint $\delta > -1$. Nevertheless, Gaussianity can be a good approximation when the amplitude of fluctuations is small, i.e. when the density fluctuations are still linear.

The fact that the likelihood of a Gaussian statistical field is so simple allows for several important analytical results, including the number density of peaks in the initial density field (Sec. 6.2). Exact renormalized bias operators can be derived for Gaussian fields, as described in Sec. 6.6 and [100]. For this, we define orthogonal polynomials [857] with respect to a given weight $p(\boldsymbol{w})$, for example the joint PDF of the 5 peak invariants. Orthogonal polynomials satisfy

$$\int d^n \boldsymbol{w}\, p(\boldsymbol{w}) O_{\boldsymbol{n}}(\boldsymbol{w}) O_{\boldsymbol{n}'}^{\star}(\boldsymbol{w}) = \delta_{\boldsymbol{n}\boldsymbol{n}'}\,, \tag{A.36}$$

where $O_{\boldsymbol{n}}^{\star}$ is the dual polynomial to $O_{\boldsymbol{n}}$, $\boldsymbol{n}, \boldsymbol{n}'$ are sets of indices, $\delta_{\boldsymbol{n}\boldsymbol{n}'}$ denotes the Kronecker symbol in the space of these sets, and $p(\boldsymbol{w})$ is the joint 1-point distribution of the 5 peak invariants. Univariate orthogonal polynomials, which depend on only one element of the vector $\boldsymbol{w}$, are always their own dual.

We now give the first few dual polynomials for the bivariate Hermite polynomials which are employed in Sec. 6.6.1 [see 196, for details]:

$$H_{n0}^{\star}(\nu, J_1) = H_n(\nu)\,, \qquad H_{0n}^{\star}(\nu, J_1) = H_n(J_1)\,,$$
$$H_{11}^{\star}(\nu, J_1) = \nu J_1 - \gamma_1\,,$$
$$H_{21}^{\star}(\nu, J_1) = \nu^2 J_1 - J_1 - 2\gamma_1 \nu\,,$$
$$H_{12}^{\star}(\nu, J_1) = \nu J_1^2 - \nu - 2\gamma_1 J_1\,,$$
$$H_{31}^{\star}(\nu, J_1) = \nu^3 J_1 - 3\nu J_1 - 3\gamma_1 \nu^2 + 3\gamma_1\,,$$
$$H_{22}^{\star}(\nu, J_1) = \nu^2 J_1^2 - \nu^2 - J_1^2 - 4\gamma_1 \nu J_1 + 1 + 2\gamma_1^2\,. \tag{A.37}$$

Note that one could work with the variable $z = (J_1 - \gamma_1 \nu)/\sqrt{1 - \gamma_1^2}$ rather than J_1 as in [492] and, thus, end up with univariate Hermite polynomials.

Appendix B   Cosmological perturbation theory

In this appendix, we review the PT approach to the description of the nonlinear gravitational evolution of dark matter density and velocity fields. On large scales, where the perturbative bias expansion is valid, the vorticity as well as baryonic pressure effects are small, and the nonlinear evolution of the cosmic matter density field can be well approximated by a self-gravitating, presureless, longitudinal-velocity single fluid. We will turn to the corrections induced by neglecting the aforementioned effects in Appendix B.3.

Appendix B.1   Standard perturbation theory

In the pressureless-fluid approximation, the time evolution of the density $\delta(\boldsymbol{x}, \tau)$ and velocity fields $\boldsymbol{v}(\boldsymbol{x}, \tau)$ are governed by the continuity, Euler and Poisson equations:

$$\frac{\partial}{\partial \tau} \delta(\boldsymbol{x}, \tau) + \boldsymbol{\nabla} \cdot \{[1 + \delta(\boldsymbol{x}, \tau)] \, \boldsymbol{v}(\boldsymbol{x}, \tau)\} = 0 \tag{B.1}$$

$$\frac{\partial}{\partial \tau} \boldsymbol{v}(\boldsymbol{x}, \tau) + [\boldsymbol{v}(\boldsymbol{x}, \tau) \cdot \boldsymbol{\nabla}] \, \boldsymbol{v}(\boldsymbol{x}, \tau) + \mathcal{H}(\tau) \boldsymbol{v}(\boldsymbol{x}, \tau) = -\boldsymbol{\nabla} \Phi(\boldsymbol{x}, \tau) \tag{B.2}$$

$$\nabla^2 \Phi(\boldsymbol{x}, \tau) = \frac{3}{2} \mathcal{H}^2 \Omega_m(\tau) \delta(\boldsymbol{x}, \tau) \,. \tag{B.3}$$

These equations are strictly valid for ΛCDM only; however, they are also highly accurate for canonical, non-clustering dark energy models where the sound horizon of the dark energy component is of order $\mathcal{H}^{-1}$ (see Sec. 8.3 for a brief discussion).

In Fourier space, we can eliminate the gravitational potential Φ and express the longitudinal velocity field in terms of the velocity divergence $\theta \equiv \boldsymbol{\nabla} \cdot \boldsymbol{v}(\boldsymbol{x}, \tau)$ to obtain the following two equations:

$$\frac{\partial \delta(\boldsymbol{k}, \tau)}{\partial \tau} + \theta(\boldsymbol{k}, \tau) = -\int_{\boldsymbol{k}_1} \int_{\boldsymbol{k}_2} (2\pi)^3 \delta_D(\boldsymbol{k} - \boldsymbol{k}_{12}) \alpha(\boldsymbol{k}_1, \boldsymbol{k}_2) \theta(\boldsymbol{k}_1, \tau) \delta(\boldsymbol{k}_2, \tau), \tag{B.4}$$

$$\frac{\partial \theta(\boldsymbol{k}, \tau)}{\partial \tau} + \mathcal{H}(\tau) \theta(\boldsymbol{k}, \tau) + \frac{3}{2} \mathcal{H}^2(\tau) \Omega_m(\tau) \delta(\boldsymbol{k}, \tau) = -\int_{\boldsymbol{k}_1} \int_{\boldsymbol{k}_2} (2\pi)^3 \delta_D(\boldsymbol{k} - \boldsymbol{k}_{12}) \beta(\boldsymbol{k}_1, \boldsymbol{k}_2) \theta(\boldsymbol{k}_1, \tau) \theta(\boldsymbol{k}_2, \tau). \tag{B.5}$$

where $\boldsymbol{k}_{ij\ldots} \equiv \boldsymbol{k}_i + \boldsymbol{k}_j + \cdots$, and

$$\alpha(\boldsymbol{k}_1, \boldsymbol{k}_2) = \frac{\boldsymbol{k}_{12} \cdot \boldsymbol{k}_1}{k_1^2}, \quad \beta(\boldsymbol{k}_1, \boldsymbol{k}_2) = \frac{k_{12}^2 (\boldsymbol{k}_1 \cdot \boldsymbol{k}_2)}{2 k_1^2 k_2^2}. \tag{B.6}$$

In SPT, Eqs. (B.4)–(B.5) are solved perturbatively in terms of the linear density contrast $\delta^{(1)}(\boldsymbol{k}, \tau)$:

$$\delta(\boldsymbol{k}, \tau) = \sum_{n=1}^{\infty} \delta^{(n)}(\boldsymbol{k}, \tau), \quad \theta(\boldsymbol{k}, \tau) = \sum_{n=1}^{\infty} \theta^{(n)}(\boldsymbol{k}, \tau), \tag{B.7}$$

where the superscript (n) indicates that a term involves n powers of the linear density contrast; we refer to such a term as being *n-th order in perturbation theory*.

On very large scales, $k \to 0$, $\delta^{(1)}$ is much less than 1, and the quadratic source term on the right-hand side of Eqs. (B.4)–(B.5) can be neglected. Then, all Fourier modes evolve independently, and we obtain a single second-order ordinary differential equation for the evolution of the linear density contrast, $\delta^{(1)}(\boldsymbol{k}, \tau)$:

$$\frac{\partial^2}{\partial \tau^2} \delta^{(1)}(\boldsymbol{k}, \tau) + \mathcal{H}(\tau) \frac{\partial}{\partial \tau} \delta^{(1)}(\boldsymbol{k}, \tau) - \frac{3}{2} \Omega_m(\tau) \mathcal{H}^2(\tau) \delta^{(1)}(\boldsymbol{k}, \tau) = 0 \,. \tag{B.8}$$

We see that the time evolution equation is scale-independent; that is, all Fourier modes evolve at the same rate, independent of $\boldsymbol{k}$. In fact, this is the reason why time derivatives of the density field can be reordered to be successively higher order in perturbation theory in Sec. 2.5.2–2.5.3. We can then factor out the *linear*

growth factor $D(\tau)$ by writing $\delta^{(1)}(\boldsymbol{k}, \tau) = [D(\tau)/D(\tau_0)]\delta^{(1)}(\boldsymbol{k}, \tau_0)$, where τ_0 is a reference time, and $D(\tau)$ obeys

$$\frac{d^2}{d\tau^2}D(\tau) + \mathcal{H}\frac{d}{d\tau}D(\tau) - \frac{3}{2}\Omega_m(\tau)\mathcal{H}^2 D(\tau) = 0\,, \tag{B.9}$$

where initial conditions are setup so that $D(\tau)$ follows the growing mode, which corresponds to $D(\tau) \propto a(\tau)$ during matter domination. Unless otherwise indicated, all relations in this review are independent of the overall normalization of the growth factor. For reference, two popular choices for the latter are *i)* $D(\tau) = a(\tau)$ during matter domination (for $a \ll 1$), which is denoted as $D_{\rm md}(\tau)$ in the text; and *ii)* $D(\tau_0) = 1$, where τ_0 corresponds to today's epoch.

The linear-theory velocity divergence field is then obtained from the linearized continuity equation [Eq. (B.4)],

$$\theta^{(1)}(\boldsymbol{k}, \tau) = -\frac{d\ln D(\tau)}{d\tau}\delta^{(1)}(\boldsymbol{k}, \tau) = -f\mathcal{H}(\tau)\delta^{(1)}(\boldsymbol{k}, \tau)\,, \tag{B.10}$$

where the logarithmic growth rate $f(\tau) \equiv d\ln D/d\ln a$. Note that the identity

$$\frac{d(\mathcal{H}f)}{d\tau} = \mathcal{H}^2\left(\frac{3}{2}\Omega_m - f - f^2\right) \tag{B.11}$$

follows from the linear equations of motion.

The equations of motion Eqs. (B.4)–(B.5) motivate an ansatz of writing the n-th order solution in Eq. (B.7) as

$$\delta^{(n)}(\boldsymbol{k}, \tau) = \int_{\boldsymbol{k}_1}\cdots\int_{\boldsymbol{k}_n}(2\pi)^3\delta_D(\boldsymbol{k} - \boldsymbol{k}_{12\cdots n})F_n(\boldsymbol{k}_1, \cdots, \boldsymbol{k}_n, \tau)\delta^{(1)}(\boldsymbol{k}_1, \tau)\cdots\delta^{(1)}(\boldsymbol{k}_n, \tau)$$

$$\theta^{(n)}(\boldsymbol{k}, \tau) = -\mathcal{H}(\tau)f(\tau)\int_{\boldsymbol{k}_1}\cdots\int_{\boldsymbol{k}_n}(2\pi)^3\delta_D(\boldsymbol{k} - \boldsymbol{k}_{12\cdots n})G_n(\boldsymbol{k}_1, \cdots, \boldsymbol{k}_n, \tau)\delta^{(1)}(\boldsymbol{k}_1, \tau)\cdots\delta^{(1)}(\boldsymbol{k}_n, \tau), \tag{B.12}$$

with symmetrized density and velocity divergence kernels, respectively, F_n and G_n. From the linear solutions, $F_1 = G_1 = 1$ is obvious. Inserting the ansatz Eq. (B.12) into Eqs. (B.4)–(B.5) yields evolution equations for the kernels (see, e.g. [250]). In general, the kernels F_n and G_n depend on time, except when Ω_m/f^2 is constant in time. This is indeed the case for the EdS Universe where $\Omega_m = f = 1$ and the kernels become particularly simple. The usual practice in SPT is to calculate the time-independent kernels in the EdS Universe and apply the same kernels to other cosmologies. In this case, the linear growth function encodes the entire cosmology dependence of the nonlinear solutions in SPT. For ΛCDM and quintessence cosmologies, this approach provides a very accurate description of the full solution [858]. Specifically, the second-order kernels F_2 and G_2 in the EdS cosmology are given by

$$F_2(\boldsymbol{k}_1, \boldsymbol{k}_2) = \frac{5}{7} + \frac{2}{7}\frac{(\boldsymbol{k}_1 \cdot \boldsymbol{k}_2)^2}{k_1^2 k_2^2} + \frac{\boldsymbol{k}_1 \cdot \boldsymbol{k}_2}{2k_1 k_2}\left(\frac{k_1}{k_2} + \frac{k_2}{k_1}\right),$$

$$G_2(\boldsymbol{k}_1, \boldsymbol{k}_2) = \frac{3}{7} + \frac{4}{7}\frac{(\boldsymbol{k}_1 \cdot \boldsymbol{k}_2)^2}{k_1^2 k_2^2} + \frac{\boldsymbol{k}_1 \cdot \boldsymbol{k}_2}{2k_1 k_2}\left(\frac{k_1}{k_2} + \frac{k_2}{k_1}\right). \tag{B.13}$$

The second-order density and velocity field can also be written in real space as

$$\delta^{(2)}(\boldsymbol{x}, \tau) = \frac{17}{21}[\delta^{(1)}]^2 + \frac{2}{7}[K_{ij}^{(1)}]^2 - s_{(1)}^i\partial_i\delta^{(1)}\,, \tag{B.14}$$

$$-\frac{1}{\mathcal{H}(\tau)f(\tau)}\theta^{(2)}(\boldsymbol{x}, \tau) = \frac{13}{21}[\delta^{(1)}]^2 + \frac{4}{7}[K_{ij}^{(1)}]^2 - s_{(1)}^i\partial_i\delta^{(1)}\,, \tag{B.15}$$

where on the right-hand side of Eq. (B.14) all terms are evaluated at $\boldsymbol{x}, \tau$. Here,

$$\boldsymbol{s}_{(1)}(\boldsymbol{q}, \tau) = \boldsymbol{x}_{\rm fl}^{(1)}(\tau) - \boldsymbol{q} = -\frac{\boldsymbol{\nabla}}{\nabla^2}\delta^{(1)}(\boldsymbol{q}, \tau) \tag{B.16}$$

is the first-order Lagrangian displacement, and

$$K_{ij}^{(1)} = \left[\frac{\partial_i \partial_j}{\nabla^2} - \frac{1}{3}\delta_{ij}\right]\delta^{(1)} \tag{B.17}$$

is the first-order tidal field [Eq. (2.21)]. The third-order expressions in real space are given in Eq. (B.46) together with Eq. (B.48) below.

Finally, we give the result for the next-to-leading (NLO, or 1-loop) contribution to the matter and velocity power spectra. We will drop the time argument for clarity. It can be easily restored in the linear power spectra and perturbation theory kernels, if desired. The NLO matter power spectrum is given by $P_{mm}(k) = P_{\rm L}(k) + P_{mm}^{\rm NLO}(k)$, where

$$P_{mm}^{\rm NLO}(k) = P_{mm}^{(22)}(k) + 2P_{mm}^{(13)}(k), \tag{B.18}$$

$$P_{mm}^{(22)}(k) \equiv \langle\delta^{(2)}(\boldsymbol{k})\delta^{(2)}(\boldsymbol{k}')\rangle' = 2\int_{\boldsymbol{p}} [F_2(\boldsymbol{p}, \boldsymbol{k}-\boldsymbol{p})]^2 P_{\rm L}(p)P_{\rm L}(|\boldsymbol{k}-\boldsymbol{p}|) = \mathcal{I}^{[\delta^{(2)},\delta^{(2)}]}(k)$$

$$P_{mm}^{(13)}(k) \equiv \langle\delta^{(1)}(\boldsymbol{k})\delta^{(3)}(\boldsymbol{k}')\rangle' = 3P_{\rm L}(k)\int_{\boldsymbol{p}} F_3(\boldsymbol{p}, -\boldsymbol{p}, \boldsymbol{k})P_{\rm L}(p), \tag{B.19}$$

where $\mathcal{I}^{[O,O']}(k)$ is defined in Eq. (4.23), and the explicit expression for $P_{mm}^{(13)}(k)$ can be found in [859]. Similarly, the NLO matter-velocity cross-power spectrum and velocity-velocity power spectrum are given by $P_{m\theta}(k) = -f\mathcal{H}P_{\rm L}(k) + P_{m\theta}^{\rm NLO}(k)$ and $P_{\theta\theta}(k) = (f\mathcal{H})^2 P_{\rm L}(k) + P_{\theta\theta}^{\rm NLO}(k)$, where

$$P_{m\theta}^{\rm NLO}(k) = P_{m\theta}^{(22)}(k) + 2P_{m\theta}^{(13)}(k), \tag{B.20}$$

$$P_{m\theta}^{(22)}(k) \equiv \langle\delta^{(2)}(\boldsymbol{k})\theta^{(2)}(\boldsymbol{k}')\rangle' = -2\mathcal{H}f\int_{\boldsymbol{p}} F_2(\boldsymbol{p}, \boldsymbol{k}-\boldsymbol{p})G_2(\boldsymbol{p}, \boldsymbol{k}-\boldsymbol{p})P_{\rm L}(p)P_{\rm L}(|\boldsymbol{k}-\boldsymbol{p}|)$$

$$P_{m\theta}^{(13)}(k) \equiv \frac{1}{2}\left[\langle\theta^{(1)}(\boldsymbol{k})\delta^{(3)}(\boldsymbol{k}')\rangle' + \langle\delta^{(1)}(\boldsymbol{k})\theta^{(3)}(\boldsymbol{k}')\rangle'\right] = -\frac{1}{2}\left[(\mathcal{H}f)P_{mm}^{(13)}(k) + (\mathcal{H}f)^{-1}P_{\theta\theta}^{(13)}(k)\right]$$

$$P_{\theta\theta}^{\rm NLO}(k) = P_{\theta\theta}^{(22)}(k) + 2P_{\theta\theta}^{(13)}(k), \tag{B.21}$$

$$P_{\theta\theta}^{(22)}(k) \equiv \langle\theta^{(2)}(\boldsymbol{k})\theta^{(2)}(\boldsymbol{k}')\rangle' = 2(\mathcal{H}f)^2\int_{\boldsymbol{p}} [G_2(\boldsymbol{p}, \boldsymbol{k}-\boldsymbol{p})]^2 P_{\rm L}(p)P_{\rm L}(|\boldsymbol{k}-\boldsymbol{p}|)$$

$$P_{\theta\theta}^{(13)}(k) \equiv \langle\theta^{(1)}(\boldsymbol{k})\theta^{(3)}(\boldsymbol{k}')\rangle' = 3(\mathcal{H}f)^2 P_{\rm L}(k)\int_{\boldsymbol{p}} G_3(\boldsymbol{p}, -\boldsymbol{p}, \boldsymbol{k})P_{\rm L}(p). \tag{B.22}$$

The accuracy of the SPT predictions for the nonlinear matter power spectrum and bispectrum has been studied extensively using simulations [283, 860, 861, 287, 862, 248, 101, 127, 269, 612, 863]. Moreover, Ref. [101] showed that the accuracy improves significantly at higher redshifts ($z > 1$) where the nonlinear scale is smaller (see below).

Appendix B.2   Feynman rules of large-scale structure

The calculation of matter and galaxy statistics in perturbation theory can be performed efficiently by using a diagrammatic representation. The Feynman rules are as follows (see [344] for a detailed description):

1. An n-point correlation function is represented by a collection of diagrams with n outgoing external legs.

2. Interaction vertices have $m \geq 2$ ingoing lines $\boldsymbol{p}_1, \cdots, \boldsymbol{p}_m$ coupling to a single outgoing line $\boldsymbol{p}$. Each such vertex is assigned a factor

$$m!F_m(\boldsymbol{p}_1, \cdots, \boldsymbol{p}_m)(2\pi)^3\delta_D(\boldsymbol{p} - \boldsymbol{p}_{1\cdots m}). \tag{B.23}$$

One assigns a positive (negative) sign to outgoing (ingoing) momenta. Each ingoing line has to be

directly connected to a propagator (linear power spectrum).[26]

3. Propagators are represented as vertices with 2 outgoing lines of opposite momentum $\pm\boldsymbol{k}$, ,
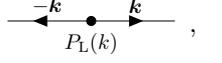
   and they are assigned a factor $P_{\mathrm{L}}(k)$.

4. All momenta that are not fixed in terms of momentum constraints are integrated over via

$$\int_{\boldsymbol{p}} \equiv \int \frac{d^3\boldsymbol{p}}{(2\pi)^3} \,. \tag{B.24}$$

   A diagram without any loop integral is said to be a leading-order (LO), or tree-level diagram.

5. Each diagram is multiplied by the *symmetry factor*, which accounts for the number of all nonequivalent labelings of external lines and degenerate configurations of the diagram.

As an example, the NLO contribution to the matter power spectrum [Eq. (B.18)] can be represented as:

$$P_{mm}^{\mathrm{NLO}}(k) = P_{mm}^{(22)}(k) + 2P_{mm}^{(13)}(k) = \quad \text{} \quad . \tag{B.25}$$
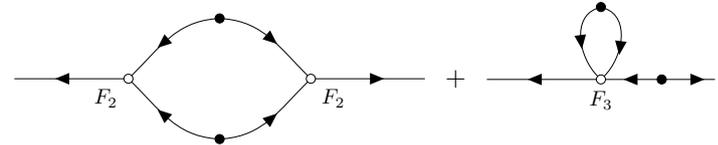
Appendix B.3   Effective field theory and the nonlinear scale

The pressureless fluid equations Eqs. (B.1)–(B.2) that we have considered so far are not strictly correct, as they neither take into account shell crossing of the dark matter, nor the presence of pressure in the baryonic component. In reality, dark matter is governed by the collisionless Boltzmann, or Vlasov equation, which predicts that multi-streaming occurs on small scales. Indeed, Eqs. (B.1)–(B.2) are obtained from the Vlasov equation by truncating the hierarchy of velocity moments, and dropping the second- and higher-order moments, which contain the effective pressure and anisotropic stress. The pressure of the baryon fluid, on the other hand, cannot be neglected on small scales. The Effective Field Theory approach to Large-Scale Structure (EFTofLSS [83, 84]) provides a rigorous approach to take into account these beyond-pressureless-perfect-fluid contributions from small-scale perturbations. Essentially, this can be seen as a bias expansion for a specific tracer that obeys stress-energy conservation. The latter in fact ensures that Eqs. (B.1)–(B.2) are only corrected by higher-derivative contributions.

The derivation of the EFT contributions proceeds by smoothing the density $\delta_\Lambda(\boldsymbol{x}, \tau)$ and velocity $\boldsymbol{v}_\Lambda(\boldsymbol{x}, \tau)$ fields on the arbitrary scale Λ, retaining only modes $k \lesssim \Lambda$ (see Sec. 2.10). While this erases the small-scale perturbations, the latter contribute stochastic terms, and moreover are modulated by δ_Λ and $\boldsymbol{v}_\Lambda$, leading to additional long-wavelength contributions. In the end, one obtains a contribution $-\partial_j \tau^{ij}/\rho_m$ on the right-hand-side of the Euler equation, where the effective stress tensor τ_{ij} captures the pressure and viscosity forces induced by the small-scale fluctuations. Expanding this to leading order in the large-scale fluctuations, the effective stress tensor can be written as [83, 84]

$$[\tau_{ij}]_\Lambda = p_{\mathrm{eff}}(\Lambda)\delta_{ij} + \overline{\rho}_m \left[c_s^2(\Lambda)\delta_\Lambda\delta_{ij} - \frac{c_{bv}^2(\Lambda)}{aH}\delta_{ij}\boldsymbol{\nabla}\cdot\boldsymbol{v}_\Lambda - \frac{3}{4}\frac{c_{sv}^2(\Lambda)}{aH}\left(\partial_i v_\Lambda^j + \partial_j v_\Lambda^i - \frac{3}{2}\delta_{ij}\boldsymbol{\nabla}\cdot\boldsymbol{v}_\Lambda \right) \right] + \cdots . \tag{B.26}$$

Here, p_{eff}, c_s, c_{bv}, c_{sv} are, respectively, effective pressure, adiabatic sound speed, bulk viscosity coefficient, and shear viscosity coefficient, which depend on Λ. Note that p_{eff} leads to a stochastic contribution to the matter velocity $v^i(\boldsymbol{k})$ which in Fourier space is proportional to ik^i (see also Sec. 2.8). Since the quantities c_s, c_{bv}, c_{sv} are due to the dependence of the small-scale density and velocity fields on the large-scale en-

---

[26]This is because diagrams that involve interaction vertices directly connected to each other are absorbed into higher-order interaction vertices.

vironment, they can in principle be calculated numerically by following a peak-background split argument similar to the one discussed in Sec. 3. Further, the effective stress tensor τ_{ij} can be expanded up to any order in terms of a basis of counter-terms that is constructed out of the same $\Pi_{ij}^{[n]}$ that are employed in the general Eulerian bias expansion in Sec. 2.5.3 [344].

Up to third order in perturbations, which is the highest order that is relevant for the expressions given in this review, there is only a single EFT contribution:

$$\delta_\Lambda^{(3)}\Big|_{\text{c.t.}}(\boldsymbol{k},\tau) = -\alpha(\tau,\Lambda)D^3(\tau)\frac{k^2}{k_{\text{NL}}^2}\delta^{(1)}(\boldsymbol{k},\tau)W_\Lambda(\boldsymbol{k})\,, \tag{B.27}$$

where "c.t." stands for counter-term, W_Λ denotes the smoothing kernel used to integrate out the small-scale modes, and k_{NL} is the nonlinear scale which we will define below. The free coefficient α cannot be predicted from perturbation theory and needs to be determined by fitting to simulations or observations. Essentially, it corresponds to an effective sound speed. We can now choose a vanishingly small smoothing scale $\Lambda^{-1} \to 0$, in which case we obtain another contribution to the NLO matter power spectrum [Eq. (B.18)],

$$P_{mm}^{\text{NLO}}\Big|_{\text{c.t.}}(k,\tau) = -2D^2(\tau)(2\pi)c_s^2\frac{k^2}{k_{\text{NL}}^2}P_{\text{L}}(k,\tau)\,, \tag{B.28}$$

where we have traded $\lim_{\Lambda^{-1}\to 0}\alpha(\Lambda)$ for c_s^2. This results in the contribution given in Eq. (4.22), where we have absorbed the growth factor $D^2(\tau)$ into $c_{s,\text{eff}}^2$. Furthermore, by allowing c_s to be fit to simulations, this correction term absorbs the dependence of the SPT contribution $P_{mm}^{(13)}$ in Eq. (B.18) on fully nonlinear small-scale modes that cannot be described by perturbation theory. Empirically, one finds values for c_s of order 1 (e.g., [85]). Note that the leading stochastic EFT contribution to the matter power spectrum scales as $k^4 P_{\varepsilon_m}^{\{4\}}(\tau)$ [346], where $\varepsilon_m \propto p_{\text{eff}}$ and $P_{\varepsilon_m}^{\{4\}}(\tau)$ is its power spectrum amplitude in the low-k limit which only depends on time. The scaling with k^4 is required by mass and momentum conservation [71, 344]. In terms of the perturbative counting of terms, this is a higher-order contribution; however, ε_m is relevant in the NLO halo-matter cross-power spectrum (Sec. 4.1.4).

We now turn to the general expectation for how large higher-order nonlinear terms to the n-point functions of matter and biased tracers are at a given scale k. The most important scale is the *nonlinear scale* which we here define as the scale where the dimensionless matter power spectrum $k^3 P_{\text{L}}(k)/(2\pi^2)$ becomes unity. Note that various different choices are possible here, since this is only a rough estimate for the scale where higher-order corrections become of order one. For our reference cosmology, this corresponds to $k_{\text{NL}}(z = 0) \simeq 0.25\,h\,\text{Mpc}^{-1}$. At and below this scale, $P_{\text{L}}(k)$ can be approximated by a power law $P_{\text{L}}(k) \propto k^n$ with $n = -1.9\cdots -1.7$ depending on the precise fit range chosen. As discussed in Sec. 4.1.4, higher-order SPT contributions to the n-point functions roughly scale as $(k/k_{\text{NL}})^{n+3} \sim (k/k_{\text{NL}})^{1.1\cdots 1.3}$, while the leading EFT term scales as $(k/k_{\text{NL}})^2$, for $c_s = \mathcal{O}(1)$.

Appendix B.4   IR resummation

The perturbative expansion and counting of terms discussed so far in this Appendix (as well as in Sec. 4.1.4) relies on the assumption of a smooth, slowly varying linear matter power spectrum. In reality, the plasma oscillations in the baryon-photon fluid prior to recombination imprint a small-amplitude oscillatory pattern in the linear matter power spectrum, the well-known BAO feature first pointed out by [41], with oscillation period $k_{\text{BAO}} = 2\pi/r_s \approx 0.06\,h\,\text{Mpc}^{-1}$, where $r_s \approx 106\,h^{-1}\,\text{Mpc}$ is the sound horizon at recombination. These oscillations are damped towards higher k by Silk damping.

Now consider a perturbation with wavenumber $p \sim k_{\text{BAO}}$. This mode induces a displacement $\boldsymbol{s}$ which moves small-scale perturbations with $k \gg p$. Perturbatively, this effect is captured at second order by the contribution $\boldsymbol{s}^i\partial_i\delta$. However, the effect of higher-order displacement terms on the oscillatory ("wiggle") part of the power spectrum is not suppressed, since the displacement itself is modulated over scales of order r_s. Fortunately, since this is a pure displacement effect, i.e., a coordinate transformation, these terms can be resummed [269, 270, 271, 272]. This resummation relies on the fact that (i) the relevant modes sourcing the displacement are in the perturbative regime ($p \lesssim 0.2\,h\,\text{Mpc}^{-1}$), and ($ii$) that the shape of the oscillatory

feature is well known. Crucially, only the long-wavelength displacement effects on this oscillatory feature are resummed, as they are the ones which are enhanced.

First, we split the linear power spectrum into smooth (s) and oscillatory parts (w) via

$$P_{\rm L}(k) = P_{\rm L}^s(k) + P_{\rm L}^w(k) \,, \tag{B.29}$$

where

$$P_{\rm L}^w(k) = \sin\left(\frac{k}{k_{\rm BAO}}\right) f_{\rm env}(k) \tag{B.30}$$

is a strictly oscillating function whose amplitude is controlled by the smooth envelope $f_{\rm env}(k)$, while $P_{\rm L}^s(k)$ is defined as the remainder. Then, the IR resummation acting on the oscillatory part yields [272][27] (see [270, 271, 864] for similar results)

$$\begin{aligned} P_{mm}^{\rm IR\text{-}resum.}(k) = P_{\rm L}^s(k) + \exp\left[-\Sigma^2(\epsilon k)k^2\right]\left(1 + \Sigma^2(\epsilon k)k^2\right) P_{\rm L}^w(k) \\ + P_{mm}^{\rm NLO}\left[P_{\rm L}(k') \to P_{\rm L}^s(k') + \exp\left(-\Sigma^2(\epsilon k')k'^2\right) P_{\rm L}^w(k')\right](k) \,. \end{aligned} \tag{B.31}$$

Here, $P_{mm}^{\rm NLO}[\cdots](k)$ stands for the NLO matter power spectrum Eq. (B.18) evaluated using the IR-resummed version of the linear power spectrum. Finally,

$$\Sigma^2(\Lambda) = \frac{1}{6\pi^2}\int_0^\Lambda dp\, P_{\rm L}(p)\left[1 - j_0(pr_s) + 2j_2(pr_s)\right] \tag{B.32}$$

is essentially the covariance $\langle s^i s^j \rangle$ of the linear Lagrangian displacement, after angle-averaging and smoothing with a sharp-k filter at the scale Λ. $\epsilon \ll 1$ is a parameter controlling down to what scale the IR resummation should occur; Ref. [271] choose $\epsilon = 1/2$. The term $+\Sigma^2 k^2$ in parentheses in the first line of Eq. (B.31) prevents the double counting of terms that appear in the NLO expression on the second line. As shown in [270, 271, 864, 272], the IR-resummed matter power spectrum and its Fourier transform, the correlation function, significantly improve the match in the shape of these correlations with full N-body simulations on the scales around the BAO feature ($r \sim r_s$ in real space; $0.05\,h\,{\rm Mpc}^{-1} \lesssim k \lesssim 0.3\,h\,{\rm Mpc}^{-1}$ in Fourier space).

At this, leading order in the wavelength $1/p$ of the soft modes that lead to displacements, the displacements are the same for both matter and LSS tracers, due to the equivalence principle (Sec. 2.7; see also [865]). This means that we can immediately make use of the IR resummation in the prediction for halo and galaxy clustering in real space, by generalizing Eq. (B.31) to:

$$\begin{aligned} P_{hm}^{\rm IR\text{-}resum.}(k) = b_1 P_{\rm L}^s(k) + b_1 \exp\left[-\Sigma^2(\epsilon k)k^2\right]\left(1 + \Sigma^2(\epsilon k)k^2\right) P_{\rm L}^w(k) \\ + P_{hm}^{\rm NLO}\left[P_{\rm L}(k') \to P_{\rm L}^s(k') + \exp\left(-\Sigma^2(\epsilon k')k'^2\right) P_{\rm L}^w(k')\right](k) \,, \end{aligned} \tag{B.33}$$

and

$$\begin{aligned} P_{hh}^{\rm IR\text{-}resum.}(k) = b_1^2 P_{\rm L}^s(k) + b_1^2 \exp\left[-\Sigma^2(\epsilon k)k^2\right]\left(1 + \Sigma^2(\epsilon k)k^2\right) P_{\rm L}^w(k) + P_\varepsilon^{\{0\}} \\ + P_{hh}^{\rm NLO}\left[P_{\rm L}(k') \to P_{\rm L}^s(k') + \exp\left(-\Sigma^2(\epsilon k')k'^2\right) P_{\rm L}^w(k')\right](k) \,, \end{aligned} \tag{B.34}$$

where the second line in each equation is obtained by replacing the linear power spectrum with its IR-resummed version in Eq. (4.22). Again, this is because the IR resummation only deals with displacement terms which, at leading order in derivatives, are only modified by biasing through the trivial factor b_1. This holds analogously for the displacement from the galaxy rest frame into redshift space (Sec. 9.3.2), whose IR contributions can similarly be resummed. Refs. [89, 180, 866, 867, 865] provide resummed results for two-point functions in redshift space.

---

[27]Eq. (7.4) there; we have dropped the 2-loop-order displacement terms $\propto D^6$ in that expression, since they are numerically small. Note also the different Fourier convention adopted in [272].

Appendix B.5   Convective SPT approach and conserved evolution at third order

In this section, we provide some details on the calculation of the evolution of a conserved biased tracer at third order in perturbation theory that is briefly presented in Sec. 2.4. The "convective SPT" system, consisting of the continuity and Euler equation Eqs. (B.1)–(B.2) for matter and the continuity equation Eq. (2.17) for the biased tracer, can be written in compact form as

$$\frac{\mathrm{D}}{\mathrm{D}\tau}\boldsymbol{\Psi} = -\boldsymbol{\sigma}\cdot\boldsymbol{\Psi} + \boldsymbol{S} \tag{B.35}$$

where

$$\boldsymbol{\Psi}(\boldsymbol{x},\tau) = \begin{pmatrix} \delta_g(\boldsymbol{x},\tau) \\ \delta(\boldsymbol{x},\tau) \\ \theta(\boldsymbol{x},\tau) \end{pmatrix}; \quad \boldsymbol{\sigma}(\tau) = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & \frac{3}{2}\Omega_m\mathcal{H}^2 & \mathcal{H} \end{pmatrix}; \quad \boldsymbol{S}(\boldsymbol{x},\tau) = \begin{pmatrix} -\delta_g\,\theta \\ -\delta\,\theta \\ -(\partial_i v_j)^2 \end{pmatrix}. \tag{B.36}$$

The fact that $\boldsymbol{\sigma}$ is degenerate already shows that the three equations are not really coupled, but rather the equation for δ_g can be integrated separately as done in Sec. 2.3. Eq. (B.35) allows for a convenient, compact derivation of both δ and δ_g. In the following, we will solve this system up to third order.

Source term

The goal is to integrate Eq. (B.35) along the fluid trajectory. However, the source term $\boldsymbol{S}$ involves some subtleties, in particular the third component $S_3 = (\partial_i v_j)^2$, since the derivative is with respect to Eulerian coordinate $\boldsymbol{x}$. The velocity shear $\partial_i v_j$ is nonlocally related to the degrees of freedom δ, θ themselves [since $\boldsymbol{v} = (\boldsymbol{\nabla}/\nabla^2)\theta$]. We can use a trick, namely the fact that the displacement is the integral of the peculiar velocity along the fluid trajectory $\boldsymbol{x}_{\mathrm{fl}}(\boldsymbol{q},\tau)$, and therefore

$$\boldsymbol{v}(\boldsymbol{x}_{\mathrm{fl}}[\boldsymbol{q},\tau],\tau) = \frac{\mathrm{D}}{\mathrm{D}\tau}\boldsymbol{s}(\boldsymbol{q},\tau)\,. \tag{B.37}$$

In the remainder of this section, we will abbreviate convective derivatives $\mathrm{D}/\mathrm{D}\tau$ as primes. Further, we need to transform the derivative from $\boldsymbol{x}$ to the fluid flow or Lagrangian coordinate $\boldsymbol{q}$ via

$$\partial_x^i = \left([\mathbf{1}+\boldsymbol{M}]^{-1}\right)^i_{\ j}\partial_q^j\,, \tag{B.38}$$

where $M^{ij} = \partial_q^i s^j$. At second order, we then obtain

$$\partial_x^i v^j = \left[\partial_q^i + (\partial_q^i s_k)\partial_q^k\right]s'^j + \mathcal{O}([\delta^{(1)}]^3)\,. \tag{B.39}$$

Here, the left-hand side is defined at the Eulerian coordinate $(\boldsymbol{x},\tau)$ while the right-hand side quantities are defined at the corresponding Lagrangian coordinate $\boldsymbol{q}$. With these expressions, we can construct source terms that are expressed purely in terms of convective time and Lagrangian spatial derivatives. Specifically, the third component of the source term, S_3, is at first and second order given by (App. B in [131])

$$(S^{(1)})_3 = -[a'(\tau)]^2\left[(K^{(1)})^2 + \frac{1}{3}(\delta^{(1)})^2\right]_{\tau_0}$$

$$(S^{(2)})_3 = 2a\,[a'(\tau)]^2\left[-\frac{2}{3}\delta^{(1)}\sigma^{(2)} - 2K^{(1)}_{ij}\mathcal{D}^{ij}\sigma^{(2)} + \frac{1}{9}(\delta^{(1)})^3 + \delta^{(1)}(K^{(1)})^2 + (K^{(1)})^3\right]_{\tau_0}, \tag{B.40}$$

where

$$\sigma^{(2)} \equiv \partial_k s^{(2)k} = \frac{1}{2}\left[-\frac{2}{7}(\delta^{[1]})^2 + \frac{3}{7}(K^{[1]})^2\right] \tag{B.41}$$

is the divergence of the second-order Lagrangian displacement.

235

Initial conditions

In order to integrate Eq. (B.35), we need to provide an expression for the initial conditions at some initial time τ_*. The quantities δ^*, θ^* are obtained by integrating only the matter part $[\boldsymbol{\Psi}]_{i=2,3}$ from $\tau = 0$ to $\tau = \tau_*$. Throughout, we only consider the fastest-growing-mode solutions for the density and velocity field. That is, the final time τ is assumed to be late enough so that all the slowly growing modes are subdominant [503, 505].

Under the assumptions explained in Sec. 2.4, the galaxy density δ_g^* at the "formation time" is given by Eq. (2.44), which we repeat here for convenience:

$$
\delta_g^* \equiv \delta_g(\boldsymbol{x}_{\mathrm{fl}}(\tau_*), \tau_*) = \sum_{n=1}^{3} \frac{b_n^*}{n!}[\delta^*]^n + b_{K^2}^* \operatorname{tr}\left[(K_{ij}^*)^2\right] + b_{K^3}^* \operatorname{tr}\left[(K_{ij}^*)^3\right] + b_{\delta K^2}\delta^* \operatorname{tr}\left[(K_{ij}^*)^2\right]
$$
$$
+ \varepsilon^* + \varepsilon_\delta^*\delta^* + \varepsilon_{\delta^2}^*[\delta^*]^2 + \varepsilon_{K^2}^* \operatorname{tr}\left[(K_{ij}^*)^2\right] , \tag{B.42}
$$

where here and throughout a superscript $*$ indicates that a quantity is evaluated at $\boldsymbol{x}_* \equiv \boldsymbol{x}_{\mathrm{fl}}(\tau_*)$ and τ_*. The stochastic fields ε^*, ε_X^* ($X = \delta$, δ^2, K^2) are assumed to be first-order random variables.

Solution

We define the matrix $\boldsymbol{A}(\tau, \tau_i)$ as the solution to the following matrix ODE with boundary condition

$$
\partial_\tau \boldsymbol{A}(\tau, \tau_i) + \boldsymbol{\sigma}(\tau)\boldsymbol{A}(\tau, \tau_i) = 0; \quad \boldsymbol{A}(\tau, \tau) = \boldsymbol{1} . \tag{B.43}
$$

In order to obtain closed analytical expressions, we restrict to an EdS Universe. In this case, $\boldsymbol{A}$ assumes a simple form constructed out of the growing and decaying modes which scale as $a(\tau) \propto \tau^2$ and $H(\tau) \propto \tau^{-3}$, respectively (App. B in [131]). In the final result, we then replace scale factors $a(\tau)$ with growth factors $D(\tau)$, as described after Eq. (B.12).

Given the initial condition for Eq. (B.35) from above, $\boldsymbol{\Psi}(\boldsymbol{x}_*, \tau_*) = (\delta_g^*, \delta^*, \theta^*)$ with Eq. (B.42), the particular solution to Eq. (B.35) is then given by

$$
\boldsymbol{\Psi}(\boldsymbol{x}, \tau) = \boldsymbol{A}(\tau, \tau_*)\begin{pmatrix} \delta_g^* \\ \delta^* \\ \theta^* \end{pmatrix} + \int_{\tau_*}^{\tau} d\tau' \, \boldsymbol{A}(\tau, \tau')\boldsymbol{S}(\boldsymbol{x}_{\mathrm{fl}}(\tau'), \tau') . \tag{B.44}
$$

The interpretation of Eq. (B.44) is clear: the density and velocity of the fluid and galaxy at position $(\boldsymbol{x}, \tau)$ is given by an integral over the fluid trajectory of the source term multiplied by the Green's function $\boldsymbol{A}(\tau, \tau')$. Eq. (B.44) leads to the SPT result for matter, in particular for the terms that are invariant under time-dependent coordinate shifts $\boldsymbol{x} \to \boldsymbol{x} + \boldsymbol{\xi}(\tau)$. The non-invariant terms, which are not captured by Eq. (B.44), are precisely the displacement terms relating the solutions in Lagrangian and Eulerian coordinates. Finally, expanding in these displacements to the same perturbative order then yields exactly the SPT result. Thus, the convective SPT calculation keeps track of the displacement terms exactly (i.e., it resums them), similar to Lagrangian perturbation theory.

At second order, we then recover our previous result, Eq. (2.33) as well as Eqs. (B.14)–(B.15). Specifically, the solution reads

$$
\boldsymbol{\Psi}^{(2)}(\boldsymbol{x}, \tau) = \begin{pmatrix} b_1^E(\tau)\delta^{(2)} + b_2^E(\tau)(\delta^{(1)})^2/2 + b_{K^2}^E(\tau)(K_{ij}^{(1)})^2 + \varepsilon_\delta^E(\tau)\delta^{(1)} \\ D^2(\tau)\left[\frac{17}{21}(\delta^{(1)})^2 + \frac{2}{7}(K_{ij}^{(1)})^2\right] \\ -D(\tau)\dot{D}(\tau)\left[\frac{13}{21}(\delta^{(1)})^2 + \frac{4}{7}(K_{ij}^{(1)})^2\right] \end{pmatrix}_{\boldsymbol{q}} . \tag{B.45}
$$

---

[28]Note that $\theta^{(3)}$ in Eq. (B23) of [131] lacks an overall minus sign.

At third order, we obtain for the matter fields [131][28]

$$\delta^{(3)} = \frac{341}{567}\delta^3 + \frac{11}{21}K^2\delta + \frac{2}{9}K^3 + \frac{1}{6}O_{\text{td}}$$

$$\theta^{(3)} = -f\mathcal{H}\left[\frac{71}{189}\delta^3 + \frac{5}{7}K^2\delta + \frac{2}{3}K^3 + \frac{1}{2}O_{\text{td}}\right], \tag{B.46}$$

where K^n stands for $\text{tr}[K_{ij}^n]$, and all quantities on the right-hand side are evaluated at linear order and at τ. O_{td} is defined in Eq. (2.50). Note that, although all quantities appearing in Eq. (B.46) are defined using derivatives with respect to $\boldsymbol{x}$, the distinction between Eulerian and Lagrangian derivatives in these terms only matters at higher order. Notice further that the coefficient of the δ^3 term in $\delta^{(3)}$ is exactly the third-order coefficient of the perturbative expansion of spherical collapse in an EdS Universe [Eq. (2.14)], as expected: for a spherically symmetric perturbation, $K_{ij} = 0$, and $\delta^{(3)}$ has to reduce to the cubic term in the perturbative expansion of the spherical collapse solution. To the authors' knowledge, despite its fairly simple form, Ref. [131] was the first to give the complete expression for $\delta^{(3)}$, $\theta^{(3)}$ in real space. Interestingly, the expression for the nonlinear density derived by Ref. [136] from ellipsoidal collapse [444, 435] does not contain a term of the form $K^{ij}\mathcal{D}_{ij}\sigma^{(2)}$, while it matches the PT result at second order. The same is found in the local tidal approximation (LTA) of [445], as well as the fully relativistic local approximation derived in [150]. The underlying reason is that the evolution equations of an isolated ellipsoidal perturbation are only approximate, since they neglect the dynamical interaction with the large-scale environment. When allowing for general configurations, the nonlinear evolution of tidal fields in cosmology is nonlocal [868, 148, 149].

Finally, in order to obtain the density at a fixed order in standard Eulerian perturbation theory, we need to displace $\boldsymbol{\Psi}$ from a fixed Lagrangian position to the Eulerian position, by expanding in the argument. Let us define the Eulerian solution $\boldsymbol{\Psi}^E$ through

$$\boldsymbol{\Psi}^E(\boldsymbol{x}) = \boldsymbol{\Psi}(\boldsymbol{x} - \boldsymbol{s}[\boldsymbol{q}, \tau]). \tag{B.47}$$

$\boldsymbol{\Psi}$ on the right-hand side contains all invariant terms, which we have obtained above. We now perform a Taylor expansion in $\boldsymbol{s}$ as well as a perturbative expansion of $\boldsymbol{s}$, noting that $\boldsymbol{s}$ is itself a function of the Lagrangian rather than Eulerian position. We obtain at third order

$$\boldsymbol{\Psi}^{E,(3)}(\boldsymbol{x}, \tau) = \boldsymbol{\Psi}^{(3)} - s_{(1)}^i \partial_i \boldsymbol{\Psi}^{(2)} - \left[s_{(2)}^i - s_{(1)}^j (\partial_j s_{(1)}^i)\right] \partial_i \boldsymbol{\Psi}^{(1)} + \frac{1}{2}s_{(1)}^i s_{(1)}^j \partial_i \partial_j \boldsymbol{\Psi}^{(1)}, \tag{B.48}$$

where on the right-hand side all quantities are evaluated at $\boldsymbol{x}$, τ, and $\boldsymbol{\Psi}^{(n)} = \{\delta_g^{(n)}, \delta^{(n)}, \theta^{(n)}\}$, are given in Eq. (B.45) for $n = 2$, and Eq. (2.49), Eq. (B.46) for $n = 3$, respectively. As expected, the displacement from Lagrangian to Eulerian position, being merely a coordinate shift, does not affect the bias relation at any order.

Appendix B.6   Conserved evolution and bias expansion beyond the EdS background

We now allow for a more general expansion history, in particular ΛCDM or quintessence (neglecting perturbations in the dark energy component), and derive how this changes the evolution equations of a conserved tracer in the convective SPT approach. For this, we transform the time coordinate in the equations of motion Eq. (2.43) to $\ln D$, where D is the linear growth factor [Eq. (B.9)]. Since $d\ln D/d\tau = \mathcal{H}d\ln D/d\ln a = \mathcal{H}f$ by definition of the growth rate f, we introduce

$$\tilde{\boldsymbol{v}} \equiv (\mathcal{H}f)^{-1}\boldsymbol{v} \qquad \text{and} \qquad \frac{\text{D}}{\text{D}\ln D} \equiv \frac{\partial}{\partial \ln D} + \tilde{v}^i \partial_i. \tag{B.49}$$

Note that $\tilde{\theta} = \partial_i \tilde{v}^i = -\delta$ at linear order. Using Eq. (B.11), we then obtain

$$\frac{\text{D}}{\text{D}\ln D}\tilde{\boldsymbol{\Psi}} = -\tilde{\boldsymbol{\sigma}} \cdot \tilde{\boldsymbol{\Psi}} + \tilde{\boldsymbol{S}} \tag{B.50}$$

$$\tilde{\boldsymbol{\Psi}}(\boldsymbol{x}, \tau) = \begin{pmatrix} \delta_g(\boldsymbol{x}, \tau) \\ \delta(\boldsymbol{x}, \tau) \\ \tilde{\theta}(\boldsymbol{x}, \tau) \end{pmatrix}; \quad \tilde{\boldsymbol{\sigma}}(\tau) = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & \frac{3}{2}\frac{\Omega_m}{f^2} & \frac{3}{2}\frac{\Omega_m}{f^2} - 1 \end{pmatrix}; \quad \tilde{\boldsymbol{S}}(\boldsymbol{x}, \tau) = \begin{pmatrix} -\delta_g\tilde{\theta} \\ -\delta\tilde{\theta} \\ -(\partial_i\tilde{\boldsymbol{v}}_j)^2 \end{pmatrix}. \tag{B.51}$$

We see that the equations maintain the same structure, and that the departure from EdS of the equations of motion is completely quantified by $\Omega_m(a)/f^2(a) - 1$. In particular, the time dependence of the second-order matter density field is, instead of $[D(\tau)]^2$, given by a second-order growth factor $D_2(\tau)$, obtained by integrating $[D(\tau)]^2$ against the Green's function corresponding to Eq. (B.51). Only starting at third order do we obtain two different time dependences in the contributions to δ at the same perturbative order; specifically, $D_{3,1}(\tau)$ and $D_{3,2}(\tau)$, obtained respectively by integrating $[D(\tau)]^3$ and $D_2(\tau)D(\tau)$ against the Green's function.

We now go back to the construction of the set of general local bias operators in Sec. 2.5.2–2.5.3. The first instance of a new operator being induced by the different time evolution in quintessence cosmology is at third order, where $\Pi_{ij}^{[3]}$ defined via Eq. (2.63) in Sec. 2.5.3 separates into two different operators $\Pi_{ij}^{[3,1]}$, $\Pi_{ij}^{[3,2]}$. Now, the reason for why we do not need to include convective time derivatives of δ itself, or equivalently $\mathrm{tr}[\Pi^{[n]}]$, in the perturbative bias expansion also holds in the more general case considered here: by including time derivatives of all other operators constructed out of $\partial_i\partial_j\Phi$, we already obtain the operators appearing in the time derivatives of δ, or equivalently in the perturbative contributions $\delta^{(2)}$, $\delta^{(3)}, \cdots$. Thus, we do not need to include $\mathrm{tr}[\Pi^{[3,i]}]$, and the first new operator in the Eulerian basis Eq. (2.64) of the bias expansion appears at fourth order, where

$$\mathrm{tr}[\Pi^{[1]}\Pi^{[3]}] \longrightarrow \mathrm{tr}[\Pi^{[1]}\Pi^{[3,1]}], \ \mathrm{tr}[\Pi^{[1]}\Pi^{[3,2]}]. \tag{B.52}$$

Conversely, in the EdS limit these two operators have to combine into $\mathrm{tr}[\Pi^{[1]}\Pi^{[3]}]$. This means that the bias coefficients of the two different fourth-order operators will be approximately equal, with a fractional difference scaling as $D_{3,1}(\tau)/D_{3,2}(\tau) - 1 \ll 1$. Thus, the bases of bias operators described in Sec. 2.5.2–2.5.3 can be straightforwardly extended beyond EdS to quintessence cosmologies. Moreover, all explicit results for galaxy and halo statistics derived in Sec. 4, which only rely on perturbative results up to third order, are independent of the EdS assumption. While the time evolution of bias parameters for conserved tracers [Eq. (2.52)] is modified from the EdS case, this is not of practical relevance if all the bias parameters are determined from the data at a fixed time.

Appendix C   Bias conventions and their relation

The general bias expansion consists of a set of operators $\{O_i^{[n]}\}$, $i = 1, \cdots N_n$, at each order n in perturbation theory (as discussed in Sec. 2.8, the stochastic terms involve the same basis of operators as well, hence we do not need to consider them explicitly here). A superscript $[n]$ indicates that the lowest-order contribution to this operator is at n-th order in perturbation theory. Of course, there are contributions at order $n + 1, n + 2$, and so on as well. Any linearly independent combination of the $O_i^{[n]}$ at a given order n leads to an equivalent bias expansion. For example, for $n = 1$, we only have a single operator $O^{[1]} = \delta$, but $O^{[1]} = \theta/\mathcal{H}$ or $O^{[1]} = \nabla^2\Phi/\mathcal{H}^2$ constitute equivalent choices. Correspondingly, several different conventions for the first few orders of the bias expansion have been used in the literature. We provide a brief summary of the relations between the most commonly used bias conventions here.

Let us denote the set of operators in a given basis at a fixed order n in perturbation theory as $\boldsymbol{O} = \{O_i^{[n]}\}_{i=1}^{N_n}$. The bias expansion at n-th order can then be treated as an N_n-dimensional vector space. Consequently, a different basis $\tilde{\boldsymbol{O}}$ can always be written as

$$\tilde{\boldsymbol{O}} = \boldsymbol{M} \cdot \boldsymbol{O}\,, \tag{C.1}$$

where $\boldsymbol{M}$ is an invertible $N_n \times N_n$ matrix. Then, the coefficients, i.e. the bias parameters which we write as $\boldsymbol{b} = \{b_{O_i^{[n]}}\}_{i=1}^{N_n}$, transform via the transpose of the inverse of $\boldsymbol{M}$:

$$\delta_g \supset \sum_{i=1}^{N_n} b_{O_i^{[n]}} O_i^{[n]} = \sum_{i=1}^{N_n} \tilde{b}_{\tilde{O}_i^{[n]}} \tilde{O}_i^{[n]} \quad \Rightarrow \quad \tilde{\boldsymbol{b}} = \left(\boldsymbol{M}^{-1}\right)^{\top} \cdot \boldsymbol{b}\,. \tag{C.2}$$

Note that even if one or more of the $\tilde{O}_i^{[n]}$ is the same as $O_i^{[n]}$, the corresponding bias parameter $\tilde{b}_{O_i^{[n]}}$ in general differs from $b_{O_i^{[n]}}$ due to the transformation in the N_n-dimensional vector space. We now derive these transformations for frequently used bias conventions at second and third order.

Appendix C.1   Second order

At second order, the basis used in this review consists of $\boldsymbol{O} = \{\delta^2, (K_{ij})^2\}$. Another frequently used basis is $\tilde{\boldsymbol{O}} = \{\delta^2, \mathcal{G}_2\}$, where

$$\mathcal{G}_2 \equiv \left[(\partial_i\partial_j\hat{\Phi})^2 - (\nabla^2\hat{\Phi})^2\right] = (K_{ij})^2 - \frac{2}{3}\delta^2\,, \tag{C.3}$$

where $\hat{\Phi} \equiv 2\Phi/(3\Omega_m\mathcal{H}^2) = \nabla^{-2}\delta$. We then obtain

$$\tilde{b}_{\delta^2} = b_{\delta^2} + \frac{2}{3}b_{K^2} \quad \text{and} \quad \tilde{b}_{\mathcal{G}_2} = b_{K^2}\,. \tag{C.4}$$

Finally, using the tensor

$$\Pi_{ij}^{[1]} \equiv \partial_i\partial_j\hat{\Phi} = K_{ij} + \frac{1}{3}\delta_{ij}\delta \tag{C.5}$$

introduced in Sec. 2.5.3, one can choose

$$\left(\text{tr}\left[\Pi^{[1]}\right]\right)^2 = \delta^2\,, \qquad \text{tr}\left[\left(\Pi^{[1]}\right)^2\right] = K^2 + \frac{1}{3}\delta^2 \tag{C.6}$$

as second-order basis. This yields

$$\tilde{b}_{(\text{tr}\,\Pi)^2} = b_{\delta^2} - \frac{1}{3}b_{K^2} \quad \text{and} \quad \tilde{b}_{\text{tr}[\Pi^2]} = b_{K^2}\,. \tag{C.7}$$

Appendix C.2   Third order

At third order, an operator nonlocally related to $\partial_i\partial_j\Phi$ appears, which can be defined in a variety of ways. In this review, we mostly use the definition in Eq. (2.50),

$$O_{\mathrm{td}}^{(3)} \equiv \frac{8}{21}K_{ij}^{(1)}\mathcal{D}^{ij}\left[(\delta^{(1)})^2 - \frac{3}{2}(K_{ij}^{(1)})^2\right]. \tag{C.8}$$

Note that, unlike the operators $\delta^2, (K_{ij})^2$, it is not obvious how the definition of this operator is to be extended to higher order in perturbations. We will return to this shortly. Let us relate $O_{\mathrm{td}}^{(3)}$ to the second-order tensor appearing in the general Eulerian bias expansion (Sec. 2.5.3),

$$\Pi_{ij}^{[2]} \equiv (\mathcal{H}f)^{-1}\frac{\mathrm{D}}{\mathrm{D}\tau}\Pi_{ij}^{[1]} - \Pi_{ij}^{[1]}. \tag{C.9}$$

Note that Eq. (C.9) is valid at any order in perturbation theory. At second order in PT, this yields

$$\Pi_{ij}^{[2]}\Big|^{(2)} = \partial_i\partial_k\hat{\Phi}\partial^k\partial_j\hat{\Phi} + \frac{10}{21}\frac{\partial_i\partial_j}{\nabla^2}\left(\delta^2 - \frac{3}{2}(K_{kl})^2\right). \tag{C.10}$$

This result can be derived in a variety of ways. Appendix C.3 gives some details on the Fourier-space derivation which makes use of the SPT kernel expansion (Appendix B.1). Eq. (C.10) immediately yields

$$\mathrm{tr}[\Pi^{[1]}\Pi^{[2]}]\Big|^{(3)} = \frac{5}{4}O_{\mathrm{td}}^{(3)} + (K_{ij})^3 + \frac{16}{21}\delta(K_{ij})^2 + \frac{17}{63}\delta^3. \tag{C.11}$$

This can now be used to define O_{td} at all orders in PT, although we do not need the result for any of the expressions presented in this review.

$O_{\mathrm{td}}^{(3)}$ can also be related to the second-order Lagrangian distortion tensor $M_{ij}^{(2)}$. Using that

$$M_{ij}^{(2)} = \frac{\partial_q^i\partial_q^j}{\nabla_q^2}(\boldsymbol{\nabla}_q \cdot \boldsymbol{s}_{(2)}) = [f\mathcal{H}\partial_\tau M_{ij} - M_{ij}]^{(2)} \tag{C.12}$$

$$\text{where}\quad \boldsymbol{\nabla}_q \cdot \boldsymbol{s}_{(2)} = -\frac{1}{7}\left[\left(\delta^{(1)}\right)^2 - \frac{3}{2}\left(K_{ij}^{(1)}\right)^2\right] = -\frac{3}{4}\left[\delta^{(2)} + \frac{\theta^{(2)}}{\mathcal{H}f}\right] \tag{C.13}$$

is the divergence of the second-order Lagrangian displacement, we obtain

$$O_{\mathrm{td}}^{(3)} = \frac{8}{3}\left[M^{(1)\,ij} - \frac{1}{3}\delta^{ij}\,\mathrm{tr}\,M^{(1)}\right]M_{ij}^{(2)}. \tag{C.14}$$

In Eq. (C.12) we have used that convective time derivatives reduce to ordinary time derivatives for operators defined in terms of Lagrangian coordinates.

Further, relating $O_{\mathrm{td}}^{(3)}$ to the velocity potential $\Phi_v \equiv -(f\mathcal{H})^{-1}\nabla^{-2}\theta$, we can also write [128, 197]

$$O_{\mathrm{td}}^{(3)} = \Gamma_3 + \frac{16}{63}\delta^3 - \frac{8}{21}\delta(K_{ij})^2,$$

$$\text{where}\quad \Gamma_3 \equiv (\partial_i\partial_j\hat{\Phi})^2 - (\nabla^2\hat{\Phi})^2 - \left[(\partial_i\partial_j\Phi_v)^2 - (\nabla^2\Phi_v)^2\right]. \tag{C.15}$$

Note that, at linear order, $\Phi_v^{(1)} = \Phi^{(1)}$ so that $\Gamma_3 = \Gamma_3^{[3]}$ starts at third order in PT. Specifically, we have at this order

$$\Gamma_3^{(3)} = 2K^{ij}\mathcal{D}_{ij}\left[\delta + (f\mathcal{H})^{-1}\theta\right]^{(2)} - \frac{4}{3}\delta\left[\delta + (f\mathcal{H})^{-1}\theta\right]^{(2)}, \tag{C.16}$$

where

$$\left[\delta + (f\mathcal{H})^{-1}\theta\right]^{(2)} = \frac{4}{21}\left[\delta^2 - \frac{3}{2}K^2\right]. \tag{C.17}$$

$M_{\tilde{O}O}$	δ^3	δK^2	K^3	$O_{\rm td}^{(3)}$
δ^3	1			
$\delta\mathcal{G}_2$	-2/3	1		
$\mathcal{G}_3$	-1/9	1/2	-1	
Γ_3	-16/63	8/21		1

$(M^{-1})^\top_{\tilde{O}O}$	b_{δ^3}	$b_{\delta K^2}$	b_{K^3}	$b_{\rm td}$
$\tilde{b}_{\delta^3}$	1	2/3	2/9	
$\tilde{b}_{\delta\mathcal{G}_2}$		1	1/2	-8/21
$\tilde{b}_{\mathcal{G}_3}$			-1	
$\tilde{b}_{\Gamma_3}$				1

Table C.14: Transformation between the third-order basis of operators $\boldsymbol{O} = \{\delta^3, \delta K^2, K^3, O_{\rm td}^{(3)}\}$ adopted in this review, and $\tilde{\boldsymbol{O}} = \{\delta^3, \delta\mathcal{G}_2, \mathcal{G}_3, \Gamma_3\}$ adopted in [197] (left). The right table gives the corresponding transformation of bias parameters. Vanishing matrix elements are indicated as empty cells for clarity.

$M_{\tilde{O}O}$	δ^3	δK^2	K^3	$O_{\rm td}^{(3)}$
δ^3	1			
δK^2		1		
st				-1/2
$\psi_{\rm MR}$	328/3969	-4/49	-8/63	-8/21

$(M^{-1})^\top_{\tilde{O}O}$	b_{δ^3}	$b_{\delta K^2}$	b_{K^3}	$b_{\rm td}$
$\tilde{b}_{\delta^3}$	1		41/63	
$\tilde{b}_{\delta K^2}$		1	-9/14	
$\tilde{b}_{st}$			6	-2
$\tilde{b}_{\psi_{\rm MR}}$			-63/8	

Table C.15: Transformation between the third-order basis of operators $\boldsymbol{O} = \{\delta^3, \delta K^2, K^3, O_{\rm td}^{(3)}\}$ adopted in this review, and $\tilde{\boldsymbol{O}} = \{\delta^3, \delta K^2, st, \psi_{\rm MR}\}$ adopted in [128] (left). The right table gives the corresponding transformation of bias parameters. Note that Ref. [128] use $b_3^{\rm MR} = 6\tilde{b}_{\delta^3}$ and $b_{\delta s^2}^{\rm MR} = 2\tilde{b}_{\delta K^2}$, and that $s^3 \equiv K^3$ can be re-expressed in terms of the other operators in $\tilde{\boldsymbol{O}}$ and hence is redundant. Vanishing matrix elements are indicated as empty cells for clarity.

Using that

$$
\begin{aligned}
\mathcal{G}_3 &\equiv -\frac{1}{2}\left[2(\partial_i\partial_j\hat{\Phi})^3 + \delta^3 - 3(\partial_i\partial_j\hat{\Phi})^2\delta\right] \\
&= -(K_{ij})^3 + \frac{1}{2}\delta(K_{ij})^2 - \frac{1}{9}\delta^3\,,
\end{aligned} \tag{C.18}
$$

as defined in [197], we then obtain the relation between the third-order operator basis adopted in [197] and the basis used here, which is summarized in Tab. C.14.

Finally, we can relate $O_{\rm td}^{(3)}$ to the operators defined in [128] through (note that s_{ij} there equals K_{ij} as defined here)

$$
\begin{aligned}
\eta &\equiv \nabla^2\Phi_v - \nabla^2\hat{\Phi}\,, \\
t_{ij} &\equiv \left(\partial_i\partial_j - \frac{1}{3}\delta_{ij}\nabla^2\right)\Phi_v - K_{ij} = \mathcal{D}_{ij}\eta\,.
\end{aligned} \tag{C.19}
$$

At second order in PT, this yields

$$
\eta^{(2)} = -\frac{4}{21}\left(\left[\delta^{(1)}\right]^2 - \frac{3}{2}\left[K_{ij}^{(1)}\right]^2\right)\,. \tag{C.20}
$$

We then have

$$
st \equiv s_{ij}t^{ij} = K_{ij}\left(\frac{\partial_i\partial_j}{\nabla^2}\right)\eta = -\frac{1}{2}O_{\rm td}^{(3)}\,. \tag{C.21}
$$

Using Eq. (B.46), the operator $\psi = \psi_{\rm MR}$ defined in [128] can eventually be expressed at third order as

$$
\begin{aligned}
\psi_{\rm MR}^{(3)} &\equiv \left[\eta - \frac{2}{7}(K_{ij})^2 + \frac{4}{21}\delta^2\right]^{(3)} \\
&= -\frac{8}{21}O_{\rm td}^{(3)} + \frac{328}{3969}\delta^3 - \frac{4}{49}\delta(K_{ij})^2 - \frac{8}{63}(K_{ij})^3\,.
\end{aligned} \tag{C.22}
$$

Tab. C.15 summarizes the relation between the third-order operator basis adopted in [128, 137] and the basis used here. Note that K^3 (denoted as s^3 in their notation) is linearly dependent on the other operators in their basis and hence not included here. Alternatively, one could replace either st or ψ_{MR} with the operator K^3.

Refs. [128, 137] introduce a bias parameter $b_{3\mathrm{nl}}$ which multiplies the term proportional to $f_{\mathrm{NLO}}(k)P_{\mathrm{L}}(k)$ in the NLO galaxy power spectrum, Eq. (4.22), where $f_{\mathrm{NLO}}(k)$ is defined in Eq. (4.23). In our notation, this parameter is given by

$$b_{3\mathrm{nl}} = -\frac{32}{21}\left(b_{K^2} + \frac{2}{5}b_{\mathrm{td}}\right), \tag{C.23}$$

where the prefactor comes from matching $f_{\mathrm{NLO}}(k)$ and the function $\sigma_3^2(k)$ defined in [128]. Inserting our results from Sec. 2.4 for Lagrangian LIMD bias [Eq. (2.34) and Eq. (2.53)], we obtain

$$b_{3\mathrm{nl}} = \frac{32}{315}b_1^L, \tag{C.24}$$

in agreement with [137]. Note, however, that $b_{3\mathrm{nl}}$ is different from the b_O discussed throughout the review, since it does not multiply a single bias operator. Instead, it is defined as a coefficient of a specific term in the NLO (1-loop) contribution to the galaxy power spectrum (indeed, we see from Eq. (C.23) that it involves both second- and third-order biases).

Appendix C.3 Derivation of $\Pi^{[2]}$

Using the definition of the convective time derivative [Eq. (2.18)], and the perturbation theory expansion of the density and velocity fields [Eq. (B.7)], $\Pi_{ij}^{[2]}$ can be formally written as a sum over perturbation-theory contributions [131]:

$$\Pi_{ij}^{[2]}(\boldsymbol{k},\tau) = \sum_{n=1}^{\infty} D^n(\tau)\Bigg[\frac{\boldsymbol{k}^i\boldsymbol{k}^j}{|\boldsymbol{k}|^2}(n-1)\delta^{(n)}(\boldsymbol{k})$$
$$+ \frac{1}{\mathcal{H}f}\sum_{m=1}^{n-1}\int_{\boldsymbol{k}_1}\int_{\boldsymbol{k}_2}(2\pi)^3\delta_D(\boldsymbol{k}-\boldsymbol{k}_{12})\frac{\boldsymbol{k}_1\cdot\boldsymbol{k}_2}{k_1}\frac{k_2^i k_2^j}{k_2^2}\theta^{(m)}(\boldsymbol{k}_1)\delta^{(n-m)}(\boldsymbol{k}_2)\Bigg]. \tag{C.25}$$

At second order, we can restrict to $n=2$ and $m=1$. This yields

$$\Pi_{ij}^{[2]}(\boldsymbol{k},\tau) = \int_{\boldsymbol{k}_1}\int_{\boldsymbol{k}_2}(2\pi)^3\delta_D(\boldsymbol{k}-\boldsymbol{k}_{12})\left[\frac{k_{12}^i k_{12}^j}{k_{12}^2}F_2(\boldsymbol{k}_1,\boldsymbol{k}_2) - \frac{1}{2}\frac{\boldsymbol{k}_1\cdot\boldsymbol{k}_2}{k_1^2 k_2^2}\left(k_1^i k_1^j + k_2^i k_2^j\right)\right]\delta(\boldsymbol{k}_1)\delta(\boldsymbol{k}_2)$$

$$= \int_{\boldsymbol{k}_1}\int_{\boldsymbol{k}_2}(2\pi)^3\delta_D(\boldsymbol{k}-\boldsymbol{k}_{12})\left[\frac{k_{12}^i k_{12}^j}{k_{12}^2}\left(\frac{17}{21} + \frac{2}{7}\left(\mu_{12}^2 - \frac{1}{3}\right) + \frac{1}{2}\boldsymbol{k}_1\cdot\boldsymbol{k}_2\left(\frac{1}{k_1^2} + \frac{1}{k_2^2}\right)\right)\right.$$
$$\left. - \frac{1}{2}\frac{\boldsymbol{k}_1\cdot\boldsymbol{k}_2}{k_1^2 k_2^2}\left(k_1^i k_1^j + k_2^i k_2^j\right)\right]\delta(\boldsymbol{k}_1)\delta(\boldsymbol{k}_2), \tag{C.26}$$

where $\mu_{12} = (\boldsymbol{k}_1\cdot\boldsymbol{k}_2)/k_1 k_2$ and we have used Eq. (B.12). The two displacement terms which are linear in $\boldsymbol{k}_1\cdot\boldsymbol{k}_2$ can be combined over a denominator $1/(k_1^2 k_2^2 |\boldsymbol{k}_{12}|^2)$ to give

$$\Pi_{ij}^{[2]}(\boldsymbol{k},\tau) = \int_{\boldsymbol{k}_1}\int_{\boldsymbol{k}_2}(2\pi)^3\delta_D(\boldsymbol{k}-\boldsymbol{k}_{12})\left[\frac{k_{12}^i k_{12}^j}{k_{12}^2}\left(\frac{17}{21} + \frac{2}{7}\left(\mu_{12}^2 - \frac{1}{3}\right) - \mu_{12}^2\right) + \frac{\boldsymbol{k}_1\cdot\boldsymbol{k}_2}{2k_1^2 k_2^2}\left(k_1^i k_1^j + k_2^i k_2^j\right)\right]\delta(\boldsymbol{k}_1)\delta(\boldsymbol{k}_2).$$

We can now read off the real-space expression corresponding to $\Pi^{[2]}$:

$$\Pi_{ij}^{[2]}(\boldsymbol{x},\tau) = \frac{10}{21}\frac{\partial_i\partial_j}{\nabla^2}\left[\delta^2 - \frac{3}{2}K^2\right] + \partial_i\partial_k\hat{\Phi}\partial^k\partial_j\hat{\Phi}. \tag{C.27}$$

This is Eq. (C.10). More generally, the following relations are useful to commute displacement terms with an inverse Laplacian:

$$
\begin{aligned}
\frac{1}{\nabla^2}\left[\partial_i A\frac{\partial^i}{\nabla^2}B + (A \leftrightarrow B)\right] &= \int_{\boldsymbol{k}_1}\int_{\boldsymbol{k}_2}(2\pi)^3\delta_D(\boldsymbol{k}-\boldsymbol{k}_{12})\left(-\frac{1}{k_{12}^2}\right)\left[\frac{1}{k_1^2}+\frac{1}{k_2^2}\right](\boldsymbol{k}_1\cdot\boldsymbol{k}_2)A(\boldsymbol{k}_1)B(\boldsymbol{k}_2) \\
&= \int_{\boldsymbol{k}_1}\int_{\boldsymbol{k}_2}(2\pi)^3\delta_D(\boldsymbol{k}-\boldsymbol{k}_{12})\left(-\frac{1}{k_{12}^2 k_1^2 k_2^2}\right)\left[k_{12}^2 - 2\boldsymbol{k}_1\cdot\boldsymbol{k}_2\right](\boldsymbol{k}_1\cdot\boldsymbol{k}_2)A(\boldsymbol{k}_1)B(\boldsymbol{k}_2) \\
&= \int_{\boldsymbol{k}_1}\int_{\boldsymbol{k}_2}(2\pi)^3\delta_D(\boldsymbol{k}-\boldsymbol{k}_{12})\left[-\frac{1}{k_1^2 k_2^2}(\boldsymbol{k}_1\cdot\boldsymbol{k}_2) + \frac{2}{k_{12}^2}\frac{(\boldsymbol{k}_1\cdot\boldsymbol{k}_2)^2}{k_1^2 k_2^2}\right]A(\boldsymbol{k}_1)B(\boldsymbol{k}_2) \\
&= \left(\frac{\partial_i}{\nabla^2}A\right)\left(\frac{\partial^i}{\nabla^2}B\right) - \frac{2}{\nabla^2}\left[\left(\frac{\partial_i\partial_j}{\nabla^2}A\right)\left(\frac{\partial^i\partial^j}{\nabla^2}B\right)\right].
\end{aligned} \tag{C.28}
$$

In particular, this yields

$$
\begin{aligned}
\frac{\partial^i\partial^j}{\nabla^2}\left[s^k\partial_k\delta\right] &= -\frac{1}{2}\partial^i\partial^j[s^2] + \frac{\partial^i\partial^j}{\nabla^2}\left[K^2 + \frac{1}{3}\delta^2\right] \\
\frac{\partial^i}{\nabla^2}\left[s^k\partial_k\delta\right] &= -\frac{1}{2}\partial^i\left[s^2\right] + \frac{\partial^i}{\nabla^2}\left[K^2 + \frac{1}{3}\delta^2\right] = -s^k(\partial_k s^i) + \frac{\partial^i}{\nabla^2}\left[K^2 + \frac{1}{3}\delta^2\right].
\end{aligned} \tag{C.29}
$$

These relations can also be used to directly derive $\Pi_{ij}^{[2]}$ from the real-space expression for $\delta^{(2)}$, Eq. (B.14).

Appendix D   Halo finding algorithms

Collections of bound particles, or dark matter halos, identified in N-body simulations are a central tool for testing the predictions of bias models. It would be desirable to process halo catalogs into weighted samples with statistical properties similar to those hosting the observed galaxies, quasars, clusters, or other tracers. However, this proves to be quite challenging owing to uncertainties in the relation between galaxies and dark matter halos. Moreover, there is freedom in defining even the most basic halo property, its mass.

Halo identification algorithms, or "halo finders", can be broadly divided into two categories: Friends-of-Friends (FoF) finders [23] and spherical overdensity (SO) finders [869]. The choice of halo finder and mass definition is somewhat arbitrary and is often made to suit specific purposes. For example, the mass assignment of SO halos is more closely connected to the predictions of the spherical collapse model, which many of the results for Lagrangian models of halos presented in Sec. 5–6 rely on, and to observable mass proxies such as the weak lensing signal or the temperature of thermal X-ray emitting gas.

Particle membership of SO halos is defined by being inside a spherical shell centered on the center-of-mass of particles which encloses a fixed interior density $\rho(< R) = \Delta_{\mathrm{SO}}\, \bar{\rho}_m(z)$. The halo mass is given by the sum of the member particle masses. Unbound particles inside this shell are sometimes pruned. A common choice for the interior density threshold is $\Delta_{\mathrm{SO}} = \Delta_{\mathrm{vir}}(z) \approx 200$, which ensures that the mean effective linear collapse threshold δ_{cr} is close to the spherical collapse prediction ($\delta_{\mathrm{cr}} = 1.686$). Larger values $\Delta_{\mathrm{SO}} = 500 - 2500$ are also adopted to obtain a closer match to observationally inferred halo masses. On the other hand, particle membership, and hence mass, of an FoF halo is determined by being separated by less than a linking length λ from another member particle, where, typically, $\lambda \sim 0.15 - 0.2$ in units of the mean interparticle distance. Such linking lengths can yield a mean effective critical threshold δ_{cr} which is lower than the spherical collapse prediction.

The question of how SO halo masses can be mapped onto friends-of-friends masses remains a matter of debate (see [870, 615] for a discussion). The halo mass definition is clearly of relevance to Lagrangian models of halos as well as universal mass functions phrased in terms of the peak significance $\nu_c \equiv \delta_{\mathrm{cr}}/\sigma(M)$: changing the mass assigned to halos also changes ν_c for each halo through $\sigma(M)$, and, thereby, changes the multiplicity $\nu_c f(\nu_c)$.

References

[1] M. Ryle and R. W. Clarke, Mon. Not. R. Astron. Soc.**122**, 349 (1961).

[2] P. J. E. Peebles, Astrophys. J.**185**, 413 (1973).

[3] M. G. Hauser and P. J. E. Peebles, Astrophys. J.**185**, 757 (1973).

[4] P. J. E. Peebles and M. G. Hauser, Astrophys. J. Supp.**28**, 19 (1974).

[5] P. J. E. Peebles, Astrophys. J.**196**, 647 (1975).

[6] E. J. Groth and P. J. E. Peebles, Astrophys. J.**217**, 385 (1977).

[7] M. Seldner and P. J. E. Peebles, Astrophys. J.**225**, 7 (1978).

[8] R. P. Kirshner, A. Oemler, Jr. and P. L. Schechter, Astron. J.**84**, 951 (1979).

[9] N. A. Bahcall and R. M. Soneira, Astrophys. J.**270**, 20 (1983).

[10] A. A. Klypin and A. I. Kopylov, Soviet Astronomy Letters **9**, 41 (1983).

[11] M. Davis and P. J. E. Peebles, Astrophys. J.**267**, 465 (1983).

[12] N. Kaiser, Astrophys. J. Lett.**284**, L9 (1984).

[13] J. M. Bardeen, J. R. Bond, N. Kaiser and A. S. Szalay, Astrophys. J.**304**, 15 (1986).

[14] S. Cole and N. Kaiser, Mon. Not. R. Astron. Soc.**237**, 1127 (1989).

[15] H. J. Mo and S. D. M. White, Mon. Not. R. Astron. Soc.**282**, 347 (1996), [arXiv:astro-ph/9512127].

[16] R. K. Sheth and G. Tormen, Mon. Not. R. Astron. Soc.**308**, 119 (1999), [arXiv:astro-ph/9901122].

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

[18] J. Tonry and M. Davis, Astron. J.**84**, 1511 (1979).

[19] S. J. Maddox, G. Efstathiou, W. J. Sutherland and J. Loveday, Mon. Not. R. Astron. Soc.**242**, 43P (1990).

[20] M. Davis, J. Huchra, D. W. Latham and J. Tonry, Astrophys. J.**253**, 423 (1982).

[21] D. J. Baumgart and J. N. Fry, Astrophys. J.**375**, 25 (1991).

[22] C. Park, J. R. Gott, III and L. N. da Costa, Astrophys. J. Lett.**392**, L51 (1992).

[23] M. Davis, G. Efstathiou, C. S. Frenk and S. D. M. White, Astrophys. J.**292**, 371 (1985).

[24] M. Tegmark *et al.*, Astrophys. J.**606**, 702 (2004), [arXiv:astro-ph/0310725].

[25] S. Cole *et al.*, Mon. Not. R. Astron. Soc.**362**, 505 (2005), [arXiv:astro-ph/0501174].

[26] D. Parkinson *et al.*, Phys. Rev. D**86**, 103518 (2012), [arXiv:1210.2130].

[27] C. L. Bennett *et al.*, Astrophys. J. Supp.**208**, 20 (2013), [arXiv:1212.5225].

[28] S. de la Torre *et al.*, Astron. Astrophys.**557**, A54 (2013), [arXiv:1303.2622].

[29] Planck Collaboration, Astron. Astrophys.**594**, A1 (2016), [arXiv:1502.01582].

[30] S. Alam *et al.*, Mon. Not. R. Astron. Soc.**470**, 2617 (2017), [arXiv:1607.03155].

[31] DES, T. M. C. Abbott *et al.*, Phys. Rev. **D98**, 043526 (2018), [arXiv:1708.01530].

[32] V. Springel, C. S. Frenk and S. D. M. White, Nature (London)**440**, 1137 (2006), [arXiv:astro-ph/0604561].

[33] A. Einstein and W. de Sitter, Proceedings of the National Academy of Science **18**, 213 (1932).

[34] G. Efstathiou, J. R. Bond and S. D. M. White, Mon. Not. R. Astron. Soc.**258**, 1P (1992).

[35] C. L. Bennett *et al.*, Astrophys. J. Lett.**464**, L1 (1996), [arXiv:astro-ph/9601067].

[36] M. Davis, G. Efstathiou, C. S. Frenk and S. D. M. White, Nature (London)**356**, 489 (1992).

[37] G. Efstathiou, W. J. Sutherland and S. J. Maddox, Nature (London)**348**, 705 (1990).

[38] P. J. E. Peebles, Astrophys. J.**284**, 439 (1984).

[39] DES, J. Elvin-Poole *et al.*, Phys. Rev. **D98**, 042006 (2018), [arXiv:1708.01536].

[40] W. Percival, Cosmological Constraints from Galaxy Clustering, in *The Invisible Universe: Dark Matter and Dark Energy*, edited by L. Papantonopoulos, , Lecture Notes in Physics, Berlin Springer Verlag Vol. 720, p. 157, 2007, [arXiv:astro-ph/0601538].

[41] D. J. Eisenstein and W. Hu, Astrophys. J.**496**, 605 (1998), [arXiv:astro-ph/9709112].

[42] D. J. Eisenstein *et al.*, Astrophys. J.**633**, 560 (2005), [arXiv:astro-ph/0501171].

[43] H.-J. Seo and D. J. Eisenstein, Astrophys. J.**598**, 720 (2003), [arXiv:astro-ph/0307460].

[44] H.-J. Seo and D. J. Eisenstein, Astrophys. J.**665**, 14 (2007), [arXiv:astro-ph/0701079].

[45] J. C. Jackson, Mon. Not. R. Astron. Soc.**156**, 1P (1972), [arXiv:0810.3908].

[46] W. L. W. Sargent and E. L. Turner, Astrophys. J. Lett.**212**, L3 (1977).

[47] N. Kaiser, Monthly Notices of the Royal Astronomical Society (ISSN 0035-8711) **227**, 1 (1987).

[48] Dark Energy Survey Collaboration, Mon. Not. R. Astron. Soc.**460**, 1270 (2016), [arXiv:1601.00329].

[49] G.-B. Zhao *et al.*, Mon. Not. R. Astron. Soc.**457**, 2377 (2016), [arXiv:1510.08216].

[50] G. J. Hill *et al.*, The Hobby-Eberly Telescope Dark Energy Experiment (HETDEX): Description and Early Pilot Survey Results, in *Panoramic Views of Galaxy Formation and Evolution*, edited by T. Kodama, T. Yamada and K. Aoki, , Astronomical Society of the Pacific Conference Series Vol. 399, p. 115, 2008, [arXiv:0806.0183].

[51] S. Miyazaki *et al.*, Astrophys. J.**807**, 22 (2015), [arXiv:1504.06974].

[52] J. T. A. de Jong, G. A. Verdoes Kleijn, K. H. Kuijken and E. A. Valentijn, Experimental Astronomy **35**, 25 (2013), [arXiv:1206.1254].

[53] M. Levi *et al.*, ArXiv e-prints (2013), [arXiv:1308.0847].

[54] L. Amendola *et al.*, Living Reviews in Relativity **16** (2013), [arXiv:1206.1225].

[55] R. S. de Jong *et al.*, 4MOST: 4-metre multi-object spectroscopic telescope, in *Ground-based and Airborne Instrumentation for Astronomy IV*, , Proceedings of the SPIE Vol. 8446, p. 84460T, 2012, [arXiv:1206.6885].

[56] LSST Science Collaboration, ArXiv e-prints (2009), [arXiv:0912.0201].

[57] M. Takada *et al.*, PASJ**66**, R1 (2014), [arXiv:1206.0737].

[58] D. Spergel *et al.*, ArXiv e-prints (2015), [arXiv:1503.03757].
[59] A. Cooray and W. Hu, Astrophys. J.**548**, 7 (2001), [arXiv:astro-ph/0004151].
[60] E. Sefusatti and R. Scoccimarro, Phys. Rev. D**71**, 063001 (2005), [arXiv:astro-ph/0412626].
[61] E. Sefusatti, M. Crocce, S. Pueblas and R. Scoccimarro, Phys. Rev. D**74**, 023522 (2006), [arXiv:astro-ph/0604505].
[62] E. Sefusatti and E. Komatsu, Phys. Rev. D**76**, 083004 (2007), [arXiv:0705.0343].
[63] P. Zhang, M. Liguori, R. Bean and S. Dodelson, Physical Review Letters **99**, 141302 (2007), [arXiv:0704.1932].
[64] R. Reyes *et al.*, Nature **464**, 256 (2010), [arXiv:1003.2185].
[65] C. Blake *et al.*, Mon. Not. R. Astron. Soc.**456**, 2806 (2016), [arXiv:1507.03086].
[66] A. R. Pullen, S. Alam, S. He and S. Ho, Mon. Not. R. Astron. Soc.**460**, 4098 (2016), [arXiv:1511.04457].
[67] R. Scoccimarro, E. Sefusatti and M. Zaldarriaga, Phys. Rev. D**69**, 103513 (2004), [arXiv:astro-ph/0312286].
[68] N. Dalal, O. Doré, D. Huterer and A. Shirokov, Phys. Rev. D**77**, 123514 (2008), [arXiv:0710.4560].
[69] J. Einasto, *Large-scale structure of the universe* (, 1992), p. 249.
[70] P. Coles, Large-scale Structure, Theory and Statistics, in *Phase Transitions in the Early Universe: Theory and Observations*, edited by H. J. de Vega, I. M. Khalatnikov and N. G. Sanchez, p. 217, 2001, [arXiv:astro-ph/0103017].
[71] P. J. E. Peebles, *The large-scale structure of the universe* (Princeton University Press, 1980).
[72] F. Bernardeau, S. Colombi, E. Gaztanaga and R. Scoccimarro, Phys.Rept. **367**, 1 (2002), [arXiv:astro-ph/0112551].
[73] M. Crocce and R. Scoccimarro, Phys. Rev. D**73**, 063519 (2006), [arXiv:astro-ph/0509418].
[74] F. Bernardeau, M. Crocce and R. Scoccimarro, Phys. Rev. D**78**, 103521 (2008), [arXiv:0806.2334].
[75] F. Montesano, A. G. Sánchez and S. Phleps, Monthly Notices of the Royal Astronomical Society **408**, 2397 (2010).
[76] P. McDonald, Phys. Rev. D**75**, 043514 (2007), [arXiv:astro-ph/0606028].
[77] A. Taruya and T. Hiramatsu, The Astrophysical Journal **674**, 617 (2008).
[78] T. Buchert, Mon. Not. R. Astron. Soc.**254**, 729 (1992).
[79] T. Matsubara, Phys. Rev. D**78**, 083519 (2008), [arXiv:0807.1733].
[80] V. Zheligovsky and U. Frisch, Journal of Fluid Mechanics **749**, 404 (2014), [arXiv:1312.6320].
[81] S. Matarrese and M. Pietroni, Journal of Cosmology and Astroparticle Physics **06**, 026 (2007).
[82] M. Pietroni, Journal of Cosmology and Astroparticle Physics **10**, 036 (2008).
[83] D. Baumann, A. Nicolis, L. Senatore and M. Zaldarriaga, JCAP **7**, 051 (2012), [arXiv:1004.2488].
[84] J. J. M. Carrasco, M. P. Hertzberg and L. Senatore, Journal of High Energy Physics **9**, 82 (2012), [arXiv:1206.2926].
[85] J. J. M. Carrasco, S. Foreman, D. Green and L. Senatore, Journal of Cosmology and Astroparticle Physics **07**, 057 (2014).
[86] M. P. Hertzberg, Physical Review D **89**, 043521 (2014).
[87] R. A. Porto, L. Senatore and M. Zaldarriaga, Journal of Cosmology and Astroparticle Physics **05**, 022 (2014).
[88] L. Senatore, JCAP **11**, 007 (2015), [arXiv:1406.7843].
[89] L. Senatore and M. Zaldarriaga, arXiv.org , 1225 (2014), [arXiv:1409.1225].
[90] T. Baldauf, L. Mercolli and M. Zaldarriaga, Phys. Rev. D**92**, 123007 (2015), [arXiv:1507.02256].
[91] D. Blas, M. Garny, M. M. Ivanov and S. Sibiryakov, JCAP **7**, 052 (2016), [arXiv:1512.05807].
[92] H. Mo, F. C. van den Bosch and S. White, *Galaxy Formation and Evolution* (Cambridge University Press, 2010).
[93] J. Guzik and U. Seljak, Mon. Not. R. Astron. Soc.**335**, 311 (2002), [arXiv:astro-ph/0201448].
[94] T. G. Brainerd, R. D. Blandford and I. Smail, Astrophys. J.**466**, 623 (1996), [arXiv:astro-ph/9503073].
[95] T. Baldauf, U. Seljak, L. Senatore and M. Zaldarriaga, JCAP **10**, 031 (2011), [arXiv:1106.5507].
[96] D. Jeong, F. Schmidt and C. M. Hirata, Phys. Rev. D**85**, 023504 (2012), [arXiv:1107.5427].
[97] Planck Collaboration, Astron. Astrophys.**594**, A13 (2016), [arXiv:1502.01589].
[98] V. Desjacques, D. Jeong and F. Schmidt, Phys. Rev. D**84**, 063512 (2011), [arXiv:1105.3628].
[99] A. S. Szalay, Astrophys. J.**333**, 21 (1988).
[100] F. Schmidt, D. Jeong and V. Desjacques, Phys. Rev. D**88**, 023515 (2013), [arXiv:1212.0868].
[101] D. Jeong and E. Komatsu, Astrophys.J. **651**, 619 (2006), [arXiv:astro-ph/0604075].
[102] F. Bernardeau, Astron. Astrophys.**312**, 11 (1996), [arXiv:astro-ph/9602072].
[103] J. N. Fry, Astrophys. J. Lett.**308**, L71 (1986).
[104] P. Coles, Mon. Not. R. Astron. Soc.**262**, 1065 (1993).
[105] R. J. Scherrer and D. H. Weinberg, Astrophys. J.**504**, 607 (1998), [arXiv:astro-ph/9712192].
[106] P. Coles, A. L. Melott and D. Munshi, Astrophys. J. Lett.**521**, L5 (1999), [arXiv:astro-ph/9904253].
[107] V. K. Narayanan, A. A. Berlind and D. H. Weinberg, Astrophys. J.**528**, 1 (2000), [arXiv:astro-ph/9812002].
[108] J. E. Gunn and J. R. Gott, III, Astrophys. J.**176**, 1 (1972).
[109] F. Bernardeau, Astrophys. J.**392**, 1 (1992).
[110] H. J. Mo, Y. P. Jing and S. D. M. White, Mon. Not. R. Astron. Soc.**284**, 189 (1997), [arXiv:astro-ph/9603039].
[111] P. Fosalba and E. Gaztanaga, Mon. Not. R. Astron. Soc.**301**, 503 (1998), [arXiv:astro-ph/9712095].
[112] C. Wagner, F. Schmidt, C.-T. Chiang and E. Komatsu, JCAP **8**, 042 (2015), [arXiv:1503.03487].
[113] T. Matsubara, Phys. Rev. D**83**, 083518 (2011), [arXiv:1102.4619].
[114] A. Nusser and M. Davis, Astrophys. J. Lett.**421**, L1 (1994), [arXiv:astro-ph/9309009].
[115] J. N. Fry, Astrophys. J. Lett.**461**, L65 (1996).
[116] M. Tegmark and P. J. E. Peebles, Astrophys. J. Lett.**500**, L79 (1998), [arXiv:astro-ph/9804067].
[117] P. Catelan, F. Lucchin, S. Matarrese and C. Porciani, Mon. Not. R. Astron. Soc.**297**, 692 (1998), [arXiv:astro-ph/9708067].
[118] K. C. Chan, R. Scoccimarro and R. K. Sheth, Phys. Rev. D**85**, 083509 (2012), [arXiv:1201.3614].
[119] Dmitriev, N.A. and Zel'dovich, Ya. B., JETP **18**, 793 (1964).

[120] J. Carlson, B. Reid and M. White, Mon. Not. R. Astron. Soc.**429**, 1674 (2013), [arXiv:1209.0780].
[121] R. K. Sheth, K. C. Chan and R. Scoccimarro, Phys. Rev. D**87**, 083002 (2013), [arXiv:1207.7117].
[122] A. Taruya, K. Koyama and J. Soda, Astrophys. J.**510**, 541 (1999), [arXiv:astro-ph/9807005].
[123] S. Matarrese, P. Coles, F. Lucchin and L. Moscardini, Mon. Not. R. Astron. Soc.**286**, 115 (1997), [arXiv:astro-ph/9608004].
[124] P. Catelan, C. Porciani and M. Kamionkowski, Mon. Not. R. Astron. Soc.**318**, L39 (2000), [arXiv:astro-ph/0005544].
[125] X. Wang and A. Szalay, Phys. Rev. D**86**, 043508 (2012), [arXiv:1204.0019].
[126] A. F. Heavens, S. Matarrese and L. Verde, Mon. Not. R. Astron. Soc.**301**, 797 (1998), [arXiv:astro-ph/9808016].
[127] R. E. Smith, R. Scoccimarro and R. K. Sheth, Phys. Rev. D**75**, 063512 (2007), [arXiv:astro-ph/0609547].
[128] P. McDonald and A. Roy, JCAP **8**, 20 (2009), [arXiv:0902.0991].
[129] T. Baldauf, U. Seljak, V. Desjacques and P. McDonald, Phys. Rev. D**86**, 083540 (2012), [arXiv:1201.4827].
[130] A. Elia, S. Kulkarni, C. Porciani, M. Pietroni and S. Matarrese, Mon. Not. R. Astron. Soc.**416**, 1703 (2011), [arXiv:1012.4833].
[131] M. Mirbabayi, F. Schmidt and M. Zaldarriaga, JCAP **7**, 030 (2015), [arXiv:1412.5169].
[132] G. Somogyi and R. E. Smith, Phys. Rev. D**81**, 023524 (2010), [arXiv:0910.5220].
[133] S. Basilakos and M. Plionis, Astrophys. J.**550**, 522 (2001), [arXiv:astro-ph/0011265].
[134] T. Buchert, Mon. Not. R. Astron. Soc.**267**, 811 (1994), [arXiv:astro-ph/9309055].
[135] F. R. Bouchet, S. Colombi, E. Hivon and R. Juszkiewicz, Astron. Astrophys.**296**, 575 (1995), [arXiv:astro-ph/9406013].
[136] Y. Ohta, I. Kayo and A. Taruya, Astrophys. J.**608**, 647 (2004), [arXiv:astro-ph/0402618].
[137] S. Saito *et al.*, Phys. Rev. **D90**, 123522 (2014), [arXiv:1405.1447].
[138] M. Biagetti, V. Desjacques, A. Kehagias and A. Riotto, Phys. Rev. **D90**, 045022 (2014), [arXiv:1405.1435].
[139] E. V. Linder, Phys. Rev. **D72**, 043529 (2005), [arXiv:astro-ph/0507263].
[140] A. Barreira and F. Schmidt, JCAP **6**, 053 (2017), [arXiv:1703.09212].
[141] T. Matsubara, Phys. Rev. D**90**, 043537 (2014), [arXiv:1304.4226].
[142] M. P. van Daalen, J. Schaye, I. G. McCarthy, C. M. Booth and C. Dalla Vecchia, Mon. Not. R. Astron. Soc.**440**, 2997 (2014), [arXiv:1310.7571].
[143] W. D. Goldberger and I. Z. Rothstein, Phys. Rev. D**73**, 104029 (2006), [arXiv:hep-th/0409156].
[144] R. A. Porto, Phys. Rep.**633**, 1 (2016), [arXiv:1601.04914].
[145] E. Pajer, F. Schmidt and M. Zaldarriaga, Phys. Rev. D**88**, 083502 (2013), [arXiv:1305.0824].
[146] L. Dai, E. Pajer and F. Schmidt, JCAP **11**, 043 (2015), [arXiv:1502.02011].
[147] P. Simon, Astron. Astrophys.**430**, 827 (2005), [arXiv:astro-ph/0409435].
[148] L. Kofman and D. Pogosyan, Astrophys. J.**442**, 30 (1995), [arXiv:astro-ph/9403029].
[149] E. Bertschinger, Cosmological dynamics: Course 1, in *Summer School on Cosmology and Large Scale Structure (Session 60) Les Houches, France, August 1-28, 1993*, 1993, [arXiv:astro-ph/9503125].
[150] H. Y. Ip and F. Schmidt, JCAP **2**, 025 (2017), [arXiv:1610.01059].
[151] T. Matsubara, Phys. Rev. D**92**, 023534 (2015), [arXiv:1505.01481].
[152] V. Desjacques, M. Crocce, R. Scoccimarro and R. K. Sheth, Phys. Rev. D**82**, 103529 (2010), [arXiv:1009.3449].
[153] M. Musso and R. K. Sheth, Mon. Not. R. Astron. Soc.**423**, L102 (2012), [arXiv:1201.3876].
[154] T. Matsubara, Astrophys. J.**525**, 543 (1999), [arXiv:astro-ph/9906029].
[155] P. Coles and P. Erdogdu, JCAP **10**, 007 (2007), [arXiv:0706.0412].
[156] V. Desjacques, Phys. Rev. D**78**, 103503 (2008), [arXiv:0806.0007].
[157] A. Babul and S. D. M. White, Mon. Not. R. Astron. Soc.**253**, 31P (1991).
[158] R. G. Bower, P. Coles, C. S. Frenk and S. D. M. White, Astrophys. J.**405**, 403 (1993).
[159] F. Schmidt and F. Beutler, Phys. Rev. D**96**, 083533 (2017), [arXiv:1705.07843].
[160] A. E. Broderick, P. Chang and C. Pfrommer, Astrophys. J.**752**, 22 (2012), [arXiv:1106.5494].
[161] A. Lamberts *et al.*, Astrophys. J.**811**, 19 (2015), [arXiv:1502.07980].
[162] J. R. Pritchard, S. R. Furlanetto and M. Kamionkowski, Mon. Not. R. Astron. Soc.**374**, 159 (2007), [arXiv:astro-ph/0604358].
[163] J. Schaye, Astrophys. J.**559**, 507 (2001), [arXiv:astro-ph/0104272].
[164] A. A. Meiksin, Reviews of Modern Physics **81**, 1405 (2009), [arXiv:0711.3358].
[165] A. Meiksin and M. White, Mon. Not. R. Astron. Soc.**350**, 1107 (2004), [arXiv:astro-ph/0307289].
[166] R. A. C. Croft, Astrophys. J.**610**, 642 (2004), [arXiv:astro-ph/0310890].
[167] A. Pontzen, Phys. Rev. D**89**, 083010 (2014), [arXiv:1402.0506].
[168] A. Pontzen, S. Bird, H. Peiris and L. Verde, Astrophys. J. Lett.**792**, L34 (2014), [arXiv:1407.6367].
[169] R. G. Carlberg, Astrophys. J.**433**, 468 (1994), [arXiv:astro-ph/9404005].
[170] P. Colín, A. A. Klypin and A. V. Kravtsov, Astrophys. J.**539**, 561 (2000), [arXiv:astro-ph/9907337].
[171] E. Jennings, C. M. Baugh and D. Hatt, Mon. Not. R. Astron. Soc.**446**, 793 (2015), [arXiv:1407.7296].
[172] S. Pueblas and R. Scoccimarro, Phys. Rev. D**80**, 043504 (2009), [arXiv:0809.4606].
[173] Y. Zheng, P. Zhang and Y. Jing, Phys. Rev. D**91**, 123512 (2015), [arXiv:1410.1256].
[174] T. Baldauf, V. Desjacques and U. Seljak, Phys. Rev. **D92**, 123507 (2015), [arXiv:1405.5885].
[175] A. Dekel and O. Lahav, Astrophys. J.**520**, 24 (1999), [arXiv:astro-ph/9806193].
[176] A. Taruya and J. Soda, Astrophys. J.**522**, 46 (1999), [arXiv:astro-ph/9809204].
[177] N. Hamaus, U. Seljak, V. Desjacques, R. E. Smith and T. Baldauf, Phys. Rev. D**82**, 043515 (2010), [arXiv:1004.5377].
[178] T. Baldauf, U. Seljak, R. E. Smith, N. Hamaus and V. Desjacques, Phys. Rev. D**88**, 083507 (2013), [arXiv:1305.2917].
[179] K. C. Chan, N. Hamaus and V. Desjacques, Phys. Rev. D**90**, 103521 (2014), [arXiv:1409.3849].

[180] A. Perko, L. Senatore, E. Jennings and R. H. Wechsler, ArXiv e-prints (2016), [arXiv:1610.09321].
[181] F. Schmidt, Phys. Rev. D**81**, 103002 (2010), [arXiv:1003.0409].
[182] D. Wands and A. Slosar, Phys. Rev. D**79**, 123507 (2009), [arXiv:0902.1084].
[183] J. Yoo, Phys. Rev. D**90**, 123507 (2014), [arXiv:1408.5137].
[184] D. Bertacca *et al.*, Classical and Quantum Gravity **32**, 175019 (2015), [arXiv:1501.03163].
[185] L. Dai, E. Pajer and F. Schmidt, JCAP **10**, 059 (2015), [arXiv:1504.00351].
[186] G. Lemaître, Annales de la Societe Scietifique de Bruxelles **53**, 51 (1933).
[187] J. D. Barrow and P. Saich, Mon. Not. R. Astron. Soc.**262**, 717 (1993).
[188] S. Cole, Mon. Not. R. Astron. Soc.**286**, 38 (1997), [arXiv:astro-ph/9604046].
[189] B. D. Sherwin and M. Zaldarriaga, Phys. Rev. D**85**, 103523 (2012), [arXiv:1202.3998].
[190] A. Challinor and A. Lewis, Phys. Rev. D**84**, 043516 (2011), [arXiv:1105.5292].
[191] S. Matarrese, S. Mollerach and M. Bruni, Phys. Rev. **D58**, 043504 (1998), [arXiv:astro-ph/9707278].
[192] D. Baumann, P. Steinhardt, K. Takahashi and K. Ichiki, Phys. Rev. D**76**, 084019 (2007), [arXiv:hep-th/0703290].
[193] L. Boubekeur, P. Creminelli, J. Norena and F. Vernizzi, JCAP **0808**, 028 (2008), [arXiv:0806.1016].
[194] F. Schmidt and D. Jeong, Phys. Rev. D**86**, 083513 (2012), [arXiv:1205.1514].
[195] F. Schmidt, E. Pajer and M. Zaldarriaga, Phys. Rev. D**89**, 083507 (2014), [arXiv:1312.5616].
[196] T. Lazeyras, M. Musso and V. Desjacques, Phys. Rev. D**93**, 063007 (2016), [arXiv:1512.05283].
[197] V. Assassi, D. Baumann, D. Green and M. Zaldarriaga, JCAP **8**, 56 (2014), [arXiv:1402.5916].
[198] G. Rosen, Am. J. Phys. **40**, 683 (1972).
[199] S. Weinberg, Phys. Rev. D**67**, 123504 (2003), [arXiv:astro-ph/0302326].
[200] P. Creminelli, J. Noreña, M. Simonović and F. Vernizzi, JCAP **12**, 025 (2013), [arXiv:1309.3557].
[201] A. Kehagias and A. Riotto, Nuclear Physics B **873**, 514 (2013), [arXiv:1302.0130].
[202] W. D. Goldberger, L. Hui and A. Nicolis, Phys. Rev. D**87**, 103520 (2013), [arXiv:1303.1193].
[203] M. Peloso and M. Pietroni, JCAP **5**, 031 (2013), [arXiv:1302.0223].
[204] M. Peloso and M. Pietroni, JCAP **4**, 011 (2014), [arXiv:1310.7915].
[205] P. Valageas, Phys. Rev. D**89**, 083534 (2014), [arXiv:1311.1236].
[206] A. Kehagias, J. Noreña, H. Perrier and A. Riotto, Nuclear Physics B **883**, 83 (2014), [arXiv:1311.0786].
[207] B. Horn, L. Hui and X. Xiao, JCAP **9**, 044 (2014), [arXiv:1406.0842].
[208] P. McDonald, Astrophys. J.**585**, 34 (2003), [arXiv:astro-ph/0108064].
[209] D. M. Goldberg and M. S. Vogeley, Astrophys. J.**605**, 1 (2004), [arXiv:astro-ph/0307191].
[210] E. Sirko, Astrophys.J. **634**, 728 (2005), [arXiv:astro-ph/0503106].
[211] M. C. Martino and R. K. Sheth, Mon. Not. R. Astron. Soc.**394**, 2109 (2009), [arXiv:0901.0757].
[212] N. Y. Gnedin, A. V. Kravtsov and D. H. Rudd, Astrophys. J. Supp.**194**, 46 (2011), [arXiv:1104.1428].
[213] Y. Li, W. Hu and M. Takada, Phys. Rev. D**89**, 083519 (2014), [arXiv:1401.0385].
[214] C. Wagner, F. Schmidt, C.-T. Chiang and E. Komatsu, Mon.Not.Roy.Astron.Soc. **448**, 11 (2015), [arXiv:1409.6294].
[215] F. Schmidt, M. V. Lima, H. Oyaizu and W. Hu, Phys. Rev. **D79**, 083518 (2009), [arXiv:0812.0545].
[216] W. Hu, C.-T. Chiang, Y. Li and M. LoVerde, Phys. Rev. D**94**, 023002 (2016), [arXiv:1605.01412].
[217] W. H. Press and P. Schechter, Astrophys. J.**187**, 425 (1974).
[218] E. Pajer and M. Zaldarriaga, JCAP **8**, 037 (2013), [arXiv:1301.7182].
[219] R. E. Angulo and S. D. M. White, Mon. Not. R. Astron. Soc.**405**, 143 (2010), [arXiv:0912.4277].
[220] V. R. Eke, S. Cole and C. S. Frenk, Mon. Not. R. Astron. Soc.**282** (1996), [arXiv:astro-ph/9601088].
[221] V. Desjacques, Phys. Rev. D**87**, 043505 (2013), [arXiv:1211.4128].
[222] R. K. Sheth and G. Lemson, Mon. Not. R. Astron. Soc.**304**, 767 (1999), [arXiv:astro-ph/9808138].
[223] R. K. Sheth, H. J. Mo and G. Tormen, Mon. Not. R. Astron. Soc.**323**, 1 (2001), [arXiv:astro-ph/9907024].
[224] R. K. Sheth and G. Tormen, Mon. Not. R. Astron. Soc.**329**, 61 (2002), [arXiv:astro-ph/0105113].
[225] J. Lee and S. F. Shandarin, Astrophys. J.**500**, 14 (1998), [arXiv:astro-ph/9709200].
[226] M. Manera, R. K. Sheth and R. Scoccimarro, Mon. Not. R. Astron. Soc.**402**, 589 (2010), [arXiv:0906.1314].
[227] K. Hoffmann, J. Bel and E. Gaztañaga, Mon. Not. R. Astron. Soc.**450**, 1674 (2015), [arXiv:1503.00313].
[228] F. Bernardeau, S. Colombi, E. Gaztañaga and R. Scoccimarro, Phys. Rep.**367**, 1 (2002), [arXiv:astro-ph/0112551].
[229] J. A. Tyson, F. Valdes, J. F. Jarvis and A. P. Mills, Jr., Astrophys. J. Lett.**281**, L59 (1984).
[230] Y. P. Jing, Y. Suto and H. J. Mo, Astrophys. J.**657**, 664 (2007), [arXiv:astro-ph/0610099].
[231] L. Gao and S. D. M. White, Mon. Not. R. Astron. Soc.**377**, L5 (2007), [arXiv:astro-ph/0611921].
[232] P. Schneider, Astrophys. J.**498**, 43 (1998), [arXiv:astro-ph/9708269].
[233] L. van Waerbeke, Astron. Astrophys.**334**, 1 (1998), [arXiv:astro-ph/9710244].
[234] E. S. Sheldon *et al.*, Astron. J.**127**, 2544 (2004), [arXiv:astro-ph/0312036].
[235] R. Mandelbaum *et al.*, Mon. Not. R. Astron. Soc.**432**, 1544 (2013), [arXiv:1207.1120].
[236] DES, J. Prat *et al.*, Phys. Rev. **D98**, 042005 (2018), [arXiv:1708.01537].
[237] D. Munshi, P. Valageas, L. van Waerbeke and A. Heavens, Phys. Rep.**462**, 67 (2008), [arXiv:astro-ph/0612667].
[238] D. N. Limber, Astrophys. J.**119**, 655 (1954).
[239] DES, E. Krause *et al.*, Submitted to: Phys. Rev. D (2017), [arXiv:1706.09359].
[240] J. N. Fry, Physical Review Letters **73**, 215 (1994).
[241] J. A. Frieman and E. Gaztanaga, Astrophys. J.**425**, 392 (1994), [arXiv:astro-ph/9306018].
[242] S. Matarrese, L. Verde and A. F. Heavens, Mon. Not. R. Astron. Soc.**290**, 651 (1997), [arXiv:astro-ph/9706059].
[243] D. Jeong and E. Komatsu, Astrophys. J.**703**, 1230 (2009), [arXiv:0904.0497].
[244] M. Schmittfull, T. Baldauf and U. Seljak, Phys. Rev. D**91**, 043530 (2015), [arXiv:1411.6595].

[245] C. Modi, E. Castorina and U. Seljak, Mon. Not. R. Astron. Soc. **472**, 3959 (2017), [arXiv:1612.01621].
[246] W. Hu, Phys. Rev. D**64**, 083005 (2001), [arXiv:astro-ph/0105117].
[247] A. Cooray, Phys. Rev. D**64**, 043516 (2001), [arXiv:astro-ph/0105415].
[248] R. Scoccimarro *et al.*, Astrophys. J.**496**, 586 (1998), [arXiv:astro-ph/9704075].
[249] E. Sefusatti, *Probing fundamental physics with large-scale structure: From galaxy formation to inflation*, PhD thesis, New York University, New York, USA, 2005.
[250] D. Jeong, *Cosmology with high (z>1) redshift galaxy surveys*, PhD thesis, University of Texas at Austin, 2010.
[251] E. Gaztañaga and F. Bernardeau, Astronomy and Astrophysics **331**, 829 (1998).
[252] J. Barriga and E. Gaztañaga, Mon. Not. R. Astron. Soc.**333**, 443 (2002), [arXiv:astro-ph/0112278].
[253] M. Takada and B. Jain, Monthly Notice of the Royal Astronomical Society **340**, 580 (2003).
[254] J. Bel, K. Hoffmann and E. Gaztañaga, Mon. Not. R. Astron. Soc.**453**, 259 (2015), [arXiv:1504.02074].
[255] A. J. Nishizawa, M. Takada and T. Nishimichi, Mon. Not. R. Astron. Soc.**433**, 209 (2013), [arXiv:1212.4025].
[256] J. E. Pollack, R. E. Smith and C. Porciani, Mon. Not. R. Astron. Soc.**440**, 555 (2014), [arXiv:1309.0504].
[257] V. Desjacques and R. K. Sheth, Phys. Rev. D**81**, 023526 (2010), [arXiv:0909.4544].
[258] N. Frusciante and R. K. Sheth, JCAP **11**, 16 (2012), [arXiv:1208.0229].
[259] A. Elia, A. D. Ludlow and C. Porciani, Mon. Not. R. Astron. Soc.**421**, 3472 (2012), [arXiv:1111.4211].
[260] M. Tegmark, Physical Review Letters **79**, 3806 (1997), [arXiv:astro-ph/9706198].
[261] H. A. Feldman, N. Kaiser and J. A. Peacock, Astrophys. J.**426**, 23 (1994), [arXiv:astro-ph/9304022].
[262] E. T. Vishniac, Monthly Notices of the Royal Astronomical Society **203**, 345 (1983).
[263] T. Baldauf, M. Mirbabayi, M. Simonović and M. Zaldarriaga, ArXiv e-prints (2016), [arXiv:1602.00674].
[264] T. Fujita, V. Mauerhofer, L. Senatore, Z. Vlah and R. Angulo, ArXiv e-prints (2016), [arXiv:1609.00717].
[265] E. O. Nadler, A. Perko and L. Senatore, Journal of Cosmology and Astro-Particle Physics **2018**, 058 (2018), [arXiv:1710.10308].
[266] P. McDonald, Phys. Rev. D**74**, 103512 (2006), [arXiv:astro-ph/0609413].
[267] R. Angulo, M. Fasiello, L. Senatore and Z. Vlah, JCAP **9**, 029 (2015), [arXiv:1503.08826].
[268] I. Mohammed, U. Seljak and Z. Vlah, Mon. Not. R. Astron. Soc.**466**, 780 (2017), [arXiv:1607.00043].
[269] M. Crocce and R. Scoccimarro, Phys. Rev. D**77**, 023533 (2008), [arXiv:0704.2783].
[270] L. Senatore and M. Zaldarriaga, JCAP **2**, 013 (2015), [arXiv:1404.5954].
[271] T. Baldauf, M. Mirbabayi, M. Simonović and M. Zaldarriaga, Phys. Rev. D**92**, 043514 (2015), [arXiv:1504.04366].
[272] D. Blas, M. Garny, M. M. Ivanov and S. Sibiryakov, JCAP **7**, 028 (2016), [arXiv:1605.02149].
[273] D. Jeong and E. Komatsu, Astrophys. J.**691**, 569 (2009), [arXiv:0805.2632].
[274] A. Taruya, T. Nishimichi, S. Saito and T. Hiramatsu, Phys. Rev. D**80**, 123503 (2009), [arXiv:0906.0507].
[275] S. M. Fall, M. J. Geller, B. J. T. Jones and S. D. M. White, Astrophys. J. Lett.**205**, L121 (1976).
[276] J. N. Fry and P. J. E. Peebles, Astrophys. J.**221**, 19 (1978).
[277] S. D. M. White, Mon. Not. R. Astron. Soc.**186**, 145 (1979).
[278] J. N. Fry, Astrophys. J.**306**, 358 (1986).
[279] R. Balian and R. Schaeffer, Astron. Astrophys.**220**, 1 (1989).
[280] E. Gaztanaga and J. A. Frieman, Astrophys. J. Lett.**437**, L13 (1994), [arXiv:astro-ph/9407079].
[281] I. Szapudi and S. Colombi, Astrophys. J.**470**, 131 (1996), [arXiv:astro-ph/9510030].
[282] J. N. Fry, Astrophys. J.**279**, 499 (1984).
[283] R. Juszkiewicz, F. R. Bouchet and S. Colombi, Astrophys. J. Lett.**412**, L9 (1993), [arXiv:astro-ph/9306003].
[284] F. R. Bouchet, R. Schaeffer and M. Davis, Astrophys. J.**383**, 19 (1991).
[285] J. N. Fry, A. L. Melott and S. F. Shandarin, Astrophys. J.**412**, 504 (1993).
[286] F. Bernardeau, Astrophys. J.**433**, 1 (1994), [arXiv:astro-ph/9312026].
[287] C. M. Baugh, E. Gaztanaga and G. Efstathiou, Mon. Not. R. Astron. Soc.**274**, 1049 (1995), [arXiv:astro-ph/9408057].
[288] S. Colombi, F. R. Bouchet and L. Hernquist, Astrophys. J.**465**, 14 (1996), [arXiv:astro-ph/9508142].
[289] F. R. Bouchet, R. Juszkiewicz, S. Colombi and R. Pellat, Astrophys. J. Lett.**394**, L5 (1992).
[290] A. Buchalter and M. Kamionkowski, Astrophys. J.**521**, 1 (1999), [arXiv:astro-ph/9903462].
[291] E. Gaztañaga, P. Fosalba and R. A. C. Croft, Mon. Not. R. Astron. Soc.**331**, 13 (2002), [arXiv:astro-ph/0107523].
[292] C. Uhlemann, S. Codis, C. Pichon, F. Bernardeau and P. Reimberg, Mon. Not. R. Astron. Soc.**460**, 1529 (2016), [arXiv:1512.05793].
[293] S. Codis, C. Pichon, F. Bernardeau, C. Uhlemann and S. Prunet, Mon. Not. R. Astron. Soc.**460**, 1549 (2016), [arXiv:1603.03347].
[294] I. Szapudi, Mon. Not. R. Astron. Soc.**300**, L35 (1998), [arXiv:astro-ph/9805090].
[295] L. Hui and E. Gaztañaga, Astrophys. J.**519**, 622 (1999), [arXiv:astro-ph/9810194].
[296] G. Kauffmann, A. Nusser and M. Steinmetz, Mon. Not. R. Astron. Soc.**286**, 795 (1997), [arXiv:astro-ph/9512009].
[297] M. Blanton, R. Cen, J. P. Ostriker and M. A. Strauss, Astrophys. J.**522**, 590 (1999), [arXiv:astro-ph/9807029].
[298] R. Cen and J. P. Ostriker, Astrophys. J.**538**, 83 (2000).
[299] J. Bel and C. Marinoni, Mon. Not. R. Astron. Soc.**424**, 971 (2012), [arXiv:1205.3200].
[300] A. Pujol *et al.*, Mon. Not. R. Astron. Soc.**462**, 35 (2016), [arXiv:1601.00160].
[301] K. Hoffmann *et al.*, Mon. Not. R. Astron. Soc.**447**, 1724 (2015), [arXiv:1403.1259].
[302] R. E. Angulo, C. M. Baugh and C. G. Lacey, Mon. Not. R. Astron. Soc.**387**, 921 (2008), [arXiv:0712.2280].
[303] A. Paranjape, R. K. Sheth and V. Desjacques, Mon. Not. R. Astron. Soc.**431**, 1503 (2013), [arXiv:1210.1483].
[304] A. Paranjape *et al.*, Mon. Not. R. Astron. Soc.**436**, 449 (2013), [arXiv:1305.5830].
[305] M. Musso, A. Paranjape and R. K. Sheth, Mon. Not. R. Astron. Soc.**427**, 3145 (2012), [arXiv:1205.3401].

[306] M. Biagetti, K. C. Chan, V. Desjacques and A. Paranjape, Mon. Not. R. Astron. Soc.**441**, 1457 (2014), [arXiv:1310.1401].
[307] M. Manera and E. Gaztañaga, Mon. Not. R. Astron. Soc.**415**, 383 (2011), [arXiv:0912.0446].
[308] N. Roth and C. Porciani, Mon. Not. R. Astron. Soc.**415**, 829 (2011), [arXiv:1101.1520].
[309] K. C. Chan and R. Scoccimarro, Phys. Rev. D**86**, 103519 (2012), [arXiv:1204.5770].
[310] K. Hoffmann, J. Bel and E. Gaztañaga, Mon. Not. R. Astron. Soc.**465**, 2225 (2017), [arXiv:1607.01024].
[311] I. Jee, C. Park, J. Kim, Y.-Y. Choi and S. S. Kim, Astrophys. J.**753**, 11 (2012), [arXiv:1204.5573].
[312] J. L. Tinker *et al.*, Astrophys. J.**724**, 878 (2010), [arXiv:1001.3162].
[313] J. L. Tinker *et al.*, Astrophys.J. **688**, 709 (2008), [arXiv:0803.2706].
[314] T. Lazeyras, C. Wagner, T. Baldauf and F. Schmidt, JCAP **2**, 018 (2016), [arXiv:1511.01096].
[315] T. Baldauf, U. Seljak, L. Senatore and M. Zaldarriaga, JCAP **9**, 007 (2016), [arXiv:1511.01465].
[316] Y. Li, W. Hu and M. Takada, Phys. Rev. D**93**, 063507 (2016), [arXiv:1511.01454].
[317] R. K. Sheth and G. Tormen, Mon.Not.Roy.Astron.Soc. **308**, 119 (1999), [arXiv:astro-ph/9901122].
[318] B. E. Robertson, A. V. Kravtsov, J. Tinker and A. R. Zentner, The Astrophysical Journal **696**, 636 (2009).
[319] A. V. Kravtsov and A. A. Klypin, Astrophys. J.**520**, 437 (1999), [arXiv:astro-ph/9812311].
[320] Y. P. Jing, Astrophys. J. Lett.**515**, L45 (1999), [arXiv:astro-ph/9901138].
[321] U. Seljak and M. S. Warren, Mon. Not. R. Astron. Soc.**355**, 129 (2004), [arXiv:astro-ph/0403698].
[322] H. Guo and Y. P. Jing, Astrophys. J.**702**, 425 (2009), [arXiv:0907.0282].
[323] J. E. Pollack, R. E. Smith and C. Porciani, Mon. Not. R. Astron. Soc.**420**, 3469 (2012), [arXiv:1109.3458].
[324] T. Lazeyras and F. Schmidt, JCAP **1809**, 008 (2018), [arXiv:1712.07531].
[325] M. M. Abidi and T. Baldauf, JCAP **1807**, 029 (2018), [arXiv:1802.07622].
[326] E. Castorina, A. Paranjape, O. Hahn and R. K. Sheth, ArXiv e-prints (2016), [arXiv:1611.03619].
[327] R. G. Mann, J. A. Peacock and A. F. Heavens, Mon. Not. R. Astron. Soc.**293**, 209 (1998), [arXiv:astro-ph/9708031].
[328] T. Okumura, U. Seljak and V. Desjacques, JCAP **11**, 014 (2012), [arXiv:1206.4070].
[329] K. C. Chan, Phys. Rev. **D92**, 123525 (2015), [arXiv:1507.04753].
[330] R. Scoccimarro, H. A. Feldman, J. N. Fry and J. A. Frieman, Astrophys.J. **546**, 652 (2001), [arXiv:astro-ph/0004087].
[331] L. Verde *et al.*, Mon.Not.Roy.Astron.Soc. **335**, 432 (2002), [arXiv:astro-ph/0112161].
[332] T. Nishimichi *et al.*, PASJ**59**, 93 (2007), [arXiv:astro-ph/0609740].
[333] F. Marín, Astrophys. J.**737**, 97 (2011), [arXiv:1011.4530].
[334] F. A. Marín *et al.*, Mon. Not. R. Astron. Soc.**432**, 2654 (2013), [arXiv:1303.6644].
[335] H. Gil-Marín *et al.*, Mon. Not. R. Astron. Soc.**451**, 539 (2015), [arXiv:1407.5668].
[336] Z. Slepian *et al.*, Mon. Not. R. Astron. Soc.**468**, 1070 (2017).
[337] C.-T. Chiang, C. Wagner, A. G. Sánchez, F. Schmidt and E. Komatsu, JCAP **9**, 028 (2015), [arXiv:1504.03322].
[338] C.-T. Chiang, C. Wagner, F. Schmidt and E. Komatsu, JCAP **5**, 048 (2014), [arXiv:1403.3411].
[339] C. Doux *et al.*, Phys. Rev. D**94**, 103506 (2016), [arXiv:1607.03625].
[340] F. Schmidt, Phys. Rev. D**93**, 063512 (2016), [arXiv:1511.02231].
[341] R. Casas-Miranda, H. J. Mo, R. K. Sheth and G. Boerner, Mon. Not. R. Astron. Soc.**333**, 730 (2002), [arXiv:astro-ph/0105008].
[342] H. Gil-Marín, C. Wagner, L. Verde, R. Jimenez and A. F. Heavens, Mon. Not. R. Astron. Soc.**407**, 772 (2010), [arXiv:1003.3238].
[343] A. Cooray and R. K. Sheth, Phys.Rept. **372**, 1 (2002), [arXiv:astro-ph/0206508].
[344] A. Akbar Abolhasani, M. Mirbabayi and E. Pajer, JCAP **5**, 063 (2016), [arXiv:1509.07886].
[345] U. Seljak, N. Hamaus and V. Desjacques, Physical Review Letters **103**, 091303 (2009), [arXiv:0904.2963].
[346] T. Baldauf, E. Schaan and M. Zaldarriaga, JCAP **3**, 007 (2016), [arXiv:1507.02255].
[347] T. Baldauf, S. Codis, V. Desjacques and C. Pichon, Mon. Not. R. Astron. Soc.**456**, 3985 (2016), [arXiv:1510.09204].
[348] D. Ginzburg, V. Desjacques and K. C. Chan, Phys. Rev. **D96**, 083528 (2017), [arXiv:1706.08738].
[349] I. Kayo, M. Takada and B. Jain, Mon. Not. R. Astron. Soc.**429**, 344 (2013), [arXiv:1207.6322].
[350] M. Tegmark and B. C. Bromley, Astrophys. J. Lett.**518**, L69 (1999), [arXiv:astro-ph/9809324].
[351] A. Taruya, H. Magira, Y. P. Jing and Y. Suto, PASJ**53**, 155 (2001), [arXiv:astro-ph/0012511].
[352] M. Sato and T. Matsubara, Phys. Rev. D**87**, 123523 (2013), [arXiv:1304.4228].
[353] J. G. Cresswell and W. J. Percival, Mon. Not. R. Astron. Soc.**392**, 682 (2009), [arXiv:0808.1101].
[354] A. Taruya, Astrophys. J.**537**, 37 (2000), [arXiv:astro-ph/9909124].
[355] M. Blanton, Astrophys. J.**544**, 63 (2000), [arXiv:astro-ph/0003228].
[356] V. Wild *et al.*, Mon. Not. R. Astron. Soc.**356**, 247 (2005), [arXiv:astro-ph/0404275].
[357] G. Lemson and G. Kauffmann, Mon. Not. R. Astron. Soc.**302**, 111 (1999).
[358] W. J. Percival, D. Scott, J. A. Peacock and J. S. Dunlop, Mon. Not. R. Astron. Soc.**338**, L31 (2003), [arXiv:astro-ph/0208457].
[359] R. K. Sheth and G. Tormen, Mon. Not. R. Astron. Soc.**349**, 1464 (2004), [arXiv:astro-ph/0402055].
[360] V. Avila-Reese, P. Colín, S. Gottlöber, C. Firmani and C. Maulbetsch, Astrophys. J.**634**, 51 (2005), [arXiv:astro-ph/0508053].
[361] G. Harker, S. Cole, J. Helly, C. Frenk and A. Jenkins, Mon. Not. R. Astron. Soc.**367**, 1039 (2006), [arXiv:astro-ph/0510488].
[362] L. Gao, V. Springel and S. D. M. White, Mon. Not. R. Astron. Soc.**363**, L66 (2005), [arXiv:astro-ph/0506510].
[363] R. H. Wechsler, A. R. Zentner, J. S. Bullock, A. V. Kravtsov and B. Allgood, Astrophys. J.**652**, 71 (2006), [arXiv:astro-ph/0512416].
[364] A. D. Ludlow *et al.*, Mon. Not. R. Astron. Soc.**441**, 378 (2014), [arXiv:1312.0945].

[365] Y.-Y. Mao, A. R. Zentner and R. H. Wechsler, Mon. Not. Roy. Astron. Soc. **474**, 5143 (2018), [arXiv:1705.03888].
[366] J. F. Navarro, C. S. Frenk and S. D. M. White, Astrophys. J. **490**, 493 (1997), [arXiv:astro-ph/9611107].
[367] N. Dalal, M. White, J. R. Bond and A. Shirokov, Astrophys. J.**687**, 12 (2008), [arXiv:0803.3453].
[368] A. R. Wetzel, J. D. Cohn, M. White, D. E. Holz and M. S. Warren, Astrophys. J.**656**, 139 (2007), [arXiv:astro-ph/0606699].
[369] A. Faltenbacher and S. D. M. White, Astrophys. J. **708**, 469 (2010), [arXiv:0909.4302].
[370] A. Paranjape and N. Padmanabhan, Mon. Not. Roy. Astron. Soc. **468**, 2984 (2017), [arXiv:1612.02833].
[371] T. Lazeyras, M. Musso and F. Schmidt, JCAP **3**, 059 (2017), [arXiv:1612.04360].
[372] A. S. Villarreal *et al.*, Mon. Not. R. Astron. Soc.**472**, 1088 (2017), [arXiv:1705.04327].
[373] A. N. Salcedo *et al.*, Mon. Not. Roy. Astron. Soc. **475**, 4411 (2018), [arXiv:1708.08451].
[374] T. Sunayama, A. P. Hearin, N. Padmanabhan and A. Leauthaud, Mon. Not. Roy. Astron. Soc. **458**, 1510 (2016), [arXiv:1509.06417].
[375] J. R. Bond, S. Cole, G. Efstathiou and N. Kaiser, The Astrophysical Journal **379**, 440 (1991).
[376] C. Lacey and S. Cole, Mon. Not. R. Astron. Soc.**262**, 627 (1993).
[377] H. Y. Wang, H. J. Mo and Y. P. Jing, Mon. Not. R. Astron. Soc.**375**, 633 (2007), [arXiv:astro-ph/0608690].
[378] J. Diemand, M. Kuhlen and P. Madau, Astrophys. J.**667**, 859 (2007), [arXiv:astro-ph/0703337].
[379] S. More, B. Diemer and A. V. Kravtsov, Astrophys. J.**810**, 36 (2015), [arXiv:1504.05591].
[380] M. Borzyszkowski, C. Porciani, E. Romano-Díaz and E. Garaldi, Mon. Not. R. Astron. Soc.**469**, 594 (2017), [arXiv:1610.04231].
[381] A. Paranjape, O. Hahn and R. K. Sheth, Mon. Not. Roy. Astron. Soc. **476**, 3631 (2018), [arXiv:1706.09906].
[382] X. Yang, H. J. Mo and F. C. van den Bosch, Astrophys. J. Lett.**638**, L55 (2006), [arXiv:astro-ph/0509626].
[383] L. Wang, S. M. Weinmann, G. De Lucia and X. Yang, Mon. Not. R. Astron. Soc.**433**, 515 (2013), [arXiv:1305.0350].
[384] H. Miyatake *et al.*, Phys. Rev. Lett. **116**, 041301 (2016), [arXiv:1506.06135].
[385] S. More *et al.*, Astrophys. J. **825**, 39 (2016), [arXiv:1601.06063].
[386] Y.-T. Lin *et al.*, Astrophys. J.**819**, 119 (2016), [arXiv:1504.07632].
[387] Y. Zu, R. Mandelbaum, M. Simet, E. Rozo and E. S. Rykoff, Mon. Not. Roy. Astron. Soc. **470**, 551 (2017), [arXiv:1611.00366].
[388] D. J. Croton, L. Gao and S. D. M. White, Mon. Not. Roy. Astron. Soc. **374**, 1303 (2007), [arXiv:astro-ph/0605636].
[389] A. R. Zentner, A. P. Hearin and F. C. van den Bosch, Mon. Not. R. Astron. Soc.**443**, 3044 (2014), [arXiv:1311.1818].
[390] A. P. Hearin, Mon. Not. R. Astron. Soc.**451**, L45 (2015), [arXiv:1501.02798].
[391] C. S. Frenk, S. D. M. White, M. Davis and G. Efstathiou, Astrophys. J.**327**, 507 (1988).
[392] N. Katz, T. Quinn and J. M. Gelb, Mon. Not. R. Astron. Soc.**265**, 689 (1993).
[393] C. Porciani, A. Dekel and Y. Hoffman, Mon. Not. R. Astron. Soc.**332**, 339 (2002), [arXiv:astro-ph/0105165].
[394] A. D. Ludlow and C. Porciani, Mon. Not. R. Astron. Soc.**413**, 1961 (2011), [arXiv:1011.2493].
[395] A. Jenkins *et al.*, Mon. Not. R. Astron. Soc.**321**, 372 (2001), [arXiv:astro-ph/0005260].
[396] A. Paranjape and R. K. Sheth, Mon. Not. R. Astron. Soc.**426**, 2789 (2012), [arXiv:1206.3506].
[397] J. E. Gunn, Astrophys. J.**218**, 592 (1977).
[398] O. Lahav, P. B. Lilje, J. R. Primack and M. J. Rees, Mon. Not. R. Astron. Soc.**251**, 128 (1991).
[399] A. Einstein and E. G. Straus, Reviews of Modern Physics **17**, 120 (1945).
[400] R. Balbinot, R. Bergamini and A. Comastri, Phys. Rev. D**38**, 2415 (1988).
[401] M. Carrera and D. Giulini, Reviews of Modern Physics **82**, 169 (2010), [arXiv:0810.2712].
[402] E. Bertschinger, Astrophys. J. Supp.**58**, 39 (1985).
[403] M. Borzyszkowski, A. D. Ludlow and C. Porciani, Mon. Not. R. Astron. Soc.**445**, 4124 (2014), [arXiv:1405.7367].
[404] P. Valageas, Astron. Astrophys.**508**, 93 (2009), [arXiv:0905.2277].
[405] P. Valageas, Astron. Astrophys.**525**, A98 (2011), [arXiv:1009.1131].
[406] C. Giocoli, J. Moreno, R. K. Sheth and G. Tormen, Mon. Not. R. Astron. Soc.**376**, 977 (2007), [arXiv:astro-ph/0611221].
[407] R. K. Sheth and R. van de Weygaert, Mon. Not. R. Astron. Soc.**350**, 517 (2004), [arXiv:astro-ph/0311260].
[408] F. Bernardeau, Astrophys. J.**427**, 51 (1994), [arXiv:astro-ph/9311066].
[409] M. C. Neyrinck, Mon. Not. R. Astron. Soc.**428**, 141 (2013), [arXiv:1204.1326].
[410] J. A. Peacock and A. F. Heavens, Monthly Notices of the Royal Astronomical Society **243**, 133 (1990).
[411] R. I. Epstein, Mon. Not. R. Astron. Soc.**205**, 207 (1983).
[412] K. Jedamzik, The Astrophysical Journal **448**, 1 (1995).
[413] M. Maggiore and A. Riotto, Astrophys. J.**711**, 907 (2010), [arXiv:0903.1249].
[414] S. Chandrasekhar, Reviews of Modern Physics **15**, 1 (1943).
[415] A. de Simone, M. Maggiore and A. Riotto, Mon. Not. R. Astron. Soc.**412**, 2587 (2011), [arXiv:1007.1903].
[416] C.-P. Ma, M. Maggiore, A. Riotto and J. Zhang, Mon. Not. R. Astron. Soc.**411**, 2644 (2011), [arXiv:1007.4201].
[417] P. S. Corasaniti and I. Achitouv, Phys. Rev. D**84**, 023009 (2011), [arXiv:1107.1251].
[418] I. Achitouv, Y. Rasera, R. K. Sheth and P. S. Corasaniti, Physical Review Letters **111**, 231303 (2013), [arXiv:1212.1166].
[419] A. Paranjape, T. Y. Lam and R. K. Sheth, Mon. Not. Roy. Astron. Soc. **420**, 1429 (2012), [arXiv:1105.1990].
[420] M. Musso and R. K. Sheth, Mon. Not. R. Astron. Soc.**438**, 2683 (2014), [arXiv:1306.0551].
[421] M. Musso and A. Paranjape, Mon. Not. R. Astron. Soc.**420**, 369 (2012), [arXiv:1108.0565].
[422] F. Lucchin and S. Matarrese, Astrophys. J.**330**, 535 (1988).
[423] S. Colafrancesco, F. Lucchin and S. Matarrese, Astrophys. J.**345**, 3 (1989).
[424] P. Coles and J. D. Barrow, Mon. Not. R. Astron. Soc.**228**, 407 (1987).
[425] S. Matarrese, L. Verde and R. Jimenez, Astrophys. J.**541**, 10 (2000), [arXiv:astro-ph/0001366].

[426] M. Lo Verde, A. Miller, S. Shandera and L. Verde, Journal of Cosmology and Astro-Particle Physics **4**, 14 (2008), [arXiv:0711.4126].
[427] M. Maggiore and A. Riotto, Astrophys. J.**717**, 526 (2010), [arXiv:0903.1251].
[428] V. Desjacques and U. Seljak, Classical and Quantum Gravity **27**, 124011 (2010), [arXiv:1003.5020].
[429] L. Appel and B. J. T. Jones, Mon. Not. R. Astron. Soc.**245**, 522 (1990).
[430] V. Desjacques, J.-O. Gong and A. Riotto, JCAP **9**, 6 (2013), [arXiv:1301.7437].
[431] A. G. Doroshkevich, Astrophysics **6**, 320 (1970).
[432] Y. P. Jing and Y. Suto, Astrophys. J.**574**, 538 (2002), [arXiv:astro-ph/0202064].
[433] Y. Hoffman, Astrophys. J.**308**, 493 (1986).
[434] P. J. E. Peebles, Astrophys. J.**365**, 27 (1990).
[435] J. R. Bond and S. T. Myers, Astrophys. J. Supp.**103**, 1 (1996).
[436] A. Del Popolo and M. Gambera, Astron. Astrophys.**337**, 96 (1998), [arXiv:astro-ph/9802214].
[437] J. Lee and U.-L. Pen, Astrophys. J. Lett.**532**, L5 (2000), [arXiv:astro-ph/9911328].
[438] J. Lee, O. Hahn and C. Porciani, Astrophys. J.**707**, 761 (2009), [arXiv:0906.5166].
[439] G. Despali, G. Tormen and R. K. Sheth, Mon. Not. R. Astron. Soc.**431**, 1143 (2013), [arXiv:1212.4157].
[440] A. D. Ludlow, M. Borzyszkowski and C. Porciani, Mon. Not. R. Astron. Soc.**445**, 4110 (2014).
[441] D. Lynden-Bell, Astrophys. J.**139**, 1195 (1964).
[442] C. C. Lin, L. Mestel and F. H. Shu, Astrophys. J.**142**, 1431 (1965).
[443] V. Icke, Astron. Astrophys.**27**, 1 (1973).
[444] S. D. M. White and J. Silk, Astrophys. J.**231**, 1 (1979).
[445] L. Hui and E. Bertschinger, Astrophys. J.**471**, 1 (1996), [arXiv:astro-ph/9508114].
[446] T. Y. Lam and R. K. Sheth, Mon. Not. R. Astron. Soc.**386**, 407 (2008), [arXiv:0711.5029].
[447] T. Y. Lam and R. K. Sheth, Mon. Not. R. Astron. Soc.**389**, 1249 (2008), [arXiv:0805.1238].
[448] P. Monaco, Astrophys. J.**447**, 23 (1995), [arXiv:astro-ph/9406029].
[449] T. Chiueh and J. Lee, Astrophys. J.**555**, 83 (2001), [arXiv:astro-ph/0010286].
[450] E. Castorina and R. K. Sheth, Mon. Not. R. Astron. Soc.**433**, 1529 (2013), [arXiv:1301.5128].
[451] M. Musso and R. K. Sheth, Mon. Not. R. Astron. Soc.**442**, 401 (2014), [arXiv:1401.3185].
[452] V. Desjacques, Mon. Not. R. Astron. Soc.**388**, 638 (2008), [arXiv:0707.4670].
[453] Y. Lithwick and N. Dalal, Astrophys. J.**734**, 100 (2011), [arXiv:1010.3723].
[454] M. Maggiore and A. Riotto, Astrophys. J.**717**, 515 (2010), [arXiv:0903.1250].
[455] S. D. M. White, Formation and Evolution of Galaxies, in *Cosmology and Large Scale Structure*, edited by R. Schaeffer, J. Silk, M. Spiro and J. Zinn-Justin, p. 349, 1996.
[456] A. R. Zentner, International Journal of Modern Physics D **16**, 763 (2007), [arXiv:astro-ph/0611454].
[457] M. Musso *et al.*, Mon. Not. Roy. Astron. Soc. **476**, 4877 (2018), [arXiv:1709.00834].
[458] A. G. Doroshkevich, Astrofizika **6**, 581 (1970).
[459] J. A. Peacock and A. F. Heavens, Mon. Not. R. Astron. Soc.**217**, 805 (1985).
[460] Y. Hoffman and J. Shaham, Astrophys. J.**297**, 16 (1985).
[461] A. Heavens and J. Peacock, Mon. Not. R. Astron. Soc.**232**, 339 (1988).
[462] B. S. Ryden, Astrophys. J.**333**, 78 (1988).
[463] R. van de Weygaert and A. Babul, Astrophys. J. Lett.**425**, L59 (1994), [arXiv:astro-ph/9402003].
[464] P. Catelan and T. Theuns, Mon. Not. R. Astron. Soc.**282**, 436 (1996), [arXiv:astro-ph/9604077].
[465] J. R. Bond and S. T. Myers, Astrophys. J. Supp.**103**, 41 (1996).
[466] J. R. Bond and S. T. Myers, Astrophys. J. Supp.**103**, 63 (1996).
[467] C.-P. Ma and E. Bertschinger, Astrophys. J.**612**, 28 (2004), [arXiv:astro-ph/0311049].
[468] B. M. Schäfer and P. M. Merkel, Mon. Not. R. Astron. Soc.**421**, 2751 (2012), [arXiv:1101.4584].
[469] R. A. C. Croft and E. Gaztanaga, Astrophys. J.**495**, 554 (1998), [arXiv:astro-ph/9701163].
[470] S. de and R. A. C. Croft, Mon. Not. R. Astron. Soc.**401**, 1989 (2010), [arXiv:0910.1310].
[471] L. Hui, N. Y. Gnedin and Y. Zhang, Astrophys. J.**486**, 599 (1997), [arXiv:astro-ph/9608157].
[472] D. Pogosyan *et al.*, Mon. Not. R. Astron. Soc.**396**, 635 (2009), [arXiv:0811.1530].
[473] S. Codis, C. Pichon and D. Pogosyan, Mon. Not. R. Astron. Soc.**452**, 3369 (2015), [arXiv:1504.06073].
[474] J. A. Peacock, S. L. Lumsden and A. F. Heavens, Mon. Not. R. Astron. Soc.**229**, 469 (1987).
[475] W. J. Percival and B. M. Schäfer, Mon. Not. R. Astron. Soc.**385**, L78 (2008), [arXiv:0712.2729].
[476] R. van de Weygaert and E. Bertschinger, Mon. Not. R. Astron. Soc.**281**, 84 (1996), [arXiv:astro-ph/9507024].
[477] C. Porciani, Mon. Not. R. Astron. Soc.**463**, 4068 (2016), [arXiv:1609.00730].
[478] J. Dubinski, Astrophys. J.**401**, 441 (1992).
[479] E. Bertschinger and B. Jain, Astrophys. J.**431**, 486 (1994), [arXiv:astro-ph/9307033].
[480] K. C. Chan, R. K. Sheth and R. Scoccimarro, Phys. Rev. D**96**, 103543 (2017), [arXiv:1511.01909].
[481] E. Regos and A. S. Szalay, Mon. Not. R. Astron. Soc.**272**, 447 (1995).
[482] V. J. Martínez and E. Saar, *Statistics of the Galaxy Distribution* (Chapman &amp, 2002).
[483] M. Kac, Bull. Am. Math. Soc. **49**, 938 (1943).
[484] S. O. Rice, Bell System Tech. J. **25**, 46 (1945).
[485] L. Verde *et al.*, Mon. Not. R. Astron. Soc.**443**, 122 (2014), [arXiv:1404.2241].
[486] R. E. Smith, Mon. Not. R. Astron. Soc.**400**, 851 (2009), [arXiv:0810.1960].
[487] P. Coles, Mon. Not. R. Astron. Soc.**238**, 319 (1989).
[488] S. L. Lumsden, A. F. Heavens and J. A. Peacock, Mon. Not. R. Astron. Soc.**238**, 293 (1989).

[489] S. Otto, H. D. Politzer and M. B. Wise, Physical Review Letters **56**, 1878 (1986).

[490] J. M. Cline, H. D. Politzer, S.-J. Rey and M. B. Wise, Communications in Mathematical Physics **112**, 217 (1987).

[491] T. Matsubara, Astrophys. J. Supp.**101**, 1 (1995), [arXiv:astro-ph/9501056].

[492] C. Gay, C. Pichon and D. Pogosyan, Phys. Rev. **D85**, 023011 (2012), [arXiv:1110.0261].

[493] J. C. Collins, *Renormalization*, Cambridge Monographs on Mathematical Physics Vol. 26 (Cambridge University Press, Cambridge, 1986).

[494] T. Matsubara and V. Desjacques, Phys. Rev. D**93**, 123522 (2016), [arXiv:1604.06579].

[495] T. Matsubara, Phys. Rev. D**86**, 063518 (2012), [arXiv:1206.0562].

[496] J. R. Bond, The formation of Cosmic Structure, in *Frontiers in Physics - From colliders to cosmology, proceedings of the Lake Louise Winter Institute*, edited by A. Astbury, B. Campbell, W. Israel, A. Kamal and F. Khanna, pp. 182–235, 1989.

[497] G. Efstathiou and M. J. Rees, Mon. Not. Roy. Astron. Soc. **230**, 5 (1988).

[498] E. Castorina, A. Paranjape and R. K. Sheth, Mon. Not. R. Astron. Soc.**468**, 3813 (2017), [arXiv:1611.03613].

[499] A. Moradinezhad Dizgah, K. C. Chan, J. Noreña, M. Biagetti and V. Desjacques, JCAP **9**, 030 (2016), [arXiv:1512.06084].

[500] A. Manrique and E. Salvador-Sole, Astrophys. J.**453**, 6 (1995), [arXiv:astro-ph/9505083].

[501] A. Manrique and E. Salvador-Sole, Astrophys. J.**467**, 504 (1996), [arXiv:astro-ph/9511012].

[502] A. Manrique *et al.*, Astrophys. J.**499**, 548 (1998), [arXiv:astro-ph/9712081].

[503] R. Scoccimarro, Mon. Not. R. Astron. Soc.**299**, 1097 (1998), [arXiv:astro-ph/9711187].

[504] M. Crocce, S. Pueblas and R. Scoccimarro, Mon.Not.Roy.Astron.Soc. **373**, 369 (2006), [arXiv:astro-ph/0606505].

[505] N. McCullagh, D. Jeong and A. S. Szalay, Mon. Not. R. Astron. Soc.**455**, 2945 (2016), [arXiv:1507.07824].

[506] A. S. Szalay and L. G. Jensen, Acta Physica Hungarica **62**, 263 (1987).

[507] M. Biagetti, V. Desjacques, A. Kehagias, D. Racco and A. Riotto, JCAP **4**, 040 (2016), [arXiv:1508.07330].

[508] Y. B. Zel'dovich, Astron. Astrophys.**5**, 84 (1970).

[509] S. F. Shandarin and Y. B. Zeldovich, Reviews of Modern Physics **61**, 185 (1989).

[510] S. Bharadwaj, Astrophys. J.**472**, 1 (1996), [arXiv:astro-ph/9606121].

[511] R. Scoccimarro and J. Frieman, Astrophys. J. Supp.**105**, 37 (1996), [arXiv:astro-ph/9509047].

[512] T. Matsubara, Phys. Rev. D**77**, 063530 (2008), [arXiv:0711.2521].

[513] V. Desjacques, D. Jeong and F. Schmidt, JCAP **1803**, 017 (2018), [arXiv:1711.06745].

[514] T. Baldauf and V. Desjacques, Phys. Rev. D**95**, 043535 (2017), [arXiv:1612.04521].

[515] R. E. Smith, R. Scoccimarro and R. K. Sheth, Phys. Rev. D**77**, 043525 (2008), [arXiv:astro-ph/0703620].

[516] N. Padmanabhan and M. White, Phys. Rev. D**80**, 063508 (2009), [arXiv:0906.1198].

[517] D. J. Eisenstein, H.-J. Seo, E. Sirko and D. N. Spergel, Astrophys. J.**664**, 675 (2007), [arXiv:astro-ph/0604362].

[518] S. Anselmi, G. D. Starkman and R. K. Sheth, Mon. Not. R. Astron. Soc.**455**, 2474 (2016), [arXiv:1508.01170].

[519] S. Anselmi, G. D. Starkman, P.-S. Corasaniti, R. K. Sheth and I. Zehavi, Phys. Rev. Lett. **121**, 021302 (2018), [arXiv:1703.01275].

[520] S. Anselmi, P.-S. Corasaniti, G. D. Starkman, R. K. Sheth and I. Zehavi, Phys. Rev. **D98**, 023527 (2018), [arXiv:1711.09063].

[521] D. Baumann and L. McAllister, *Inflation and String Theory* (Cambridge University Press, 2015), arXiv:1404.2601.

[522] N. Bartolo, E. Komatsu, S. Matarrese and A. Riotto, Phys.Rept. **402**, 103 (2004), [arXiv:astro-ph/0406398].

[523] M. Liguori, E. Sefusatti, J. R. Fergusson and E. P. S. Shellard, Advances in Astronomy **2010** (2010), [arXiv:1001.4707].

[524] X. Chen, Advances in Astronomy **2010**, 638979 (2010), [arXiv:1002.1416].

[525] S. Matarrese, F. Lucchin and S. A. Bonometto, Astrophys. J. Lett.**310**, L21 (1986).

[526] B. Grinstein and M. B. Wise, Astrophys. J.**310**, 19 (1986).

[527] R. J. Scherrer and E. Bertschinger, Astrophys. J.**381**, 349 (1991).

[528] D. H. Weinberg and S. Cole, Mon. Not. R. Astron. Soc.**259**, 652 (1992).

[529] J. N. Fry and R. J. Scherrer, Astrophys. J.**429**, 36 (1994).

[530] R. Scoccimarro, Astrophys. J.**542**, 1 (2000), [arXiv:astro-ph/0002037].

[531] L. Verde, L. Wang, A. F. Heavens and M. Kamionkowski, Mon. Not. R. Astron. Soc.**313**, 141 (2000), [arXiv:astro-ph/9906301].

[532] L. Verde, Advances in Astronomy **2010**, 768675 (2010), [arXiv:1001.5217].

[533] J. M. Bardeen, Phys. Rev. D**22**, 1882 (1980).

[534] V. F. Mukhanov, H. A. Feldman and R. H. Brandenberger, Phys. Rep.**215**, 203 (1992).

[535] E. Komatsu and D. N. Spergel, Phys. Rev. D**63**, 063002 (2001), [arXiv:astro-ph/0005036].

[536] S. Matarrese and L. Verde, Astrophys. J. Lett.**677**, L77 (2008), [arXiv:0801.4826].

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

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

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

[540] V. Assassi, D. Baumann and F. Schmidt, JCAP **12**, 043 (2015), [arXiv:1510.03723].

[541] T. Baldauf, U. Seljak and L. Senatore, JCAP **4**, 006 (2011), [arXiv:1011.1513].

[542] L. Verde and S. Matarrese, Astrophys. J. Lett.**706**, L91 (2009), [arXiv:0909.3224].

[543] D. Tseliakhovich, C. Hirata and A. Slosar, Phys. Rev. **D82**, 043531 (2010), [arXiv:1004.3302].

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

[545] V. Desjacques and U. Seljak, Advances in Astronomy **2010**, 908640 (2010), [arXiv:1006.4763].

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

[547] E. Sefusatti, M. Crocce and V. Desjacques, Mon. Not. R. Astron. Soc.**425**, 2903 (2012), [arXiv:1111.6966].

[548] F. Schmidt and L. Hui, Physical Review Letters **110**, 011301 (2013), [arXiv:1210.2965].
[549] K. M. Smith and M. Zaldarriaga, Mon. Not. R. Astron. Soc.**417**, 2 (2011), [arXiv:astro-ph/0612571].
[550] J. Gleyzes, R. de Putter, D. Green and O. Doré, JCAP **4**, 002 (2017), [arXiv:1612.06366].
[551] M. Shiraishi, E. Komatsu, M. Peloso and N. Barnaby, JCAP **1305**, 002 (2013), [arXiv:1302.3056].
[552] F. Schmidt, N. E. Chisari and C. Dvorkin, JCAP **10**, 032 (2015), [arXiv:1506.02671].
[553] D. Jeong and M. Kamionkowski, Physical Review Letters **108**, 251301 (2012).
[554] N. E. Chisari, C. Dvorkin, F. Schmidt and D. N. Spergel, Phys. Rev. D**94**, 123507 (2016), [arXiv:1607.05232].
[555] K. Smith, M. LoVerde and M. Zaldarriaga, Phys. Rev. Lett. **107**, 191301 (2011), [arXiv:1108.1805].
[556] D. Baumann, S. Ferraro, D. Green and K. Smith, JCAP **1305**, 001 (2013), [arXiv:1209.2173].
[557] T. J. Allen, B. Grinstein and M. B. Wise, Physics Letters B **197**, 66 (1987).
[558] F. Schmidt, Phys. Rev. D**87**, 123518 (2013), [arXiv:1304.1817].
[559] M. Tellarini, A. J. Ross, G. Tasinato and D. Wands, JCAP **6**, 014 (2016), [arXiv:1603.06814].
[560] D. Babich, P. Creminelli and M. Zaldarriaga, JCAP **8**, 009 (2004), [arXiv:astro-ph/0405356].
[561] P. Creminelli, A. Nicolis, L. Senatore, M. Tegmark and M. Zaldarriaga, Journal of Cosmology and Astro-Particle Physics **5**, 4 (2006), [arXiv:astro-ph/0509029].
[562] P. Creminelli, L. Senatore, M. Zaldarriaga and M. Tegmark, JCAP **3**, 005 (2007), [arXiv:astro-ph/0610600].
[563] X. Chen and Y. Wang, Phys. Rev. D**81**, 063511 (2010), [arXiv:0909.0496].
[564] X. Chen and Y. Wang, JCAP **4**, 27 (2010), [arXiv:0911.3380].
[565] J. Maldacena, Journal of High Energy Physics **5**, 13 (2003), [arXiv:astro-ph/0210603].
[566] X. Chen, R. Easther and E. A. Lim, JCAP **4**, 010 (2008), [arXiv:0801.3295].
[567] R. Flauger and E. Pajer, JCAP **1**, 017 (2011), [arXiv:1002.0833].
[568] F.-Y. Cyr-Racine and F. Schmidt, Phys. Rev. D**84**, 083505 (2011), [arXiv:1106.2806].
[569] G. Cabass, E. Pajer and F. Schmidt, JCAP **1809**, 003 (2018), [arXiv:1804.07295].
[570] T. Tanaka and Y. Urakawa, JCAP **5**, 014 (2011), [arXiv:1103.1251].
[571] P. Creminelli, G. D'Amico, M. Musso and J. Noreña, JCAP **11**, 038 (2011), [arXiv:1106.1462].
[572] R. de Putter, O. Doré and D. Green, JCAP **10**, 024 (2015), [arXiv:1504.05935].
[573] S. Camera, R. Maartens and M. G. Santos, Mon. Not. R. Astron. Soc.**451**, L80 (2015), [arXiv:1412.4781].
[574] N. Bartolo *et al.*, Physics of the Dark Universe **13**, 30 (2016), [arXiv:1506.00915].
[575] E. Villa, L. Verde and S. Matarrese, Classical and Quantum Gravity **31**, 234005 (2014), [arXiv:1409.4738].
[576] X. Chen, M.-x. Huang, S. Kachru and G. Shiu, JCAP **1**, 002 (2007), [arXiv:hep-th/0605045].
[577] P. D. Meerburg, J. P. van der Schaar and P. Stefano Corasaniti, Journal of Cosmology and Astro-Particle Physics **5**, 18 (2009), [arXiv:0901.4044].
[578] I. Agullo and L. Parker, Phys. Rev. **D83**, 063526 (2011), [arXiv:1010.5766].
[579] M. H. Namjoo, H. Firouzjahi and M. Sasaki, Europhys. Lett. **101**, 39001 (2013), [arXiv:1210.3692].
[580] Planck Collaboration, Astron. Astrophys.**571**, A24 (2014), [arXiv:1303.5084].
[581] C. Cheung, A. L. Fitzpatrick, J. Kaplan, L. Senatore and P. Creminelli, Journal of High Energy Physics **3**, 014 (2008), [arXiv:0709.0293].
[582] L. Senatore, K. M. Smith and M. Zaldarriaga, JCAP **1**, 28 (2010), [arXiv:0905.3746].
[583] G. Cabass, E. Pajer and F. Schmidt, JCAP **1**, 003 (2017), [arXiv:1612.00033].
[584] C. T. Byrnes and K.-Y. Choi, Advances in Astronomy **2010**, 724525 (2010), [arXiv:1002.3110].
[585] D. Baumann and D. Green, Phys. Rev. D**85**, 103520 (2012), [arXiv:1109.0292].
[586] M. Mirbabayi and M. Simonović, JCAP **3**, 056 (2016), [arXiv:1507.04755].
[587] T. Noumi, M. Yamaguchi and D. Yokoyama, Journal of High Energy Physics **6**, 51 (2013), [arXiv:1211.1624].
[588] N. Arkani-Hamed and J. Maldacena, ArXiv e-prints (2015), [arXiv:1503.08043].
[589] X. Chen, M. H. Namjoo and Y. Wang, JCAP **2**, 013 (2016), [arXiv:1509.03930].
[590] H. Lee, D. Baumann and G. L. Pimentel, Journal of High Energy Physics **12**, 40 (2016), [arXiv:1607.03735].
[591] R. Flauger, M. Mirbabayi, L. Senatore and E. Silverstein, JCAP **1710**, 058 (2017), [arXiv:1606.00513].
[592] S. Endlich, A. Nicolis and J. Wang, JCAP **10**, 11 (2013), [arXiv:1210.0569].
[593] E. Dimastrogiovanni, M. Fasiello, D. Jeong and M. Kamionkowski, JCAP **12**, 050 (2014), [arXiv:1407.8204].
[594] P. Adshead and M. Wyman, Physical Review Letters **108**, 261302 (2012), [arXiv:1202.2366].
[595] I. Agullo and S. Shandera, JCAP **9**, 7 (2012), [arXiv:1204.4409].
[596] A. Ashoorioon, R. Casadio and T. Koivisto, JCAP **12**, 002 (2016), [arXiv:1605.04758].
[597] M. Shiraishi, D. Nitta, S. Yokoyama and K. Ichiki, JCAP **3**, 41 (2012), [arXiv:1201.0376].
[598] M. Shiraishi, JCAP **6**, 15 (2012), [arXiv:1202.2847].
[599] M. Biagetti, T. Lazeyras, T. Baldauf, V. Desjacques and F. Schmidt, Mon. Not. R. Astron. Soc.**468**, 3277 (2017), [arXiv:1611.04901].
[600] V. Desjacques, D. Jeong and F. Schmidt, Phys. Rev. D**84**, 061301 (2011), [arXiv:1105.3476].
[601] N. Roth and C. Porciani, Mon. Not. Roy. Astron. Soc. **425**, L81 (2012), [arXiv:1205.3165].
[602] H. D. Politzer and M. B. Wise, Astrophys. J. Lett.**285**, L1 (1984).
[603] L. G. Jensen and A. S. Szalay, Astrophys. J. Lett.**305**, L5 (1986).
[604] A. Taruya, K. Koyama and T. Matsubara, Phys. Rev. D**78**, 123534 (2008), [arXiv:0808.4085].
[605] E. Sefusatti, Phys. Rev. D**80**, 123002 (2009), [arXiv:0905.0717].
[606] V. Desjacques and U. Seljak, Phys. Rev. D**81**, 023006 (2010), [arXiv:0907.2257].
[607] S. Shandera, N. Dalal and D. Huterer, JCAP **3**, 017 (2011), [arXiv:1010.3722].
[608] M. Biagetti and V. Desjacques, Mon. Not. Roy. Astron. Soc. **451**, 3643 (2015), [arXiv:1501.04982].

[609] V. Desjacques, U. Seljak and I. T. Iliev, Mon. Not. R. Astron. Soc. **396**, 85 (2009), [arXiv:0811.2748].
[610] M. Grossi *et al.*, Mon. Not. R. Astron. Soc.**398**, 321 (2009), [arXiv:0902.2013].
[611] A. Pillepich, C. Porciani and O. Hahn, Mon. Not. R. Astron. Soc.**402**, 191 (2010), [arXiv:0811.4176].
[612] T. Nishimichi, A. Taruya, K. Koyama and C. Sabiu, JCAP **7**, 002 (2010), [arXiv:0911.4768].
[613] V. Assassi, D. Baumann, E. Pajer, Y. Welling and D. van der Woude, JCAP **11**, 024 (2015), [arXiv:1505.06668].
[614] N. Hamaus, U. Seljak and V. Desjacques, Phys. Rev. D**84**, 083509 (2011), [arXiv:1104.2321].
[615] S. More, A. V. Kravtsov, N. Dalal and S. Gottlöber, Astrophys. J. Supp.**195**, 4 (2011), [arXiv:1103.0005].
[616] C. Wagner and L. Verde, JCAP **3**, 002 (2012), [arXiv:1102.3229].
[617] B. Diemer, S. More and A. V. Kravtsov, Astrophys. J.**766**, 25 (2013), [arXiv:1207.0816].
[618] G. Despali *et al.*, Mon. Not. Roy. Astron. Soc. **456**, 2486 (2016), [arXiv:1507.05627].
[619] N. Afshordi and A. J. Tolley, Phys. Rev. D**78**, 123507 (2008), [arXiv:0806.1046].
[620] B. A. Reid, L. Verde, K. Dolag, S. Matarrese and L. Moscardini, JCAP **7**, 013 (2010), [arXiv:1004.1637].
[621] K. M. Smith, S. Ferraro and M. LoVerde, JCAP **3**, 32 (2012), [arXiv:1106.0503].
[622] K. M. Smith and M. LoVerde, JCAP **11**, 009 (2011), [arXiv:1010.0055].
[623] A. J. Ross *et al.*, Mon. Not. R. Astron. Soc.**428**, 1116 (2013), [arXiv:1208.1491].
[624] B. Leistedt and H. V. Peiris, Mon. Not. R. Astron. Soc.**444**, 2 (2014), [arXiv:1404.6530].
[625] S. Ho *et al.*, JCAP **5**, 040 (2015), [arXiv:1311.2597].
[626] T. Giannantonio *et al.*, Phys. Rev. **D89**, 023511 (2014), [arXiv:1303.1349].
[627] B. Leistedt, H. V. Peiris and N. Roth, Phys. Rev. Lett. **113**, 221301 (2014), [arXiv:1405.4315].
[628] C. Carbone, O. Mena and L. Verde, JCAP **7**, 020 (2010), [arXiv:1003.0456].
[629] N. Agarwal, S. Ho and S. Shandera, JCAP **1402**, 038 (2014), [arXiv:1311.2606].
[630] R. de Putter and O. Doré, arXiv:1412.3854.
[631] A. Raccanelli *et al.*, JCAP **1501**, 042 (2015), [arXiv:1406.0010].
[632] S. Camera, M. G. Santos and R. Maartens, Mon. Not. Roy. Astron. Soc. **448**, 1035 (2015), [arXiv:1409.8286].
[633] O. Doré *et al.*, ArXiv e-prints (2014), [arXiv:1412.4872].
[634] D. Alonso and P. G. Ferreira, Phys. Rev. **D92**, 063525 (2015), [arXiv:1507.03550].
[635] A. Raccanelli *et al.*, arXiv:1507.05903.
[636] A. Moradinezhad Dizgah and C. Dvorkin, JCAP **1801**, 010 (2018), [arXiv:1708.06473].
[637] T. Giannantonio, C. Porciani, J. Carron, A. Amara and A. Pillepich, Mon. Not. R. Astron. Soc.**422**, 2854 (2012), [arXiv:1109.0958].
[638] M. Alvarez *et al.*, ArXiv e-prints (2014), [arXiv:1412.4671].
[639] R. de Putter and O. Doré, ArXiv e-prints (2014), [arXiv:1412.3854].
[640] A. Raccanelli, O. Doré and N. Dalal, JCAP **8**, 034 (2015), [arXiv:1409.1927].
[641] U. Seljak, Physical Review Letters **102**, 021302 (2009), [arXiv:0807.1770].
[642] A. Slosar, JCAP **3**, 004 (2009), [arXiv:0808.0044].
[643] P. McDonald and U. Seljak, JCAP **10**, 007 (2009), [arXiv:0810.0323].
[644] C. Cunha, D. Huterer and O. Doré, Phys. Rev. D**82**, 023004 (2010), [arXiv:1003.2416].
[645] Particle Data Group, C. Patrignani *et al.*, Chin. Phys. **C40**, 100001 (2016).
[646] M. Shoji and E. Komatsu, Monthly Notices of the Royal Astronomical Society **81**, 123516 (2010).
[647] J. Lesgourgues and S. Pastor, Phys. Rep.**429**, 307 (2006), [arXiv:astro-ph/0603494].
[648] W. Hu and D. J. Eisenstein, Astrophys. J.**498**, 497 (1998), [arXiv:astro-ph/9710216].
[649] W. Hu, D. J. Eisenstein and M. Tegmark, Phys. Rev. Lett. **80**, 5255 (1998).
[650] J. Brandbyge and S. Hannestad, JCAP **1**, 021 (2010), [arXiv:0908.1969].
[651] S. Bird, M. Viel and M. G. Haehnelt, Mon. Not. R. Astron. Soc.**420**, 2551 (2012), [arXiv:1109.4416].
[652] C. Wagner, L. Verde and R. Jimenez, Astrophys. J. Lett.**752**, L31 (2012), [arXiv:1203.5342].
[653] Y. Y. Y. Wong, JCAP **10**, 035 (2008), [arXiv:0809.0693].
[654] S. Saito, M. Takada and A. Taruya, Phys. Rev. D**80**, 083528 (2009), [arXiv:0907.2922].
[655] K. Ichiki and M. Takada, Phys. Rev. D**85**, 063521 (2012), [arXiv:1108.4688].
[656] F. Villaescusa-Navarro *et al.*, JCAP **3**, 011 (2014), [arXiv:1311.0866].
[657] E. Castorina, E. Sefusatti, R. K. Sheth, F. Villaescusa-Navarro and M. Viel, JCAP **2**, 049 (2014), [arXiv:1311.1212].
[658] M. Costanzi *et al.*, JCAP **12**, 012 (2013), [arXiv:1311.1514].
[659] A. Ringwald and Y. Y. Y. Wong, JCAP **12**, 005 (2004), [arXiv:hep-ph/0408241].
[660] M. LoVerde, Phys. Rev. D**90**, 083530 (2014), [arXiv:1405.4855].
[661] M. LoVerde, Phys. Rev. D**90**, 083518 (2014), [arXiv:1405.4858].
[662] S. Singh and C.-P. Ma, Phys. Rev. D**67**, 023506 (2003), [arXiv:astro-ph/0208419].
[663] M. Viel, M. G. Haehnelt and V. Springel, JCAP **6**, 015 (2010), [arXiv:1003.2422].
[664] F. Marulli, C. Carbone, M. Viel, L. Moscardini and A. Cimatti, Mon. Not. R. Astron. Soc.**418**, 346 (2011), [arXiv:1103.0278].
[665] Y. Ali-Haïmoud and S. Bird, Mon. Not. R. Astron. Soc.**428**, 3375 (2013), [arXiv:1209.0461].
[666] E. Castorina, C. Carbone, J. Bel, E. Sefusatti and K. Dolag, JCAP **7**, 043 (2015), [arXiv:1505.07148].
[667] A. Banerjee and N. Dalal, JCAP **11**, 015 (2016), [arXiv:1606.06167].
[668] F. Villaescusa-Navarro *et al.*, Astrophys. J. **861**, 53 (2018), [arXiv:1708.01154].
[669] C.-T. Chiang, W. Hu, Y. Li and M. LoVerde, Phys. Rev. D**97**, 123526 (2018), [arXiv:1710.01310].
[670] L. Hui and K. P. Parfrey, Phys. Rev. D**77**, 043527 (2008), [arXiv:0712.1162].
[671] K. Parfrey, L. Hui and R. K. Sheth, Phys. Rev. D**83**, 063511 (2011), [arXiv:1012.1335].

[672] L. Senatore and M. Zaldarriaga, ArXiv e-prints (2017), [arXiv:1707.04698].
[673] M. LoVerde, Phys. Rev. D**93**, 103526 (2016), [arXiv:1602.08108].
[674] C.-T. Chiang, M. LoVerde and F. Villaescusa-Navarro, arXiv e-prints , arXiv:1811.12412 (2018), [arXiv:1811.12412].
[675] M. Shoji and E. Komatsu, Astrophys. J.**700**, 705 (2009), [arXiv:0903.2669].
[676] F. Bernardeau, N. Van de Rijt and F. Vernizzi, Phys. Rev. D**87**, 043530 (2013), [arXiv:1209.3662].
[677] M. Lewandowski, A. Perko and L. Senatore, JCAP **5**, 019 (2015), [arXiv:1412.5049].
[678] F. Schmidt, Phys. Rev. **D94**, 063508 (2016), [arXiv:1602.09059].
[679] K. Ahn, Astrophys. J.**830**, 68 (2016), [arXiv:1603.09356].
[680] J. A. Blazek, J. E. McEwen and C. M. Hirata, Physical Review Letters **116**, 121303 (2016), [arXiv:1510.03554].
[681] R. Barkana and A. Loeb, Mon. Not. R. Astron. Soc.**415**, 3113 (2011), [arXiv:1009.1393].
[682] D. Grin, O. Doré and M. Kamionkowski, Phys. Rev. D**84**, 123003 (2011), [arXiv:1107.5047].
[683] M. T. Soumagnac *et al.*, Physical Review Letters **116**, 201302 (2016), [arXiv:1602.01839].
[684] D. Tseliakhovich and C. Hirata, Phys. Rev. D**82**, 083520 (2010), [arXiv:1005.2416].
[685] N. Dalal, U.-L. Pen and U. Seljak, JCAP **11**, 7 (2010), [arXiv:1009.4704].
[686] J. Yoo, N. Dalal and U. Seljak, JCAP **7**, 18 (2011), [arXiv:1105.3732].
[687] J. Yoo and U. Seljak, Phys. Rev. D**88**, 103520 (2013), [arXiv:1308.1401].
[688] Z. Slepian and D. J. Eisenstein, Mon. Not. R. Astron. Soc.**448**, 9 (2015), [arXiv:1411.4052].
[689] F. Beutler, U. Seljak and Z. Vlah, Mon. Not. R. Astron. Soc.**470**, 2723 (2017), [arXiv:1612.04720].
[690] D. Tseliakhovich, R. Barkana and C. M. Hirata, Mon. Not. R. Astron. Soc.**418**, 906 (2011), [arXiv:1012.2574].
[691] E. Visbal, R. Barkana, A. Fialkov, D. Tseliakhovich and C. M. Hirata, Nature (London)**487**, 70 (2012), [arXiv:1201.1005].
[692] C. Popa, S. Naoz, F. Marinacci and M. Vogelsberger, Mon. Not. R. Astron. Soc.**460**, 1625 (2016), [arXiv:1512.06862].
[693] Z. Slepian *et al.*, Mon. Not. Roy. Astron. Soc. **474**, 2109 (2018), [arXiv:1607.06098].
[694] J. Frieman, M. Turner and D. Huterer, Ann. Rev. Astron. Astrophys. **46**, 385 (2008), [arXiv:0803.0982].
[695] E. J. Copeland, M. Sami and S. Tsujikawa, Int. J. Mod. Phys. **D15**, 1753 (2006), [arXiv:hep-th/0603057].
[696] P. Creminelli, G. D'Amico, J. Noreña, L. Senatore and F. Vernizzi, JCAP **3**, 27 (2010), [arXiv:0911.2701].
[697] M. Lewandowski, A. Maleknejad and L. Senatore, JCAP **5**, 038 (2017), [arXiv:1611.07966].
[698] C.-T. Chiang, Y. Li, W. Hu and M. LoVerde, Phys. Rev. D**94**, 123502 (2016), [arXiv:1609.01701].
[699] T. Clifton, P. G. Ferreira, A. Padilla and C. Skordis, PhysRep **513**, 1 (2012).
[700] A. Joyce, B. Jain, J. Khoury and M. Trodden, Phys. Rep.**568**, 1 (2015), [arXiv:1407.0059].
[701] A. Joyce, L. Lombriser and F. Schmidt, Annual Review of Nuclear and Particle Science **66**, 95 (2016), [arXiv:1601.06133].
[702] C. Brans and R. H. Dicke, Phys. Rev. **124**, 925 (1961).
[703] S. M. Carroll, V. Duvvuri, M. Trodden and M. S. Turner, Phys. Rev. **D70**, 043528 (2004), [arXiv:astro-ph/0306438].
[704] A. A. Starobinsky, JETP Lett. **86**, 157 (2007), [arXiv:0706.2041].
[705] T. Chiba, Phys. Lett. **B575**, 1 (2003), [arXiv:astro-ph/0307338].
[706] J. Wang, L. Hui and J. Khoury, Phys. Rev. Lett. **109**, 241301 (2012), [arXiv:1208.4612].
[707] J. Neyman and E. L. Scott, Astrophys. J.**116**, 144 (1952).
[708] C.-P. Ma and J. N. Fry, Astrophys. J.**543**, 503 (2000), [arXiv:astro-ph/0003343].
[709] U. Seljak, Mon. Not. R. Astron. Soc.**318**, 203 (2000), [arXiv:astro-ph/0001493].
[710] J. A. Peacock and R. E. Smith, Mon. Not. R. Astron. Soc.**318**, 1144 (2000), [arXiv:astro-ph/0005010].
[711] R. Scoccimarro, R. K. Sheth, L. Hui and B. Jain, Astrophys. J.**546**, 20 (2001), [arXiv:astro-ph/0006319].
[712] A. A. Berlind and D. H. Weinberg, Astrophys. J.**575**, 587 (2002), [arXiv:astro-ph/0109001].
[713] A. V. Kravtsov *et al.*, Astrophys. J.**609**, 35 (2004), [arXiv:astro-ph/0308519].
[714] Z. Zheng *et al.*, Astrophys. J.**633**, 791 (2005), [arXiv:astro-ph/0408564].
[715] A. Vale and J. P. Ostriker, Mon. Not. R. Astron. Soc.**353**, 189 (2004), [arXiv:astro-ph/0402500].
[716] P. Martini and D. H. Weinberg, Astrophys. J.**547**, 12 (2001), [arXiv:astro-ph/0002384].
[717] A. J. Benson, S. Cole, C. S. Frenk, C. M. Baugh and C. G. Lacey, Mon. Not. R. Astron. Soc.**311**, 793 (2000), [arXiv:astro-ph/9903343].
[718] L. R. Abramo, I. Balmès, F. Lacasa and M. Lima, Mon. Not. R. Astron. Soc.**454**, 2844 (2015), [arXiv:1506.02315].
[719] Z. Zheng, A. L. Coil and I. Zehavi, Astrophys. J.**667**, 760 (2007), [arXiv:astro-ph/0703457].
[720] H. Guo *et al.*, Mon. Not. R. Astron. Soc.**453**, 4368 (2015), [arXiv:1505.07861].
[721] S. More, Astrophys. J.**741**, 19 (2011), [arXiv:1107.1498].
[722] A. Vale and J. P. Ostriker, Mon. Not. R. Astron. Soc.**371**, 1173 (2006), [arXiv:astro-ph/0511816].
[723] C. Conroy, R. H. Wechsler and A. V. Kravtsov, Astrophys. J.**647**, 201 (2006), [arXiv:astro-ph/0512234].
[724] S. A. Rodríguez-Torres *et al.*, Mon. Not. R. Astron. Soc.**460**, 1173 (2016), [arXiv:1509.06404].
[725] A. Leauthaud *et al.*, Astrophys. J.**744**, 159 (2012), [arXiv:1104.0928].
[726] P. S. Behroozi, R. H. Wechsler and C. Conroy, Astrophys. J.**770**, 57 (2013), [arXiv:1207.6105].
[727] P. Monaco, Galaxies **4**, 53 (2016), [arXiv:1605.07752].
[728] M. Vogelsberger *et al.*, Mon. Not. R. Astron. Soc.**444**, 1518 (2014), [arXiv:1405.2921].
[729] Y. Dubois *et al.*, Mon. Not. R. Astron. Soc.**444**, 1453 (2014), [arXiv:1402.1165].
[730] T. Di Matteo *et al.*, Astrophys. J. Lett.**745**, L29 (2012), [arXiv:1107.1253].
[731] M. Velliscig *et al.*, Mon. Not. R. Astron. Soc.**454**, 3328 (2015), [arXiv:1507.06996].
[732] A. J. S. Hamilton, Astrophys. J. Lett.**331**, L59 (1988).
[733] G. Boerner, H. Mo and Y. Zhou, Astron. Astrophys.**221**, 191 (1989).
[734] C. N. A. Willmer, L. N. da Costa and P. S. Pellegrini, Astron. J.**115**, 869 (1998), [arXiv:astro-ph/9803118].
[735] C. Benoist *et al.*, Astrophys. J.**514**, 563 (1999).

[736] P. Norberg *et al.*, Mon. Not. R. Astron. Soc.**332**, 827 (2002), [arXiv:astro-ph/0112043].
[737] I. Zehavi *et al.*, Astrophys. J.**621**, 22 (2005), [arXiv:astro-ph/0411557].
[738] C. Beisbart, M. Kerscher and K. Mecke, Mark Correlations: Relating Physical Properties to Spatial Distributions, in *Morphology of Condensed Matter*, edited by K. Mecke and D. Stoyan, , Lecture Notes in Physics, Berlin Springer Verlag Vol. 600, pp. 358–390, 2002, [arXiv:physics/0201069].
[739] R. K. Sheth, Mon. Not. R. Astron. Soc.**364**, 796 (2005), [arXiv:astro-ph/0511772].
[740] R. Skibba, R. K. Sheth, A. J. Connolly and R. Scranton, Mon. Not. R. Astron. Soc.**369**, 68 (2006), [arXiv:astro-ph/0512463].
[741] R. A. Skibba and R. K. Sheth, Mon. Not. R. Astron. Soc.**392**, 1080 (2009), [arXiv:0805.0310].
[742] C. M. Hirata, Mon. Not. R. Astron. Soc.**399**, 1074 (2009), [arXiv:0903.4929].
[743] E. Krause and C. M. Hirata, Mon. Not. R. Astron. Soc.**410**, 2730 (2011), [arXiv:1004.3611].
[744] W. Fang, L. Hui, B. Ménard, M. May and R. Scranton, Phys. Rev. D**84**, 063012 (2011), [arXiv:1105.3421].
[745] P. Catelan, M. Kamionkowski and R. D. Blandford, Mon. Not. R. Astron. Soc.**320**, L7 (2001), [arXiv:astro-ph/0005470].
[746] C. M. Hirata *et al.*, Mon. Not. R. Astron. Soc.**381**, 1197 (2007), [arXiv:astro-ph/0701671].
[747] T. Okumura and Y. P. Jing, Astrophys. J. Lett.**694**, L83 (2009), [arXiv:0812.2935].
[748] S. Singh, R. Mandelbaum and S. More, Mon. Not. Roy. Astron. Soc. **450**, 2195 (2015), [arXiv:1411.1755].
[749] Z. Zheng, R. Cen, H. Trac and J. Miralda-Escudé, Astrophys. J.**726**, 38 (2011), [arXiv:1003.4990].
[750] J. S. B. Wyithe and M. Dijkstra, Mon. Not. R. Astron. Soc.**415**, 3929 (2011), [arXiv:1104.0712].
[751] C. Behrens, C. Byrohl, S. Saito and J. C. Niemeyer, Astron. Astrophys. **614**, A31 (2018), [arXiv:1710.06171].
[752] B. Greig, E. Komatsu and J. S. B. Wyithe, Mon. Not. R. Astron. Soc.**431**, 1777 (2013), [arXiv:1212.0977].
[753] V. Desjacques, D. Jeong and F. Schmidt, JCAP **1812**, 035 (2018), [arXiv:1806.04015].
[754] S. R. Furlanetto, S. P. Oh and F. H. Briggs, Phys. Rep.**433**, 181 (2006), [arXiv:astro-ph/0608032].
[755] E. Visbal and A. Loeb, JCAP **11**, 016 (2010), [arXiv:1008.3178].
[756] D. Jeong and F. Schmidt, Classical and Quantum Gravity **32**, 044001 (2015), [arXiv:1407.7979].
[757] C. Alcock and B. Paczyński, Nature (London)**281**, 358 (1979).
[758] N. Padmanabhan and M. White, Phys. Rev. D**77**, 123540 (2008), [arXiv:0804.0799].
[759] M. Shoji, D. Jeong and E. Komatsu, Astrophys. J.**693**, 1404 (2009), [arXiv:0805.4238].
[760] F. Schmidt and D. Jeong, Physical Review D **86**, 083527 (2012).
[761] J. Yoo, A. L. Fitzpatrick and M. Zaldarriaga, Phys. Rev. D**80**, 083514 (2009), [arXiv:0907.0707].
[762] J. Yoo, Phys. Rev. D**82**, 083508 (2010), [arXiv:1009.3021].
[763] D. Bertacca, R. Maartens and C. Clarkson, JCAP **9**, 037 (2014), [arXiv:1405.4403].
[764] D. Jeong and F. Schmidt, Phys. Rev. D**89**, 043519 (2014), [arXiv:1305.1299].
[765] R. Moessner and B. Jain, Mon. Not. R. Astron. Soc.**294**, L18 (1998), [arXiv:astro-ph/9709159].
[766] F. Schmidt, E. Rozo, S. Dodelson, L. Hui and E. Sheldon, Physical Review Letters **103**, 051301 (2009), [arXiv:0904.4702].
[767] M. Bruni *et al.*, Phys. Rev. D**85**, 041301 (2012), [arXiv:1106.3999].
[768] J. Yoo, N. Hamaus, U. c. v. Seljak and M. Zaldarriaga, Phys. Rev. D **86**, 063514 (2012), [arXiv:1206.5809].
[769] C. Bonvin, L. Hui and E. Gaztañaga, Phys. Rev. D**89**, 083535 (2014), [arXiv:1309.1321].
[770] J. Yoo and M. Zaldarriaga, Phys. Rev. D**90**, 023513 (2014), [arXiv:1406.4140].
[771] E. Di Dio, R. Durrer, G. Marozzi and F. Montanari, JCAP **1**, 016 (2016), [arXiv:1510.04202].
[772] O. Umeh, S. Jolicoeur, R. Maartens and C. Clarkson, JCAP **3**, 034 (2017), [arXiv:1610.03351].
[773] E. Di Dio *et al.*, JCAP **3**, 006 (2017), [arXiv:1611.03720].
[774] D. Jeong and F. Schmidt, Phys. Rev. D**86**, 083512 (2012), [arXiv:1205.1512].
[775] L. Hui, E. Gaztañaga and M. Loverde, Phys. Rev. D**76**, 103502 (2007), [arXiv:0706.1071].
[776] L. Hui, E. Gaztañaga and M. Loverde, Phys. Rev. D**77**, 063526 (2008), [arXiv:0710.4191].
[777] F. Schmidt, A. Vallinotto, E. Sefusatti and S. Dodelson, Phys. Rev. D**78**, 043513 (2008), [arXiv:0804.0373].
[778] R. Scoccimarro, Phys. Rev. D**70**, 083007 (2004), [arXiv:astro-ph/0407214].
[779] A. J. S. Hamilton, Astrophysical Journal **385**, L5 (1992).
[780] T. Matsubara, Astrophys. J.**615**, 573 (2004), [arXiv:astro-ph/0408349].
[781] W. J. Percival and M. White, Mon. Not. R. Astron. Soc.**393**, 297 (2009), [arXiv:0808.0003].
[782] L. Guzzo *et al.*, Nature (London)**451**, 541 (2008), [arXiv:0802.1944].
[783] C. Blake *et al.*, Mon. Not. R. Astron. Soc.**415**, 2876 (2011), [arXiv:1104.2948].
[784] B. A. Reid *et al.*, Mon. Not. R. Astron. Soc.**426**, 2719 (2012), [arXiv:1203.6641].
[785] F. Beutler *et al.*, Mon. Not. R. Astron. Soc.**423**, 3430 (2012), [arXiv:1204.4725].
[786] L. Samushia *et al.*, Mon. Not. R. Astron. Soc.**439**, 3504 (2014), [arXiv:1312.4899].
[787] S. Satpathy *et al.*, Mon. Not. R. Astron. Soc.**469**, 1369 (2017), [arXiv:1607.03148].
[788] G. M. Bernstein and Y.-C. Cai, Mon. Not. R. Astron. Soc.**416**, 3009 (2011), [arXiv:1104.3862].
[789] L. R. Abramo, Mon. Not. R. Astron. Soc.**420**, 2042 (2012), [arXiv:1108.5449].
[790] C. Blake *et al.*, Mon. Not. R. Astron. Soc.**436**, 3089 (2013), [arXiv:1309.5556].
[791] A. J. S. Hamilton, Linear Redshift Distortions: a Review, in *The Evolving Universe*, edited by D. Hamilton, , Astrophysics and Space Science Library Vol. 231, p. 185, 1998, [arXiv:astro-ph/9708102].
[792] J. E. Gunn and B. A. Peterson, Astrophys. J.**142**, 1633 (1965).
[793] R. A. C. Croft, D. H. Weinberg, N. Katz and L. Hernquist, Astrophys. J.**495**, 44 (1998), [arXiv:astro-ph/9708018].
[794] R. A. C. Croft, D. H. Weinberg, M. Pettini, L. Hernquist and N. Katz, Astrophys. J.**520**, 1 (1999), [arXiv:astro-ph/9809401].
[795] M. Viel, M. G. Haehnelt and V. Springel, Mon. Not. R. Astron. Soc.**354**, 684 (2004), [arXiv:astro-ph/0404600].

[796] P. McDonald, U. Seljak, R. Cen, P. Bode and J. P. Ostriker, Mon. Not. R. Astron. Soc.**360**, 1471 (2005), [arXiv:astro-ph/0407378].

[797] P. McDonald *et al.*, Astrophys. J. Supp.**163**, 80 (2006), [arXiv:astro-ph/0405013].

[798] A. Slosar *et al.*, JCAP **9**, 001 (2011), [arXiv:1104.5244].

[799] U. Seljak, JCAP **3**, 004 (2012), [arXiv:1201.0594].

[800] N. Palanque-Delabrouille *et al.*, Astron. Astrophys.**559**, A85 (2013), [arXiv:1306.5896].

[801] S. Gontcho A Gontcho, J. Miralda-Escudé and N. G. Busca, Mon. Not. R. Astron. Soc.**442**, 187 (2014), [arXiv:1404.7425].

[802] K.-G. Lee *et al.*, Astrophys. J. Lett.**795**, L12 (2014), [arXiv:1409.5632].

[803] E. Gaztañaga and R. Scoccimarro, Mon. Not. R. Astron. Soc.**361**, 824 (2005), [arXiv:astro-ph/0501637].

[804] L. Verde, A. F. Heavens, S. Matarrese and L. Moscardini, Mon. Not. R. Astron. Soc.**300**, 747 (1998), [arXiv:astro-ph/9806028].

[805] R. Scoccimarro, H. M. P. Couchman and J. A. Frieman, Astrophys. J.**517**, 531 (1999), [arXiv:astro-ph/9808305].

[806] N. McCullagh and A. S. Szalay, Astrophys. J.**798**, 137 (2015), [arXiv:1411.1249].

[807] E. Hivon, F. R. Bouchet, S. Colombi and R. Juszkiewicz, Astron. Astrophys.**298**, 643 (1995), [arXiv:astro-ph/9407049].

[808] M. Sato and T. Matsubara, Phys. Rev. D**84**, 043501 (2011), [arXiv:1105.5007].

[809] A. Taruya, T. Nishimichi and S. Saito, Phys. Rev. D**82**, 063522 (2010), [arXiv:1006.0699].

[810] L. Fonseca de la Bella, D. Regan, D. Seery and D. Parkinson, ArXiv e-prints (2018), [arXiv:1805.12394].

[811] J. L. Tinker, D. H. Weinberg and Z. Zheng, Mon. Not. R. Astron. Soc.**368**, 85 (2006), [arXiv:astro-ph/0501029].

[812] J. L. Tinker, Mon. Not. R. Astron. Soc.**374**, 477 (2007), [arXiv:astro-ph/0604217].

[813] B. A. Reid, H.-J. Seo, A. Leauthaud, J. L. Tinker and M. White, Mon. Not. R. Astron. Soc.**444**, 476 (2014), [arXiv:1404.3742].

[814] M. Shoji and J. Lee, ArXiv e-prints (2012), [arXiv:1203.0869].

[815] N. Hamaus, P. M. Sutter, G. Lavaux and B. D. Wandelt, JCAP **11**, 036 (2015), [arXiv:1507.04363].

[816] C.-H. Chuang *et al.*, Phys. Rev. D**95**, 063528 (2017), [arXiv:1605.05352].

[817] Y.-C. Cai, A. Taylor, J. A. Peacock and N. Padilla, Mon. Not. R. Astron. Soc.**462**, 2465 (2016), [arXiv:1603.05184].

[818] A. J. Hawken *et al.*, Astron. Astrophys.**607**, A54 (2017), [arXiv:1611.07046].

[819] F. Simpson, J. B. James, A. F. Heavens and C. Heymans, Physical Review Letters **107**, 271301 (2011), [arXiv:1107.5169].

[820] F. Simpson *et al.*, Phys. Rev. D**93**, 023525 (2016), [arXiv:1505.03865].

[821] E. Hubble, Astrophys. J.**79**, 8 (1934).

[822] A. J. S. Hamilton, Astrophys. J. Lett.**292**, L35 (1985).

[823] P. Coles and B. Jones, Mon. Not. R. Astron. Soc.**248**, 1 (1991).

[824] B. L. Falck, M. C. Neyrinck, M. A. Aragon-Calvo, G. Lavaux and A. S. Szalay, Astrophys. J.**745**, 17 (2012), [arXiv:1111.4466].

[825] N. McCullagh, M. C. Neyrinck, I. Szapudi and A. S. Szalay, Astrophys. J. Lett.**763**, L14 (2013), [arXiv:1211.3130].

[826] J. Carron and I. Szapudi, Mon. Not. R. Astron. Soc.**434**, 2961 (2013), [arXiv:1306.1230].

[827] A. S. Szalay, T. Matsubara and S. D. Landy, Astrophys. J. Lett.**498**, L1 (1998), [arXiv:astro-ph/9712007].

[828] T. Matsubara, Astrophys. J.**535**, 1 (2000), [arXiv:astro-ph/9908056].

[829] S. Bharadwaj, Astrophys.J. **516**, 507 (1999), [arXiv:astro-ph/9812274].

[830] I. Szapudi, Astrophys.J. **614**, 51 (2004), [arXiv:astro-ph/0404477].

[831] P. Pápai and I. Szapudi, Mon. Not. R. Astron. Soc.**389**, 292 (2008), [arXiv:0802.2940].

[832] A. Raccanelli, L. Samushia and W. J. Percival, Mon. Not. R. Astron. Soc.**409**, 1525 (2010), [arXiv:1006.1652].

[833] J. Yoo and V. Desjacques, Phys. Rev. D**88**, 023502 (2013), [arXiv:1301.4501].

[834] A. Raccanelli, D. Bertacca, O. Doré and R. Maartens, JCAP **8**, 022 (2014), [arXiv:1306.6646].

[835] L. Verde, A. F. Heavens and S. Matarrese, Mon. Not. R. Astron. Soc.**318**, 584 (2000), [arXiv:astro-ph/0002240].

[836] A. J. S. Hamilton, Astrophys. J.**417**, 19 (1993).

[837] C. Bonvin and R. Durrer, Phys. Rev. D**84**, 063505 (2011), [arXiv:1105.5280].

[838] J. Binney and T. Quinn, Mon. Not. R. Astron. Soc.**249**, 678 (1991).

[839] K. B. Fisher, O. Lahav, Y. Hoffman, D. Lynden-Bell and S. Zaroubi, Mon. Not. R. Astron. Soc.**272**, 885 (1995), [arXiv:astro-ph/9406009].

[840] A. F. Heavens and A. N. Taylor, Mon. Not. R. Astron. Soc.**275**, 483 (1995), [arXiv:astro-ph/9409027].

[841] K. B. Fisher, C. A. Scharf and O. Lahav, Mon. Not. R. Astron. Soc.**266**, 219 (1994), [arXiv:astro-ph/9309027].

[842] A. Rassat and A. Refregier, Astron. Astrophys.**540**, A115 (2012), [arXiv:1112.3100].

[843] C. Shapiro, R. G. Crittenden and W. J. Percival, Mon. Not. R. Astron. Soc.**422**, 2341 (2012), [arXiv:1109.1981].

[844] A. Nicola, A. Refregier, A. Amara and A. Paranjape, Phys. Rev. D**90**, 063515 (2014), [arXiv:1405.3660].

[845] S. Salazar-Albornoz *et al.*, Mon. Not. R. Astron. Soc.**468**, 2938 (2017), [arXiv:1607.03144].

[846] A. J. S. Hamilton and M. Culhane, Mon. Not. R. Astron. Soc.**278**, 73 (1996), [arXiv:astro-ph/9507021].

[847] A. A. de Laix and G. D. Starkman, ArXiv Astrophysics e-prints (1998), [arXiv:astro-ph/9802270].

[848] W. Hu, Phys. Rev. D **62**, 043007 (2000).

[849] W. J. Percival *et al.*, Astrophys. J.**657**, 645 (2007), [arXiv:astro-ph/0608636].

[850] C. Blake *et al.*, Mon. Not. R. Astron. Soc.**406**, 803 (2010), [arXiv:1003.5721].

[851] B. R. Granett *et al.*, Mon. Not. R. Astron. Soc.**421**, 251 (2012), [arXiv:1112.0008].

[852] F. Beutler *et al.*, Mon. Not. R. Astron. Soc.(2016), [arXiv:1607.03149].

[853] P. Sarkar, J. Yadav, B. Pandey and S. Bharadwaj, Mon. Not. R. Astron. Soc.**399**, L128 (2009), [arXiv:0906.3431].

[854] M. I. Scrimgeour *et al.*, Mon. Not. R. Astron. Soc.**425**, 116 (2012), [arXiv:1205.6812].

[855] J. Carron, Astrophys. J.**738**, 86 (2011), [arXiv:1105.4467].

[856] D. Jeong, F. Schmidt and E. Sefusatti, Phys. Rev. D**83**, 123005 (2011), [arXiv:1104.0926].

[857] C. S. Withers, Journal of Statistics Applications & Probability Letters **47**, 165 (2000).

[858] R. Takahashi, Progress of Theoretical Physics **120**, 549 (2008).

[859] N. Makino, M. Sasaki and Y. Suto, Phys. Rev. D**46**, 585 (1992).

[860] P. Catelan and L. Moscardini, Astrophys. J.**426**, 14 (1994), [arXiv:astro-ph/9308002].

[861] E. Gaztanaga and C. M. Baugh, Mon. Not. R. Astron. Soc.**273**, L1 (1995), [arXiv:astro-ph/9409062].

[862] R. Scoccimarro and J. A. Frieman, Astrophys. J.**473**, 620 (1996), [arXiv:astro-ph/9602070].

[863] A. Taruya, T. Nishimichi and F. Bernardeau, Phys. Rev. D**87**, 083509 (2013), [arXiv:1301.3624].

[864] Z. Vlah, U. Seljak, M. Yat Chu and Y. Feng, JCAP **3**, 057 (2016), [arXiv:1509.02120].

[865] M. M. Ivanov and S. Sibiryakov, JCAP **1807**, 053 (2018), [arXiv:1804.05080].

[866] L. Fonseca de la Bella, D. Regan, D. Seery and S. Hotchkiss, JCAP **11**, 039 (2017), [arXiv:1704.05309].

[867] Z. Ding *et al.*, Mon. Not. Roy. Astron. Soc. **479**, 1021 (2018), [arXiv:1708.01297].

[868] E. Bertschinger and A. J. S. Hamilton, Astrophys. J.**435**, 1 (1994), [arXiv:astro-ph/9403016].

[869] C. Lacey and S. Cole, Mon. Not. R. Astron. Soc.**271**, 676 (1994), [arXiv:astro-ph/9402069].

[870] Z. Lukić, D. Reed, S. Habib and K. Heitmann, Astrophys. J.**692**, 217 (2009), [arXiv:0803.3624].