Alma Mater Studiorum · Università di Bologna

SCHOOL OF SCIENCE
Department of Physics and Astronomy
Master Degree Programme in Astrophysics and Cosmology

A revisited Correction to the Halo Mass Function for local-type Primordial non-Gaussianity

Graduation Thesis

Supervisor:
Prof. Federico Marulli
Co-Supervisors:
Dr. Sofia Contarini
Dr. Ariel Sánchez
Prof. Marco Baldi

Submitted by:
Luca Fiorino

Academic Year 2023–2024

"I have no special talent.
I am only passionately curious."
- Albert Einstein

Contents

Abstract

The theory of inflation assumes that the early Universe underwent a period of rapid expansion, giving rise to the initial density perturbations that eventually led to the formation of the large-scale structures that we observe today (Guth & Pi, 1982). Among the most promising observational signatures for inflation are primordial non-Gaussianities (PNG), as nearly all inflationary models predict their existence. The term PNG refers to deviations from a Gaussian in the distribution of primordial perturbations, typically quantified by the parameter $f_{\rm NL}^{\rm X}$, where "X" indicates different shapes corresponding to various inflationary scenarios, each imprinting distinct features on the primordial bispectrum (Takahashi, 2014). Local-type PNG — quantified by the parameter $f_{\rm NL}^{\rm loc}$ — has attracted particular attention because single-field slow-roll inflation models predict very small values for it, and detecting a significantly larger value would rule out these models.

Local-type PNG is expected to influence the abundance of dark matter halos (or, alternatively, galaxy clusters), and various theoretical models have been developed over the past decades to predict its effect on the halo mass function (see e.g. Matarrese et al., 2000; LoVerde et al., 2008; D'Amico et al., 2011a). Furthermore, local-type PNG modifies the mass tracer bias, making it scale dependent and causing deviations from the Gaussian case, particularly on large scales (Dalal et al., 2008). Both of these statistical properties of dark matter halos offer a crucial observational tool for detecting deviations from the standard Λ *cold dark matter* (ΛCDM) model.

In this work, we focus in particular on the impact of local-type PNG on halo number counts. We measure the halo mass function in a large set of cosmological N-body simulations with different values of the $f_{\rm NL}^{\rm loc}$ parameter and show that current theoretical models fail to adequately describe the non-Gaussian halo mass function, with discrepancies from the measurements increasing as the density threshold for halo identification rises. We explain how these discrepancies are related to variations in the density profile of dark matter halos, finding that the internal profile steepness (i.e. the compactness) of halos depends on the value of $f_{\rm NL}^{\rm loc}$. Specifically, we demonstrate that positive (negative) values of $f_{\rm NL}^{\rm loc}$ lead to halos with higher (lower) internal densities compared to the standard ΛCDM model, affecting the mass value associated to dark matter halos for different thresholds.

To address these discrepancies, we introduce a correction factor κ that modifies the linear density threshold for collapse, $\delta_{\rm c}(z)$, according to the density threshold used to identify halos, $\Delta_{\rm b}(z)$. The latter is defined with respect to the background density of the Universe, but other definitions can be applied using specific conversions (see e.g. Tinker et al., 2008). We model the function $\kappa(\Delta_{\rm b})$ using a second-degree polynomial, performing a Bayesian analysis to determine the polynomial

coefficients. Notably, we verify that these coefficients are independent of both the sign and magnitude of $f_{\rm NL}^{\rm loc}$.

Finally, we demonstrate that this re-parametrization is not affected by resolution effects and avoids biased constraints on $f_{\rm NL}^{\rm loc}$, correcting for significant systematic errors, particularly for halos identified with high overdensity thresholds. This improvement is crucial for deriving accurate cosmological constraints from the non-Gaussian halo mass function using real data, as surveys such as *Euclid* (Euclid Collaboration et al., 2024) and the *Dark Energy Spectroscopic Instrument* (DESI Collaboration et al., 2016) may provide galaxy cluster catalogs with different mass definitions depending on the observation strategy.

All the research conducted for this Thesis has been summarized in Fiorino et al. (2024), a scientific paper available on the arXiv platform, and currently under review at the *Journal of Cosmology and Astroparticle Physics*.

Chapter 1

Cosmological Framework

In the opening chapter, we lay the groundwork for understanding the various topics discussed throughout this Thesis by exploring the cosmological framework. Specifically, we start with a brief overview of the Theory of General Relativity (GR), the foundation upon which modern cosmology is constructed. We then introduce the Friedmann-Lemaître-Robertson-Walker (FLRW) metric, which is crucial for describing the spacetime of a Universe that is homogeneous and isotropic. Following this, we discuss the Hubble-Lemaître law, we define the cosmological redshift, and we derive the Friedmann equations through the solution of Einstein's field equations. Finally, we outline the primary features of the ΛCDM, the standard cosmological model currently in use, and address its associated tensions.

1.1 Elements of General Relativity

In the context of Cosmology, the force of gravity plays a pivotal role in shaping the interactions between the various constituents of the Universe. The theory of GR, introduced by Albert Einstein (Einstein, 1915), stands as the cornerstone for understanding these gravitational interactions. This theory, which extends the principles of Special Relativity to incorporate gravity, posits that spacetime is warped by the presence of mass and energy. Consequently, gravity is not perceived as a force in the traditional sense but as the manifestation of the curvature of spacetime. Mathematically, the spacetime is described as a 4-dimensional differentiable manifold in which a point is promoted to an *event* characterized by four coordinates (one time-like and three space-like). The geometrical properties of the spacetime are delineated by the *metric tensor* $g_{\mu\nu}$, which formally defines the distance $\mathrm{d}s^2$ between two infinitesimally close events through the following relation:

$$\mathrm{d}s^2 = \sum_{\mu,\,\nu=0}^{3} g_{\mu\nu}\mathrm{d}x^\mu\mathrm{d}x^\nu \equiv g_{\mu\nu}\mathrm{d}x^\mu\mathrm{d}x^\nu\,, \tag{1.1}$$

where $x^\mu = (ct, x, y, z)$ and $x^\nu = x^\mu + \mathrm{d}x^\mu = (c(t + \mathrm{d}t), x + \mathrm{d}x, y + \mathrm{d}y, z + \mathrm{d}z)$, with c representing the speed of light in vacuum. The first index pertains to the time-like coordinate ($x^0 = ct$) while the last three correspond to spatial coordinates. Exploiting the symmetry of the metric tensor and setting henceforth $c = 1$, the infinitesimal displacement can be expanded as:

$$\mathrm{d}s^2 = g_{00}\mathrm{d}t^2 + 2g_{0i}\mathrm{d}t\mathrm{d}x^i + g_{ij}\mathrm{d}x^i\mathrm{d}x^j\,, \tag{1.2}$$

where $g_{00}\mathrm{d}t^2$ is the time-time component, $g_{ij}\mathrm{d}x^i\mathrm{d}x^j$ are the space-space components and $2g_{0i}\mathrm{d}t\mathrm{d}x^i$ the mixed ones.

In the context of GR, we define a *geodesic* as the shortest curve connecting two events of spacetime. It generalizes the concept of straight lines in the well-known flat Minkowskian space. The equation describing a geodesic can be derived by minimising the length $L[\gamma]$ of a curve γ defined on the spacetime

$$\delta L[\gamma] = \delta \int \mathrm{d}s = \delta \int \sqrt{g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu}\mathrm{d}\lambda = 0\,, \tag{1.3}$$

where λ is the scalar (*e.g.* the proper time) which parameterises the curve γ, and $\dot{x}^\mu$ denotes the derivative with respect to λ. This path is followed by any particle in the absence of any force apart from gravity, and it is obtained from Eq. 1.3 by solving the geodesic equation:

$$\frac{\mathrm{d}^2x^\mu}{\mathrm{d}\lambda^2} + \Gamma^\mu_{\alpha\beta}\frac{\mathrm{d}x^\alpha}{\mathrm{d}\lambda}\frac{\mathrm{d}x^\beta}{\mathrm{d}\lambda} = 0\,. \tag{1.4}$$

The symbol $\Gamma^\mu_{\alpha\beta}$ represents the Christoffel's symbols, which are fully determined by the metric tensor through the expression

$$\Gamma^\mu_{\alpha\beta} = \frac{g^{\mu\nu}}{2}\left[\frac{\partial g_{\alpha\nu}}{\partial x^\beta} + \frac{\partial g_{\beta\nu}}{\partial x^\alpha} - \frac{\partial g_{\alpha\beta}}{\partial x^\nu}\right]\,, \tag{1.5}$$

where $g^{\mu\nu}$ is the inverse of $g_{\mu\nu}$, so that $g^{\mu\nu}g_{\alpha\nu} = \delta^\mu_\alpha$ is the Kronecker delta, which is equal to unity if $\mu = \alpha$ and 0 otherwise.

At this juncture, we present another fundamental aspect of GR: the stress or energy-momentum tensor $T_{\mu\nu}$. In essence, it characterizes the density, the flux energy, and the momentum of a generic fluid pervading the Universe. Assuming isotropy, the tensor can be succinctly expressed in terms of pressure (p) and energy density (ρ) as follows:

$$T^\mu_\nu = \begin{pmatrix} -\rho & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix}\,. \tag{1.6}$$

Within GR, the laws of energy and momentum conservation can be derived by enforcing the nullity of the covariant derivative

$$\nabla_\mu T^\mu_\nu = \frac{\partial T^\mu_\nu}{\partial x^\mu} + \Gamma^\mu_{\alpha\mu}T^\alpha_\nu - \Gamma^\alpha_{\nu\mu}T^\mu_\alpha = 0\,. \tag{1.7}$$

Building upon the Christoffel's symbols, we proceed to define the Riemann curvature tensor

$$R^\lambda_{\mu\nu\rho} = \frac{\partial\Gamma^\lambda_{\nu\rho}}{\partial x^\mu} - \frac{\partial\Gamma^\lambda_{\mu\rho}}{\partial x^\nu} + \Gamma^\lambda_{\mu\sigma}\Gamma^\sigma_{\nu\rho} - \Gamma^\lambda_{\nu\sigma}\Gamma^\sigma_{\mu\rho}\,, \tag{1.8}$$

which inherently depends on the first and second derivatives of the metric tensor. Subsequently, the Riemann tensor is contracted to obtain the Ricci tensor $R_{\mu\nu}$ and the Ricci scalar R:

$$R_{\mu\nu} = R^{\alpha}_{\mu\alpha\nu}\,, \tag{1.9}$$

$$R \equiv R^{\mu}_{\mu} = g^{\mu\nu} R_{\mu\nu}\,. \tag{1.10}$$

From these definitions Einstein formulated his own tensor

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R\,. \tag{1.11}$$

We now possess all the requisite components to establish a connection between the metric and the constituents of the Universe. This relationship is encapsulated within the Einstein field equations, which correlate the Einstein tensor describing the geometry to the energy-momentum tensor. These set of equations can be summarized as the following tensor equality

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu}\,, \tag{1.12}$$

where G is the Newtonian gravitational constant and Λ is the cosmological constant. The term $8\pi G$ ensures the recovery of Newtonian gravitational theory in the weak gravitation field limit. These equations hold paramount significance as they dictate both the evolution of the Universe and the formation of structures within it.

Regarding the cosmological constant, there is no impediment to relocating the Λ term to the right-hand side of Eq. 1.12, as it only involves the metric tensor. We can formally define the cosmological constant contribution to the energy-momentum tensor:

$$T_{(\Lambda)}{}^{\mu}_{\nu} = -\frac{\Lambda}{8\pi G}\delta^{\mu}_{\nu} = \begin{pmatrix} -\rho_\Lambda & 0 & 0 & 0 \\ 0 & -\rho_\Lambda & 0 & 0 \\ 0 & 0 & -\rho_\Lambda & 0 \\ 0 & 0 & 0 & -\rho_\Lambda \end{pmatrix}, \quad \text{where} \quad \rho_\Lambda = \frac{\Lambda}{8\pi G} \tag{1.13}$$

is the effective energy density of the cosmological constant. The concept of a cosmological constant as an additional energy component, referred to as *dark energy* (DE), gained prominence because of its repulsive effect. This concept aims to elucidate the current accelerated expansion witnessed in the Universe, as evidenced by observations of the flux of distant type Ia supernovae (SNIa) (Riess et al., 1998; Perlmutter et al., 1999).

1.2 Friedmann-Lemaître-Robertson-Walker Metric

Most of the prevailing modern cosmological models are constructed upon the *cosmological principle* (CP), which posits the homogeneity and isotropy of the Universe on sufficiently large scale[1]. Additionally, there exists compelling evidence supporting

[1]Nowadays, hundreds of Mpc, where 1 Mpc $\simeq 3.09 \cdot 10^{24}$ cm.

the notion that the Universe is undergoing expansion. This indicates that in the early stages of its history, the distance between us and distant galaxies was smaller than it is at present. The effect of this expansion is taken into account by introducing the scale factor $a(t)$, whose present value is set to unity by convention. Assuming the validity of the CP, the metric tensor describing the expanding Universe takes the form:

$$g_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & \dfrac{a^2(t)}{1-\kappa r^2} & 0 & 0 \\ 0 & 0 & a^2(t)r^2 & 0 \\ 0 & 0 & 0 & a^2(t)r^2\sin^2\theta \end{pmatrix}, \tag{1.14}$$

where κ is the adimensional curvature parameter. Any event of the Universe is described by three spatial polar coordinates (r, θ, φ), known as *comoving coordinates*, and one temporal coordinate t, referred as *proper* or *cosmic time*, which are defined in a reference system at rest with the Universe expansion. By substituting the expression for the metric tensor into Eq. 1.2, we derive the general form of the *Friedmann-Lemaître-Robertson-Walker* (FLRW) metric:

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2(t)\left[\frac{\mathrm{d}r^2}{1-\kappa r^2} + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\varphi^2\right)\right]. \tag{1.15}$$

To comprehend the history of the universe, it is imperative to ascertain the evolution of the scale factor a with cosmic time t. This derivation stems from the Einstein's field equations, provided the energy momentum tensor and the geometry of the universe are known. In the standard simply connected topology, the latter is determined by the value of the curvature parameter:

- $\kappa = -1 \rightarrow$ Hyperbolic geometry: space is open and infinite;
- $\kappa = 0 \rightarrow$ Flat geometry: space is Euclidean and infinite;
- $\kappa = +1 \rightarrow$ Spherical geometry: space is closed and unbounded.

1.3 The Hubble-Lemaître Law and the Cosmological Redshift

Measuring distances in an expanding Universe presents challenges. The *comoving distance* remains fixed as the Universe expands because it is tied solely to the coordinate grid. Consider the comoving distance between a light source and us: in a small time interval $\mathrm{d}t$, light travels a comoving distance $\mathrm{d}\chi = \mathrm{d}t/a(t)$, resulting in the total comoving distance traveled by light given by

$$\chi = \int_0^r \frac{\mathrm{d}r'}{\sqrt{1-\kappa r'^2}} = \int_t^{t_0} \frac{\mathrm{d}t'}{a(t')}, \tag{1.16}$$

where we have considered the FLRW metric and the fact that photons move along null geodesics ($\mathrm{d}s^2 = 0$). On the other hand, the *proper distance* $d(t)$ is defined as the physical distance between two points, which varies with time due to the scale factor $a(t)$, and it is expressed as

$$d(t) = a(t)\chi = a(t)\int_0^r \frac{\mathrm{d}r'}{\sqrt{1-\kappa r'^2}}\,. \tag{1.17}$$

From this definition, it is evident that the comoving and proper distances coincide today ($a(t_0) = 1$, by convention).

The expansion of the Universe leads to a continuous increase in physical separation between two arbitrary points in space. The *radial velocity* between these points can be computed as the derivative of the proper distance $d(t)$ with respect to cosmic time

$$v_r = \frac{\mathrm{d}}{\mathrm{d}t}d(t) = \frac{\mathrm{d}}{\mathrm{d}t}[a(t)\chi] = \dot{a}(t)\chi + a(t)\dot{\chi}\,. \tag{1.18}$$

In the absence of any comoving motion, the second term vanishes, simplifying the radial velocity to

$$v_r = \dot{a}(t)\chi = \frac{\dot{a}(t)}{a(t)}d = H(t)d\,, \tag{1.19}$$

which represents the well-known *Hubble-Lemaître law*. The function $H(t)$, known as the *Hubble parameter*, is defined as $H(t) = \dot{a}(t)/a(t)$ and is suppose to have the same value across the Universe at any given cosmic time, provided sufficiently large scales are considered. Its value at the present time $H(t_0) = H_0$ is termed the *Hubble constant* and describes the expansion rate of the Universe. Conventionally, the Hubble constant is written in terms of a dimensionless parameter h as follows

$$H_0 \equiv 100h \text{ km s}^{-1}\text{ Mpc}^{-1}\,. \tag{1.20}$$

Current measurements yield $h \simeq 0.7$ but the determination of the Hubble constant remains a subject of debate within the scientific community due to discrepancies arising from different probes. For example: $H_0 = 67.4 \pm 0.5$ km s^{-1} Mpc^{-1} from the CMB angular spectrum (Planck Collaboration et al., 2020a), $H_0 = 67.7^{+4.3}_{-0.42}$ km s^{-1} Mpc^{-1} from the analysis of gravitational waves (Mukherjee et al., 2020) and $H_0 = 74.03 \pm 1.42$ km s^{-1} Mpc^{-1} by using distance ladders as Cepheids or SNIa (Riess et al., 2019). Given that H_0 is expressed in units of s^{-1}, it can offer a rough estimate of the age of the Universe, assuming a constant expansion rate.

The global motion of objects in the Universe relative to each other is referred to as the *Hubble Flow*. Its main implication is the reddening of observed spectra of astrophysical objects, known as the *cosmological redshift*. This reddening effect, similar to the *Doppler effect*, results in observed wavelengths being shifted towards longer wavelengths. Let λ_{em} represent the wavelength of the light emitted by a source in its reference system, and λ_{obs} denote the shifted wavelength of light received by an observer. The relative difference between the two electromagnetic radiations can be expressed as

$$z \equiv \frac{\lambda_{\mathrm{obs}} - \lambda_{\mathrm{em}}}{\lambda_{\mathrm{em}}}\,. \tag{1.21}$$

In principle, this value can be less than zero if the source is approaching (*blueshift*), or greater than zero if the source is receding (*redshift*).

Let us consider a scenario in which a source located at $(r, 0, 0)$ emits a light signal at $t = t_{\mathrm{em}}$, which is then received at some time $t = t_{\mathrm{obs}}$ by an observer located at the origin of the reference system. By considering the FLRW metric we have

$$\int_{t_{\mathrm{em}}}^{t_{\mathrm{obs}}} \frac{\mathrm{d}t}{a(t)} = \int_0^r \frac{\mathrm{d}r'}{\sqrt{1 - \kappa r'^2}} \,. \tag{1.22}$$

Now, suppose that another signal is emitted at $t = t_{\mathrm{em}} + \delta t_{\mathrm{em}}$ and observed at $t = t_{\mathrm{obs}} + \delta t_{\mathrm{obs}}$. Since the right-hand side of Eq. 1.22 is independent on the expansion of the Universe, the difference in terms of photon paths is determined solely by time

$$\int_{t_{\mathrm{em}}+\delta t_{\mathrm{em}}}^{t_{\mathrm{obs}}+\delta t_{\mathrm{obs}}} \frac{\mathrm{d}t}{a(t)} = \int_0^r \frac{\mathrm{d}r'}{\sqrt{1 - \kappa r'^2}} = \int_{t_{\mathrm{em}}}^{t_{\mathrm{obs}}} \frac{\mathrm{d}t}{a(t)} \,. \tag{1.23}$$

Therefore, if the time intervals δt_{em} and δt_{obs} are sufficiently small, $a(t)$ can be considered constant, and the equivalence reported in Eq. 1.23 leads to

$$\frac{\delta t_{\mathrm{obs}}}{a(t_{\mathrm{obs}})} = \frac{\delta t_{\mathrm{em}}}{a(t_{\mathrm{em}})} \,. \tag{1.24}$$

Recalling that $\delta t = 1/\nu$, the relationship between wavelength and frequency, and the definition of z in Eq. 1.21, we find

$$1 + z = \frac{a(t_{\mathrm{obs}})}{a(t_{\mathrm{em}})} = \frac{1}{a(t_{\mathrm{em}})} \,, \tag{1.25}$$

assuming that the light signal is observed today. This relation enables the use of redshift measurements (utilizing spectroscopic or photometric techniques) to estimate the distance of extragalactic sources. However, it is important to note that the wavelength of light could also be influenced by interaction with other particles and other relativistic effects.

1.4 The Friedmann Equations

Assuming the validity of the CP and adopting an energy-momentum tensor as outlined in Eq. 1.6, we can utilize the FLRW metric to resolve the Einstein's field equations. This yields a pair of equations derived from the time-time and space-space components of the field equations, known as the *first* and *second Friedmann equation* respectively. These equations elucidate the temporal evolution of the scale factor $a(t)$ and are expressed as follows:

$$\left(\frac{\dot{a}}{a}\right)^2 + \frac{\kappa}{a^2} = \frac{8\pi G}{3}\rho \,, \tag{1.26}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + 3p\right) \,. \tag{1.27}$$

Here, the energy density ρ and the pressure p must encompass all the diverse components of the Universe, such as non-relativistic matter and radiation (photons and relativistic particles). Referring to Eq. 1.26, it becomes evident that to reconstruct the evolution of $a(t)$, we require an understanding of how the density ρ changes over cosmic time. This comprehension can be attained by examining the conservation law of energy and momentum (Eq. 1.7), considering $\nu = 0$ and the FLRW metric:

$$\frac{\partial T^{\mu}_{0}}{\partial x^{\mu}} + \Gamma^{\mu}_{\alpha\mu} T^{\alpha}_{0} - \Gamma^{\alpha}_{0\mu} T^{\mu}_{\alpha} = 0 \quad \Longrightarrow \quad \frac{\partial \rho}{\partial t} + 3\frac{\dot{a}}{a}(\rho + p) = 0\,. \tag{1.28}$$

Non-relativistic matter has effectively zero pressure, implying that the energy density of matter follows $\rho_{\rm m} \propto a^{-3}$. This behaviour arises because particle mass remains constant while the number density scales inversely with volume. In contrast, radiation is characterized by $p_{\rm r} = \rho_{\rm r}/3$, which implies that the energy density of radiation $\rho_{\rm r} \propto a^{-4}$. This scaling accounts for the decrease in energy per particle as the Universe expands. We can summarize the cases of matter and radiation, and generalize the evolution results to other constituents, by defining the *equation of state* (EoS) parameter $w_{\rm s}$

$$w_{\rm s} \equiv \frac{p_{\rm s}}{\rho_{\rm s}}\,, \tag{1.29}$$

where the subscript "s" stands for any constituent of the Universe. For matter, $w_{\rm m} = 0$, for radiation, $w_{\rm r} = 1/3$, and for the cosmological constant, from Eq. 1.13, $w_\Lambda = -1$. However, the EoS parameter is not necessarily constant over cosmic time. By integrating Eq. 1.28, the temporal evolution of any component of the Universe with a time-varying EoS parameter can be described as follows

$$\rho_{\rm s} \propto \exp\left\{-3\int^{a} \frac{{\rm d}a'}{a'}\left[1 + w_{\rm s}(a')\right]\right\} \overset{w_{\rm s}=\text{const}}{\propto} a^{-3(1+w_{\rm s})} \propto (1+z)^{3(1+w_{\rm s})}\,. \tag{1.30}$$

The second proportionality holds if $w_{\rm s}$ is independent on time. As a result of the preceding relation, it can be affirmed that various components have exerted dominance over each other across the cosmic epochs.

Starting from the first Friedmann equation, we can rearrange the terms to derive an equation for the curvature parameter

$$\frac{\kappa}{a^2} = H^2(t)\left(\frac{8\pi G\rho(t)}{3H^2(t)} - 1\right) = H^2(t)\left(\frac{\rho(t)}{\rho_{\rm crit}(t)} - 1\right)\,, \tag{1.31}$$

where we introduce the *critical density* parameter as

$$\rho_{\rm crit}(t) \equiv \frac{3H^2(t)}{8\pi G}\,. \tag{1.32}$$

From Eq. 1.31, it is apparent that the critical density represents the density required for a flat geometry ($\kappa = 0$). If the Universe's density is less than the critical density ($\rho < \rho_{\rm crit}$), we observe an open geometry characterized by perpetual expansion. Conversely, if the density exceeds the critical density ($\rho > \rho_{\rm crit}$), a closed geometry ensues, leading to expansion followed by contraction. The present-day critical density depends on the Hubble constant H_0 and its value is given by

$$\rho_{0,\rm crit} \equiv \rho_{\rm crit}(t_0) = \frac{3H_0^2}{8\pi G} \simeq 1.9 \times 10^{-29}\ h^2\ {\rm g\ cm^{-3}}\,. \tag{1.33}$$

By virtue of the definition of the critical density, we can introduce for each component a dimensionless parameter Ω, denoted as the *density parameter*, and defined as follow:

$$\Omega_{\rm s}(t) \equiv \frac{\rho_{\rm s}(t)}{\rho_{\rm crit}(t)}\,. \tag{1.34}$$

According to this definition, the *total density parameter* is simply the summation of all individual density parameters

$$\Omega_{\rm tot}(t) \equiv \sum_{\rm s} \Omega_{\rm s}(t)\,. \tag{1.35}$$

Hence, in a scenario of a flat Universe, it is straightforward to deduce that $\Omega_{\rm tot} = 1$, whereas in the case of an open and closed Universe. $\Omega_{\rm tot} < 1$ and $\Omega_{\rm tot} > 1$, respectively.

With the introduction of the newly defined $\Omega_{\rm tot}$ we can reformulate Eq. 1.31 in the following manner

$$\Omega_{\rm tot}(t) - 1 = \frac{\kappa}{a^2(t)H^2(t)}\,. \tag{1.36}$$

The sign of the right-hand side of this equation is entirely determined by the value of κ and remains invariant throughout cosmic time. Consequently, the sign of the left-hand side also remains unchanged. Therefore, a Universe described by the Friedmann equations cannot alter its geometry as it evolves.

The first Friedmann equation can be conveniently reformulated in terms of H, Ω and z, which offer a more representative characterization of the observable Universe. By utilizing the definitions of the density parameter (Eq. 1.34) and redshift (Eq. 1.21), the expression can be rewritten as follows

$$H^2(z) = H_0^2(1+z)^2\left(\Omega_{0,\kappa} + \sum_{\rm s} \Omega_{0,\rm s}(1+z)^{1+3w_{\rm s}}\right) \equiv H_0^2 E^2(z)\,. \tag{1.37}$$

Here, $\Omega_{0,\rm s}$ denotes the present value of the density parameter, and $\Omega_{0,\kappa} \equiv 1 - \sum_{\rm s} \Omega_{0,\rm s}$ represents the so-called *curvature density parameter*.

Now, let us examine the second Friedmann equation for a Universe consisting of a single component characterized by an EoS parameter $w_{\rm s}$. The equation reduces to

$$\ddot{a} = -\frac{4\pi G}{3}\rho_{\rm s}(1+3w_{\rm s})a\,. \tag{1.38}$$

From this relation, it is evident that for typical cosmological components such as matter ($w_{\rm m} = 0$) and radiation ($w_{\rm r} = 1/3$), the corresponding mono-component Universe undergoes decelerated expansion, as $\ddot{a} < 0$. Additionally, due to the expansion, the scale factor $a(t)$ increases monotonically in time. Consequently, going back in time, there exists a moment at which a approaches zero, marking a point where temperature, density, and expansion rate diverge:

$$\lim_{t\to 0} \rho_{\rm s}(a) \propto \lim_{t\to 0} a(t)^{-3(1+w_{\rm s})} = \infty\,. \tag{1.39}$$

This event is called *Big Bang* (BB) and is a characteristics of all the cosmological models assuming a single component with $-1/3 < w_{\rm s} < 1$. However, the precise

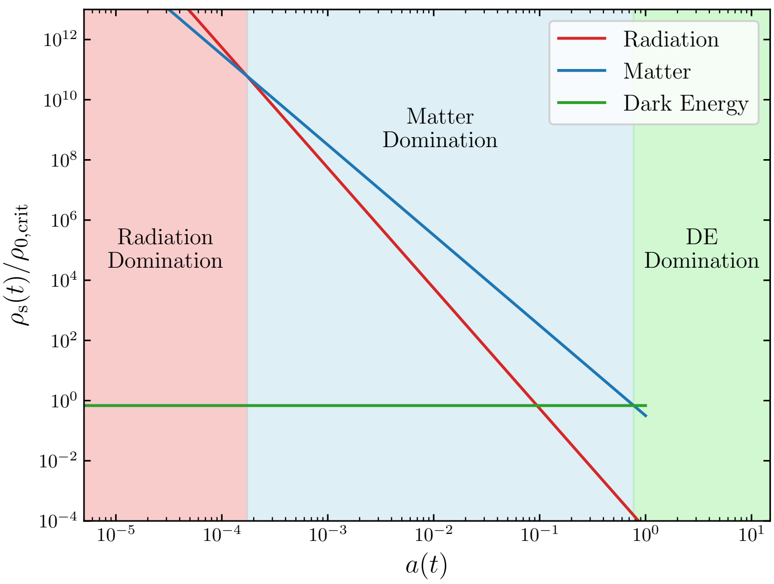

Figure 1.1: Trends with time of the three main Universe's components: radiation, matter and DE. As the Universe expands, different components start to predominate on the others because of the relative change in energy density.

physical conditions at the time of BB remain unknown because quantum corrections to gravity must be considered for times $t < t_{\rm P}$, where $t_{\rm P} \simeq 10^{-43}$ s denotes the *Planck time*. Currently, a universally accepted treatment of quantum gravity has not yet been established.

1.5 Single-Component Universe

This model is based on the hypothesis that the Universe is permeated by a single component and has a flat geometry ($\kappa = 0$ or, equivalently, $\Omega_{\rm tot} = \Omega = 1$). If this component is matter ($w_{\rm m} = 0$) the model reduces to the well-known *Einstein-de Sitter model* (EdS). With these assumptions, Eq. 1.37 simplifies to:

$$H(z) = H_0(1+z)^{\frac{3(1+w)}{2}} \,. \tag{1.40}$$

The densities of the various components of the Universe (matter, radiation and cosmological constant) evolve according to Eq. 1.30. This equation reveals that each component undergoes a period of dominance during specific cosmic epochs. Consequently, it is plausible to consider our Universe as predominantly consisting of a single component at any given time. Hence, it is common-use to delineate the history of the Universe into epochs based on the prevailing dominant component, as shown in Fig. 1.1. In particular, during the early times, radiation emerges as the dominant component, marking the *radiation-dominated era*, while at later times, the matter component assumes greater significance, defining the *matter-dominated era*. In particular, the density of DE remains constant throughout cosmic time and

Generic w_s	Matter $w_m = 0$	Radiation $w_r = 1/3$
$a(t) = \left(\frac{t}{t_0}\right)^{\frac{2}{3(1+w_s)}}$	$a(t) = \left(\frac{t}{t_0}\right)^{\frac{2}{3}}$	$a(t) = \left(\frac{t}{t_0}\right)^{\frac{1}{2}}$
$t = t_0\,(1+z)^{-\frac{3(1+w_s)}{2}}$	$t = t_0\,(1+z)^{-\frac{3}{2}}$	$t = t_0\,(1+z)^{-2}$
$H(t) = \frac{2}{3(1+w_s)t}$	$H(t) = \frac{2}{3t}$	$H(t) = \frac{1}{2t}$
$t_0 = \frac{2}{3(1+w_s)H_0}$	$t_0 = \frac{2}{3H_0}$	$t_0 = \frac{1}{2H_0}$
$\rho_s = \frac{1}{6\pi G(1+w_s)^2 t^2}$	$\rho_m = \frac{1}{6\pi G t^2}$	$\rho_r = \frac{3}{32\pi G t^2}$

Table 1.1: Dependencies obtained for the single-component Universe in the case of a a generic component (*first column*), for a matter-dominated Universe (*second column*), and for a radiation-dominated Universe (*third column*).

only becomes prominent at very recent times, delineating the *DE-dominated era*. It is noteworthy to mention that the single-component approximation accurately applies only during periods far from the moments of equivalence, i.e. the transitions at which one component starts to prevail on the others. Throughout the history of the universe, we can delineate two equivalences:

- *matter-radiation equivalence* → By definition, it is the moment at which the density of matter and radiation are equal. From this definition, we can determine the redshift z_{eq} of matter-radiation equivalence as follows

$$\rho_{0,m}(1+z_{eq})^3 = \rho_{0,r}(1+z_{eq})^4 \quad \Longrightarrow \quad z_{eq} = \frac{\rho_{0,m}}{\rho_{0,r}} - 1 \simeq 3\times 10^4\,; \tag{1.41}$$

- *DE-matter equivalence* → By definition, it is the moment at which the density of matter and DE are the same. Therefore, we can determine the redshift $z_{eq,\Lambda}$ of DE-matter equivalence as follows

$$\rho_{0,m}(1+z_{eq,\Lambda})^3 = \rho_\Lambda \quad \Longrightarrow \quad z_{eq,\Lambda} = \left(\frac{\rho_\Lambda}{\rho_{0,m}}\right)^{1/3} - 1 \simeq 0.3\,. \tag{1.42}$$

Finally, in Table 1.1 we present a compilation of useful relationships derived under the assumption of the single-component model, describing the behavior of key quantities characterizing this type of Universe. These dependencies are initially expressed for a generic component with the EoS parameter w_s, and subsequently computed for both the matter-dominated epoch ($w_m = 0$) and the radiation-dominated epoch ($w_r = 1/3$).

1.6 The Standard Cosmological Model

As highlighted in Sect. 1.5, the contemporary Universe is primarily governed by DE, yet matter maintains a significant role in terms of its contribution to energy density. Thus, relying solely on a single-model component proves insufficient for accurately depicting contemporary phenomena. Since the early 21^{st} century, the ΛCDM model has become widely recognized as the *standard cosmological model*, serving as a fundamental framework for comprehending our Universe. This designation stems from its extensive support by observational evidence and its ability to provide a solid foundation for understanding the formation of cosmic structures. It portrays the Universe as nearly flat, adhering to the CP, and its evolution is governed by the Friedmann Equations and, therefore, by GR.

The evolution of the Universe, as described by the ΛCDM model, is intricately tied to its temperature. Initially, the temperature was significantly higher than it is today, with photons currently maintaining a temperature of $T = 2.7255 \pm 0.0006$ K (Planck Collaboration et al., 2020a), constituting the *cosmic microwave background* (CMB) radiation. The CMB represents a relic from the last scattering surface, occurring approximately 3.8×10^5 years after the BB ($z_{\rm LS} \simeq 1100$). Preceding this scattering event, the Universe was populated by a hot plasma of fully ionized protons and electrons, wherein electromagnetic radiation interacted continuously with baryonic matter, maintaining the Universe opaque and in a condition of thermal equilibrium. However, as the Universe expanded, around $z \simeq 1500$, the temperature dropped sufficiently for protons and electrons to recombine, leading to the decoupling of the two cosmological components and enabling photons to propagate freely. The redshift $z_{\rm LS} \simeq 1100$ of the last scattering marks the peak probability for a photon to undergo its final scattering by the primordial plasma. This moment heralds the emergence of the CMB radiation, observable today, though heavily redshifted due to the Universe's expansion.

The ΛCDM model derives its name from the two principal constituents characterizing the current state of the Universe: DE, linked to a cosmological constant and possessing a density parameter $\Omega_{0,\Lambda} \simeq 0.7$, and CDM, a non-collisional, non-relativistic matter element with a density parameter $\Omega_{0,\rm cdm} \simeq 0.25$. Additionally, there exist a minor contribution from baryonic matter (i.e. composed by baryons) with $\Omega_{0,\rm b} \simeq 0.05$, and from radiation with $\Omega_{0,\rm r} \simeq 10^{-5}$. These energy density values align with the condition of flatness, as evidenced by $\Omega_{0,\rm tot} \simeq 1$.

A robust characterisation of this model involves the definition of six fundamental parameters (Planck Collaboration et al., 2020a):

- $\Omega_{0,\rm m} \rightarrow$ Total matter density parameter ($\Omega_{0,\rm m} h^2 = 0.143 \pm 0.001$);
- $\Omega_{0,\rm b} \rightarrow$ Baryonic matter density parameter ($\Omega_{0,\rm b} h^2 = 0.0224 \pm 0.0001$);
- $H_0 \rightarrow$ Hubble constant ($H_0 = 67.4 \pm 0.5$ km s^{-1} Mpc^{-1});
- $A_{\rm s} \rightarrow$ Primordial power spectrum amplitude ($\ln(10^{10} A_{\rm s}) = 3.04 \pm 0.01$);
- $n_{\rm s} \rightarrow$ Spectral index of the primordial power spectrum ($n_{\rm s} = 0.965 \pm 0.004$);

- $\tau \rightarrow$ Re-ionisation optical depth ($\tau = 0.054 \pm 0.007$).

Currently, the ΛCDM model is widely accepted, yet certain aspects of its theoretical framework remain enigmatic. Notably, the fundamental essence of its principal constituents, *dark matter* (DM) and DE, remains beyond complete comprehension. While DE serves as a convenient concept to explain the Universe's accelerated expansion, it lacks a definitive and universally agreed-upon physical description. Moreover, it defies association with any know form of energy, exhibiting an exceptionally low density compared to other components. Similarly, DM remains elusive in terms of its composition and detectability, as it interacts solely through gravity, without engaging with radiation. However, its existence finds validation in various observational phenomena, including gravitational lensing by galaxy clusters and redshift-space distortions in large-scale mass distributions. DM can be categorized into two primary types:

- *hot dark matter* (HDM) consists of low-mass relativistic particles, with massive neutrinos emerging as the most promising candidates;
- *cold dark matter* comprises massive non-relativistic particles, with the current focus on weakly interacting massive particles (WIMPs) as the leading contenders.

However, models of structure formation and evolution strongly suggest that the predominant portion of the DM component must exhibit characteristics of coldness.

1.7 Tensions in the ΛCDM

Regarding observational challenges, the enhanced precision of contemporary cosmological and astrophysical measurements, alongside more refined data modeling techniques, have led to notable tensions in the derived cosmological parameter values from different probes. These tensions are particularly pronounced when considering probes that encompass distinct ranges of redshift: those focused on local measurements (referred to as *late* or *low-redshift* probes) and those centered on the measurement of CMB anisotropies (referred to as *early* or *high-redshift* probes). The most perplexing tensions observed today (see Di Valentino et al., 2021a,b,c, for a detailed review) include:

- Hubble tension $\rightarrow$ As discussed in Sect. 1.3, this tension emerges from local direct measurements of H_0 employing the distance ladder approach (see e.g. Riess et al., 2016; Freedman et al., 2019), which are roughly 4.4σ divergent from the value obtained through CMB indirect measurements (Planck Collaboration et al., 2020a).
- Growth of structures tension $\rightarrow$ It arises when direct measurements of the growth rate of cosmological perturbations, derived from weak lensing and clustering techniques (Erben et al., 2013; Abbott et al., 2018; Troxel et al., 2018; Hildebrandt et al., 2020), indicate a slower growth rate compared to that inferred from the *Planck* data, with a discrepancy at $2-3\sigma$. This tension is often

quantified using the parameter $S_8 \equiv \sigma_8\sqrt{\Omega_\mathrm{m}/0.3}$, where σ_8 will be defined in Sect. 3.4.

- Curvature tension → *Planck* data indicate a preference at 3.4σ for a closed Universe (Planck Collaboration et al., 2020a; Di Valentino et al., 2020; Handley, 2021), contradicting the ΛCDM scenario, which assumes a flat space geometry. This discrepancy with the predictions of the flat ΛCDM model is linked to the anomalously higher lensing contribution in the CMB power spectrum, characterized by the A_L parameter (Calabrese et al., 2008; Planck Collaboration et al., 2020a), which exhibits strong degeneracy with $\Omega_{0,\kappa}$.

- Age of the Universe tension → The age of the Universe derived from local measurements using ancient objects, such as the first stars in the Milky Way or populations of stars in globular clusters (see e.g., Bond et al., 2013; Schlaufman et al., 2018; Jimenez et al., 2019; Valcin et al., 2020), seems slightly greater than the age determined from CMB *Planck* data within the framework of the ΛCDM cosmology (Planck Collaboration et al., 2020a).

The emergence of these tensions within the ΛCDM model suggests a potential departure from the assumed standard scenario and raises the possibility of undiscovered physics. Consequently, in recent years, numerous alternatives to the standard cosmological model have been proposed to reconcile theoretical frameworks with observational data. These alternative models generally fall into two main categories used to describe accelerating cosmologies: *dark energy models* and *modified gravity models* (for comprehensive reviews, see Yoo & Watanabe, 2012; Joyce et al., 2016).

Chapter 2

Inflation and Primordial non-Gaussianities

In order to properly understand how cosmic structures form, we need to know the initial conditions that characterize the primordial Universe. This quest for initial conditions leads to an entirely new realm of physics, the theory of inflation (Guth, 1981; Linde, 1982; Starobinsky, 1982; Albrecht & Steinhardt, 1982). Inflation was initially introduced to explain how regions that could not have been in causal contact have the same temperature — in other words — why the Universe is so homogeneous on large scales. It was soon realized that the very mechanism that explains the uniformity of the temperature can also account for the origin of perturbations in the Universe.

2.1 The Horizon Problem and the Inflationary Solution

When the Universe was approximately 3.8×10^5 years old, it displayed exceptional uniformity, with CMB temperature fluctuations of only 1 part in 10^5. At this time, photons and baryons were almost in thermal equilibrium. However, given an arbitrary initial distribution of matter and radiation, one would expect significant inhomogeneities.

A natural explanation for this uniformity is thermalization: in a highly inhomogeneous Universe, regions in thermal contact should eventually reach equilibrium. However, regions observed in the CMB were too distant from each other at recombination to be in causal contact, making thermalization unlikely. To quantify this, we define the comoving horizon (or conformal time) as

$$\eta(t) = \int_0^t \frac{\mathrm{d}t'}{a(t')} \,. \tag{2.1}$$

At recombination, η_* represents the comoving distance light could travel from $\eta = 0$ to η_*, which can be compared to the comoving distance between two regions observed on the CMB sky today. In the standard cosmological model, assuming the Universe was filled with only matter and radiation, the comoving horizon at recombination

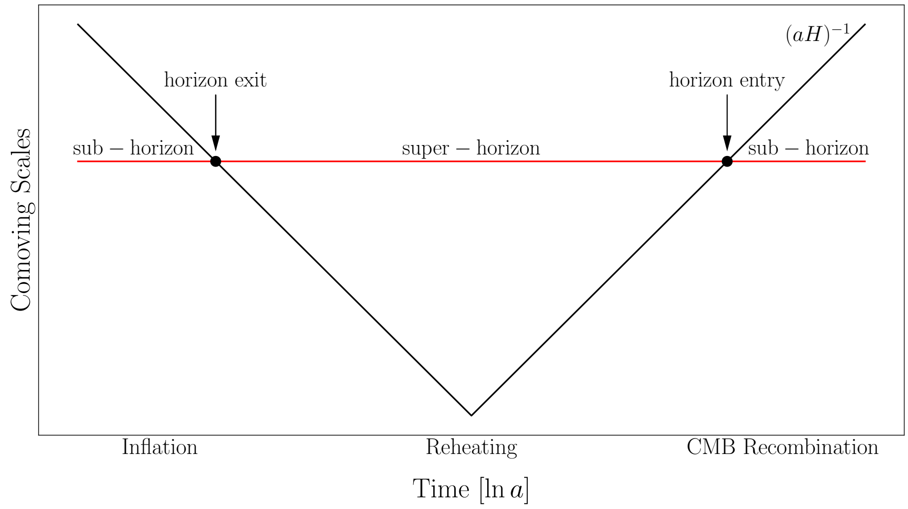

Figure 2.1: The evolution of the comoving Hubble radius $(aH)^{-1}$ as a function of the scale factor a. Very early on during inflation, all scales of interest were smaller than the Hubble radius and, therefore, within causal contact. Based on Nandi (2017).

is $\eta_* = \eta(a_*) \approx 281$ Mpc. For small angular separations θ, the comoving distance between patches on the CMB today is given by

$$\chi(\theta) \approx \chi_* \theta = (\eta_0 - \eta_*)\theta \,, \tag{2.2}$$

where $\eta_0 \approx 14200$ Mpc. Consequently, two CMB regions separated by an angle

$$\theta \geq \frac{\eta_*}{\eta_0 - \eta_*} \approx 1.2^\circ \tag{2.3}$$

could not have been in thermal contact at recombination.

The comoving horizon η can also be expressed as an integral over the scale factor:

$$\eta(a) = \int_0^a \frac{1}{a' H(a')} \, \mathrm{d} \ln a' \,. \tag{2.4}$$

Thus, the comoving horizon is the logarithmic integral of the *comoving Hubble radius* $(aH)^{-1}$, which approximates the distance light can travel in one expansion time. In the presence of matter or radiation, the comoving Hubble radius always increases, meaning causally disconnected regions could not have interacted. A potential solution is an early phase in which the comoving Hubble radius decreased, allowing a much larger region to come into thermal contact. Such phase requires accelerated expansion ($\ddot{a} > 0$) and is referred to as *inflation*. As shown in Fig. 2.1, during inflation the comoving Hubble radius shrinks, ensuring all scales of interest were once causally connected.

Assuming a substance exists that keeps the Hubble rate $H = H_{\mathrm{inf}}$ nearly constant, the scale factor evolves as

$$a(t) = a_e e^{H_{\mathrm{inf}}(t - t_e)} \qquad (t < t_e) \,, \tag{2.5}$$

where t_e is the time when inflation ends. As inflation proceeds, the Universe becomes dominated by the smooth substance driving the acceleration, transforming a chaotic and inhomogeneous region into a much larger, homogeneous, and empty space. During this process, everything within the patch become irrelevant due to dilution, and spacetime perturbations are rapidly smoothed out.

To solve the horizon problem, the comoving Hubble radius before inflation must have been larger than the current comoving radius H_0^{-1}. At the end of inflation, the comoving Hubble radius was $1/a_e H_e$, where $H_e \equiv H(t_e)$. To estimate this magnitude, we assume the temperature after inflation was $T_e = 10^{14}$ GeV and we use the relations for radiation domination (see Table 1.1). Thus, the ratio of the comoving Hubble radius at the end of inflation to today is

$$\frac{a_0 H_0}{a_e H_e} = \frac{a_e}{a_0} \overset{T \propto 1/a}{\simeq} \frac{T_0}{10^{14}\ \text{GeV}} \simeq 10^{-27}\,. \tag{2.6}$$

This implies that the comoving Hubble radius at the end of inflation was 27 orders of magnitude smaller than it is today. Therefore, the scale factor had to increase by a factor of $10^{27} \simeq e^{62}$ during inflation for the current comoving Hubble radius to be smaller than it was at the start of inflation. If the Hubble rate remained constant, this exponential growth implies a period of inflation lasting approximately 60 e-folds.

2.2 Single-field Slow-roll Inflation

We know from Eq. 1.38 that accelerated expansion implies that the quantity $\rho + 3p$ must be negative. However, both non-relativistic matter and radiation are characterized by null and positive pressure, respectively. Thus, whatever drives inflation is not ordinary matter or radiation and cannot be a cosmological constant either: a cosmological constant would lead to perpetual rapid inflation, while we need inflation to end and transition to the radiation- and then matter-dominated phases.

The simplest way to generate such a transitory epoch of accelerated expansion is through the potential energy of a scalar field $\phi(\boldsymbol{x}, t)$, known as *inflaton*. Sometimes several fields may be used to drive inflation, and consequently the corresponding models are referred to as *multi-field* (Byrnes & Wands, 2006; Gong, 2017). The energy momentum tensor of a canonical scalar field with potential $V(\phi)$ is

$$T^\alpha_\beta = g^{\alpha\nu} \frac{\partial\phi}{\partial x^\nu} \frac{\partial\phi}{\partial x^\beta} - \delta^\alpha_\beta \left[\frac{1}{2} g^{\mu\nu} \frac{\partial\phi}{\partial x^\mu} \frac{\partial\phi}{\partial x^\nu} + V(\phi) \right] . \tag{2.7}$$

Let us assume that the field is homogeneous to zeroth-order, meaning that $\phi = \phi(\boldsymbol{x})$. In this case, by evaluating the components T^0_0 and T^i_i (with $i = 1, 2, 3$) we obtain the energy density and the pressure associated to the scalar field:

$$\rho_\phi = \frac{1}{2}\dot{\phi}^2 + V(\phi) \quad \text{and} \quad p_\phi = \frac{1}{2}\dot{\phi}^2 - V(\phi)\,, \tag{2.8}$$

with $\dot{\phi}^2/2$ being the kinetic energy density of the field. A configuration with negative pressure is therefore one with more potential energy than kinetic. This is equivalent to have an equation of state

$$w = \frac{p_\phi}{\rho_\phi} = \frac{\dot{\phi}^2/2 - V(\phi)}{\dot{\phi}^2/2 + V(\phi)} \tag{2.9}$$

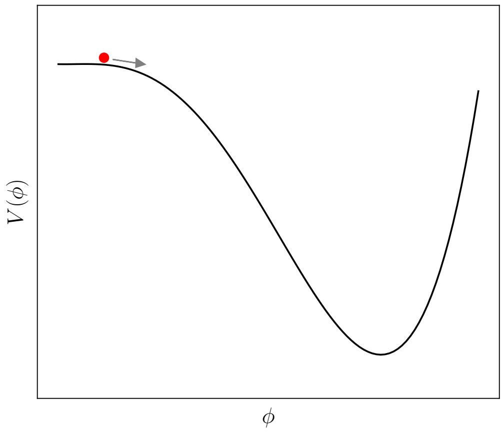

Figure 2.2: The scalar field ϕ slowly rolling down a potential $V(\phi)$. Thanks to the condition of slow-rolling, potential energy dominates over kinetic one, leading to negative pressure and accelerated expansion. The inflationary epoch ends once the field has reached the minimum of the potential. Based on Dodelson & Schmidt (2020).

that is close to -1.

One of the most popular scenario of inflation assumes that the scalar field slowly rolls toward its true ground state (Linde, 1982; Albrecht & Steinhardt, 1982). The potential energy of such a field is very close to constant so it quickly comes to dominate over the kinetic energy. Inflation ends once the field has reached the minimum of the potential, where it oscillates and decays into lighter particles (Fig. 2.2).

To determine the evolution of the inflaton ϕ for a general potential, consider the conservation law of energy and momentum (Eq. 1.7)

$$\frac{\partial \rho_\phi}{\partial t} + 3H(\rho_\phi + p_\phi) = 0\,. \tag{2.10}$$

Applying this to the density and pressure obtained above yields

$$\ddot{\phi} + 3H\dot{\phi} + V_{,\phi}(\phi) = 0\,, \tag{2.11}$$

where $V_{,\phi} \equiv \mathrm{d}V/\mathrm{d}\phi$. We can then use the conformal time η as time variable given that $\mathrm{d}\eta = \mathrm{d}t/a$. It is straightforward to show that

$$\phi'' + 2aH\phi' + a^2 V_{,\phi} = 0 \tag{2.12}$$

where the superscript " $'$ " indicates the derivative with respect to the conformal time.

Most models of inflation are slow-roll models, in which the zeroth-order field, and hence the Hubble rate, vary slowly. Therefore, a simple relation between the conformal time η and the expansion rate holds. In particular, during inflation

$$\eta = \int_{a_e}^{a} \frac{1}{a'H} \mathrm{d}\ln a' = \int_{a_e}^{a} \frac{\mathrm{d}a'}{H(a')^2} \simeq \frac{1}{H}\int_{a_e}^{a} \frac{\mathrm{d}a'}{(a')^2} \simeq -\frac{1}{aH}\,, \tag{2.13}$$

where the second equality holds because H in nearly constant, and the second due to the fact that the scale factor at the end of inflation is much larger than that in the middle. To quantify slow-roll, it is common to define two variables known as *slow-roll parameters*:

$$\epsilon_{\rm sr} = \frac{\rm d}{\rm dt}\left(\frac{1}{H}\right) = -\frac{H'}{aH^2}\,, \tag{2.14}$$

$$\delta_{\rm sr} = \frac{1}{H}\frac{\ddot{\phi}}{\dot{\phi}} = -\frac{1}{aH\phi'}\left(aH\phi' - \phi''\right) = -\frac{1}{aH\phi'}\left(3aH\phi' + a^2V_{,\phi}\right)\,, \tag{2.15}$$

which assume values that are typically very small during inflation.

The slow-roll phase cannot last indefinitely, since inflation must end at some point. This point is reached when the scalar field reaches the minimum of the potential. At that point, the field is no longer slowly rolling, but has significant kinetic energy, so it starts oscillating around the minimum. Then, the equation of state (Eq. 2.9) is no longer close to -1, but close to zero, so that the Universe has transitioned to an epoch characterized by a decelerated expansion. Then, finally, the inflaton ϕ decays into lighter particles, leading to an almost completely homogeneous and radiation-dominated Universe. This transition period is also known as *reheating*.

2.3 Cosmological Perturbations

Inflationary models have the merit that they do not only explain the homogeneity of the Universe on large scales, but also provide a theory for explaining the observed level of anisotropy. During the inflationary period, quantum fluctuations of the field were driven to scales much larger than the Hubble horizon. Then, in this process, the fluctuations were frozen and turned into metric perturbations (Mukhanov & Chibisov, 1981). The metric perturbations created during inflation can be described in two terms. The scalar, or curvature, perturbations are coupled with matter in the Universe and form the initial seeds of structure observed in galaxies today. Instead, the tensor perturbations do not couple to matter, they are associated to the generation of primordial gravitational waves. These perturbations are seen as important components to the CMB anisotropy (Hu & Dodelson, 2002).

Let us consider a homogeneous scalar field that is perturbed, so that the general expression for $\phi(\boldsymbol{x}, t)$ becomes

$$\phi(\boldsymbol{x}, t) = \bar{\phi}(t) + \delta\phi(\boldsymbol{x}, t)\,. \tag{2.16}$$

The equation governing the behavior of $\delta\phi$ in an unperturbed and expanding Universe is obtained starting from the energy and momentum conservation law (Eq. 1.7) with $\nu = 0$. In Fourier space and considering first-order terms, it is possible to derive the equation

$$\delta\phi'' + 2aH\delta\phi' + k^2\delta\phi = 0\,. \tag{2.17}$$

This equation can be written as an harmonic oscillator equation. Moving to the quantum world, it is possible to quantify the amplitude of the perturbations by introducing the so called *power spectrum*, defined by the relation

$$\langle\delta\phi(\boldsymbol{k}, t)\delta\phi^*(\boldsymbol{k}', t)\rangle = (2\pi)^3 P_{\delta\phi}\delta_{\rm D}^{(3)}(\boldsymbol{k} - \boldsymbol{k}')\,, \tag{2.18}$$

where $\boldsymbol{k}$ is the wavevector associated to the perturbation, $\delta\phi^*$ is the complex conjugate and $\delta_{\rm D}^{(3)}$ is the 3D Dirac delta function. It is possible to demonstrate that the power spectrum of fluctuations in $\delta\phi$ is equal to

$$P_{\delta\phi} = \frac{H^2}{2k^3}\,. \tag{2.19}$$

Until now, we have neglected perturbations to the metric that however become important on super-horizon scales, where the wavenumber of the perturbations is $k \lesssim aH$. In Newtonian conformal gauge, we can write the perturbed FLRW metric as

$$\mathrm{d}s^2 = -(1+2\Psi)\mathrm{d}t^2 + a^2(t)\delta_{ij}\,(1+2\Phi)\,\mathrm{d}x^i\mathrm{d}x^j\,, \tag{2.20}$$

where Ψ and Φ are called the scalar gravitational potential. When the anisotropic stresses of both photons and neutrinos are not important (as during matter domination), the two potentials are related by $\Psi \simeq -\Phi$. Typically, primordial scalar fluctuations are expressed in terms of the *curvature perturbation* $\mathcal{R}$, that is a conserved quantity on super-horizon scales. During inflation, it is related to the scalar field ϕ by the relation

$$\mathcal{R} = -\frac{aH}{\bar{\phi}'}\delta\phi\,. \tag{2.21}$$

After inflation ends, during the radiation-dominated era, the curvature perturbation simply becomes

$$\mathcal{R} = -\frac{3}{2}\Psi = \frac{3}{2}\Phi\,. \tag{2.22}$$

Thus, we can relate the post-inflation power spectrum of Ψ (or, equivalently, of Φ) to the spectrum of $\delta\phi$ derived in Eq. 2.19 at horizon crossing

$$P_\Psi(k) = P_\Phi(k) = \frac{4}{9}\left(\frac{aH}{\bar{\phi}'}\right)^2 P_{\delta\phi}\Big|_{k=aH} = \frac{8\pi G}{9k^3}\frac{H^2}{\epsilon_{\rm sr}}\Big|_{k=aH}\,. \tag{2.23}$$

The scalar perturbations generated during inflation are nowadays most commonly parametrized in terms of the power spectrum of the curvature perturbation $\mathcal{R}$. From Eqs. 2.21 and 2.23, we have

$$P_\mathcal{R}(k) = \frac{9}{4}P_\Phi(k) = \frac{2\pi G H^2}{\epsilon_{\rm sr}k^3}\Big|_{k=aH} \equiv 2\pi^2 A_{\rm s} k^{-3}\left(\frac{k}{k_{\rm p}}\right)^{n_{\rm s}-1}\,, \tag{2.24}$$

where $A_{\rm s}$ is the variance of curvature perturbations in a logarithmic wavenumber interval centered around the pivot scale $k_{\rm p}$, and $n_{\rm s}$ is the scalar spectral index (Sect. 1.6). A spectrum such that $k^3P(k)$ is constant (i.e., does not depend on k) is called a scale-invariant or scale-free spectrum. Since $n_{\rm s} \approx 1$ from CMB measurments, we conclude that the spectrum of perturbations generated by inflation is close to be scale-free.

2.4 Primordial non-Gaussianities

Some of the most prominent theories of inflation, such as the single-field slow-roll model, predict that primordial perturbations follow a nearly Gaussian distribution.

Deviations from this prediction, known as PNG, serve as one of the most promising probes to constrain the properties of the inflationary epoch, as almost every non standard model predicts their existence (Liguori et al., 2003; Seery & Lidsey, 2005a,b).

In the context of PNG, the statistical properties of the curvature perturbation $\mathcal{R}$ can be analyzed through its correlation functions, such as the power spectrum $P_{\mathcal{R}}$, the bispectrum $B_{\mathcal{R}}$, and the trispectrum $T_{\mathcal{R}}$:

$$\langle \mathcal{R}(\boldsymbol{k}_1)\mathcal{R}(\boldsymbol{k}_2)\rangle = (2\pi)^3 P_{\mathcal{R}}(k_1)\delta_{\mathrm{D}}^{(3)}(\boldsymbol{k}_1+\boldsymbol{k}_2)\,, \tag{2.25}$$

$$\langle \mathcal{R}(\boldsymbol{k}_1)\mathcal{R}(\boldsymbol{k}_2)\mathcal{R}(\boldsymbol{k}_3)\rangle = (2\pi)^3 B_{\mathcal{R}}(k_1,k_2,k_3)\delta_{\mathrm{D}}^{(3)}(\boldsymbol{k}_1+\boldsymbol{k}_2+\boldsymbol{k}_3)\,, \tag{2.26}$$

$$\langle \mathcal{R}(\boldsymbol{k}_1)\mathcal{R}(\boldsymbol{k}_2)\mathcal{R}(\boldsymbol{k}_3)\mathcal{R}(\boldsymbol{k}_4)\rangle = (2\pi)^3 T_{\mathcal{R}}(k_1,k_2,k_3,k_4)\delta_{\mathrm{D}}^{(3)}(\boldsymbol{k}_1+\boldsymbol{k}_2+\boldsymbol{k}_3+\boldsymbol{k}_4)\,. \tag{2.27}$$

Unlike the power spectrum, which depends on a single wavenumber, the bispectrum and trispectrum depend on three and four wavenumbers, respectively. The specific form of these higher-order statistics depends on the underlying inflationary mechanism, which gives rise to different characteristic shapes, such as the *local*, *equilateral*, and *orthogonal* templates. These shapes will be briefly introduced in the following subsection.

2.4.1 Bispectrum

Focusing on the bispectrum, its amplitude is typically described by the nonlinearity parameter $f_{\mathrm{NL}}^{\mathrm{X}}$, where the superscript "X" indicates the specific shape. Since the shape of the bispectrum is directly linked to the physical mechanism responsible for generating primordial fluctuations, it provides a powerful tool to distinguish between different models. Below, we summarize some of the most studied bispectrum shapes:

- Local type $\rightarrow$ This type of bispectrum arises from nonlinearities generated on super-horizon scales in the primordial adiabatic fluctuations. The curvature perturbation in this case can be expressed as (Komatsu & Spergel, 2001)

  $$\mathcal{R} = \mathcal{R}_{\mathrm{G}} + \frac{3}{5} f_{\mathrm{NL}}^{\mathrm{loc}} \mathcal{R}_{\mathrm{G}}^2 + \dots\,, \tag{2.28}$$

  where $\mathcal{R}_{\mathrm{G}}$ represents the Gaussian component of the perturbation. This expansion leads to a bispectrum of the form:

  $$B_{\mathcal{R}}^{\mathrm{loc}}(k_1,k_2,k_3) = \frac{6}{5} f_{\mathrm{NL}}^{\mathrm{loc}} \left[P_{\mathcal{R}}(k_1)P_{\mathcal{R}}(k_2) + P_{\mathcal{R}}(k_2)P_{\mathcal{R}}(k_3) + P_{\mathcal{R}}(k_3)P_{\mathcal{R}}(k_1)\right] \tag{2.29}$$

  with $P_{\mathcal{R}}(k)$ defined in Eq. 2.24. As illustrated in Fig. 2.3, this shape of the bispectrum peaks in the squeezed limit $k_3 \ll k_1 \simeq k_2$. The curvature perturbation $\mathcal{R}$ can be related to the gravitational potential Φ during the matter-dominated era through the relation $\mathcal{R} = 5\Phi/3$ (valid on super-horizon scales). Examples of models that produce local-type PNG include the curvaton scenario (Moroi & Takahashi, 2001; Lyth & Wands, 2002; Enqvist & Sloth, 2002), modulated reheating (Kofman, 2003; Dvali et al., 2004), mixed inflaton-curvaton models (Langlois & Vernizzi, 2004; Lazarides et al., 2004; Moroi et al., 2005), multi-field inflation (Byrnes et al., 2008; Byrnes & Choi, 2010), and others.

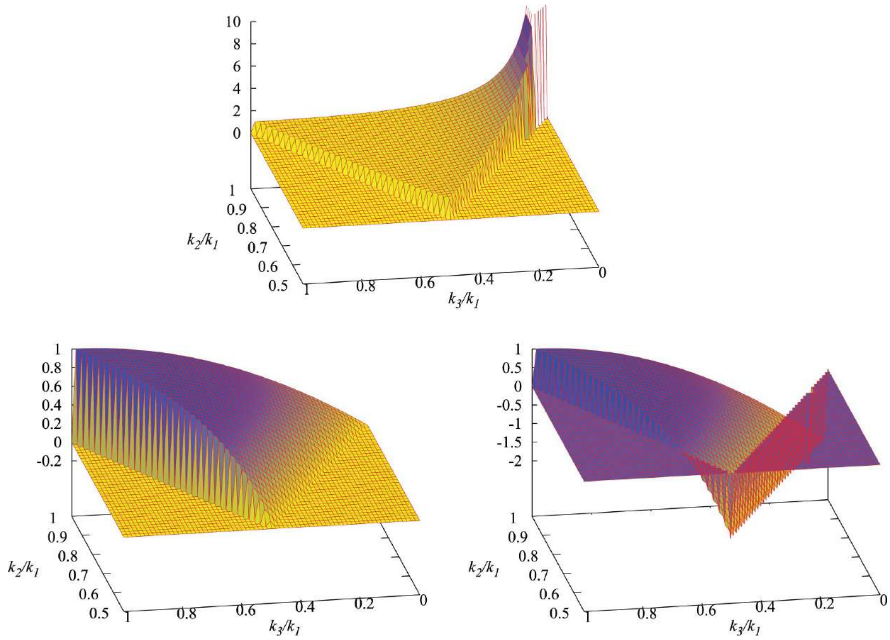

Figure 2.3: Shapes of the curvature bispectrum: local (*top*), equilateral (*bottom left*) and orthogonal (*bottom right*). Credits to Takahashi (2014).

- Equilateral type → This shape of bispectrum is characterized by the following functional form (Creminelli et al., 2006):

$$
\begin{aligned}
B_{\mathcal{R}}^{\text{equil}}(k_1,k_2,k_3) = \frac{3}{5} f_{\text{NL}}^{\text{equil}} \Big[& -3P_{\mathcal{R}}(k_1)P_{\mathcal{R}}(k_2) - 2P_{\mathcal{R}}^{2/3}(k_1)P_{\mathcal{R}}^{2/3}(k_2) \times \\
& \times P_{\mathcal{R}}^{2/3}(k_3) + 6P_{\mathcal{R}}^{1/3}(k_1)P_{\mathcal{R}}^{2/3}(k_2)P_{\mathcal{R}}(k_3) + (5 \text{ permutations}) \Big] \, ,
\end{aligned}
\tag{2.30}
$$

  which peaks around $k_1 \simeq k_2 \simeq k_3$, as shown in Fig. 2.3. This shape of the bispectrum can be obtained by considering a non-canonical form of the scalar field ϕ. Alternatively, other models of inflation capable of generating equilateral-type PNG are: k-inflation (Armendáriz-Picón et al., 1999; Garriga & Mukhanov, 1999), Dirac-Born-Infeld inflation (Alishahiha et al., 2004) and ghost inflation (Arkani-Hamed et al., 2004; Izumi & Mukohyama, 2010).

- Orthogonal type → In some models, another type of bispectrum can arise, which is orthogonal to the local and equilateral forms. Hence, it is referred to as the orthogonal shape. Its bispectrum form is given by (Senatore et al., 2010):

$$
\begin{aligned}
B_{\mathcal{R}}^{\text{ortho}}(k_1,k_2,k_3) = \frac{3}{5} f_{\text{NL}}^{\text{ortho}} \Big[& -9P_{\mathcal{R}}(k_1)P_{\mathcal{R}}(k_2) - 8P_{\mathcal{R}}^{2/3}(k_1)P_{\mathcal{R}}^{2/3}(k_2) \times \\
& \times P_{\mathcal{R}}^{2/3}(k_3) + 18P_{\mathcal{R}}^{1/3}(k_1)P_{\mathcal{R}}^{2/3}(k_2)P_{\mathcal{R}}(k_3) + (5 \text{ permutations}) \Big] \, .
\end{aligned}
\tag{2.31}
$$

  This shape gives a positive amplitude in the limit $k_1 \simeq k_2 \simeq k_3$, and negative in the limit $k_1 \simeq 2k_2 \simeq 2k_3$, as shown in Fig. 2.3.

2.5 Primordial non-Gaussianity from Single-field Inflation

Many models over the past year were proposed in the attempt of characterize the epoch of inflation. In most models, the predicted value of $f_{\rm NL}$ can vary from $f_{\rm NL} \sim \mathcal{O}(1)$ to $f_{\rm NL} \gg \mathcal{O}(1)$ (see Takahashi, 2014, for a detailed review of the topic). Here, we discuss about the predictions of the single-field inflation model.

Let us consider an action of the form (Garriga & Mukhanov, 1999)

$$S = \int \mathrm{d}^4x\sqrt{-g}\left[\frac{1}{2}M_{\rm pl}^2 R + P(X,\phi)\right] , \tag{2.32}$$

where $X \equiv -(1/2)g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi$, R is the Ricci scalar and $M_{\rm pl} \equiv (8\pi G)^{-1/2}$ is the Planck mass. With this action we can define the power spectrum as

$$\mathcal{P}_{\mathcal{R}}(k) = \frac{k^3}{2\pi^2}P_{\mathcal{R}}(k) = \frac{1}{8\pi^2 M_{\rm pl}^2}\frac{H^2}{c_{\rm s}\epsilon} , \tag{2.33}$$

with $c_{\rm s}$ being the sound speed, defined by

$$c_{\rm s}^2 = \frac{P_{,X}}{P_{,X} + 2XP_{,XX}} . \tag{2.34}$$

Here, $P_{,X}$ and $P_{,XX}$ indicate first and second derivatives with respect to the quantity X. Additionally, the slow-roll parameters can be derived as follows:

$$\epsilon = -\frac{\dot{H}}{H^2} , \quad \eta_H = \frac{\dot{\epsilon}}{\epsilon H} , \quad s = \frac{\dot{c}_{\rm s}}{c_{\rm s}H} . \tag{2.35}$$

In this case, we can write the bispectrum as (Chen et al., 2007)

$$\begin{aligned} B_{\mathcal{R}}(k_1,k_2,k_3) = (2\pi)^4\mathcal{P}_{\mathcal{R}}^2\frac{1}{\prod_{i=1}^3 k_i^3}\Bigg[&\left(\frac{1}{c_{\rm s}^2} - 1 - \frac{2\lambda}{\Sigma}\right)\frac{3k_1^2k_2^2k_3^2}{2K_{\rm t}^3} + \\ + \left(\frac{1}{c_{\rm s}^2} - 1\right)&\left(-\frac{1}{K_{\rm t}}\sum_{i<j}k_i^2k_j^2 + \frac{1}{2K_{\rm t}^2}\sum_{i\neq j}k_i^2k_j^3 + \frac{1}{8}\sum_i k_i^3\right) + \\ &A_O + A_\epsilon + A_\eta + A_s\Bigg] , \end{aligned} \tag{2.36}$$

where $K_{\rm t} \equiv k_1 + k_2 + k_3$, $A_{O,s,\epsilon,\eta}$ depend on the slow-roll parameters defined in Eq. 2.35 and

$$\Sigma = XP_{,X} + 2X^2P_{,XX} = \frac{H^2\epsilon}{c_{\rm s}^2} , \quad \lambda = X^2P_{,XX} + \frac{2}{3}X^3P_{,XXX} . \tag{2.37}$$

To describe the amplitude of the bispectrum, one can generalize the definition of the parameter $f_{\rm NL}$ as

$$\langle\mathcal{R}(k_1)\mathcal{R}(k_2)\mathcal{R}(k_3)\rangle = (2\pi)^7\delta_{\rm D}^{(3)}(\boldsymbol{k_1} + \boldsymbol{k_2} + \boldsymbol{k_3})\frac{\sum_{i=1}^3 k_i^3}{\prod_{i=1}^3 k_i^3}\mathcal{P}_{\mathcal{R}}^2\left(\frac{3}{10}f_{\rm NL}\right) . \tag{2.38}$$

Taking the equilater limit ($k_1 = k_2 = k_3 = k$), one can derive $f_{\rm NL}^{\rm equil}$ for the expression Eq. 2.36 as

$$f_{\rm NL}^{\rm equil} = \frac{85}{324}\left(1 - \frac{1}{c_{\rm s}^2}\right) - \frac{10}{81}\frac{\lambda}{\Sigma}\,. \tag{2.39}$$

Now, let us consider a standard slow-roll single-field inflation model characterized by a canonical kinetic term $P = X - V(\phi)$. The functional form of the bispectrum has been calculated as (Maldacena, 2003)

$$\begin{aligned} B_{\mathcal{R}}(k_1, k_2, k_3) = \frac{(2\pi)^4}{8}\mathcal{P}_{\mathcal{R}}^2 \frac{1}{\prod k_i^3}\Bigg[&(3\epsilon - 2\eta_\phi)\sum_i k_i^3 + \epsilon \sum_{i\neq j} k_i k_j^2 + \\ &+ 8\epsilon \frac{\sum_{i>j} k_i^2 k_j^2}{K_{\rm t}}\Bigg]\,, \end{aligned} \tag{2.40}$$

where $\eta_\phi \equiv M_{\rm pl}^2 (V_{,\phi\phi}/V)$ is the slow-roll parameter defined with respect to the potential V of the inflaton. We note that the two parameters η_ϕ and η_H are related by $\eta_H = -2\eta_\phi + 4\epsilon$. In the squeezed limit where $k_3 \to 0$, the above bispectrum reduces to

$$B_{\mathcal{R}}(k_1, k_2, k_3) = 2(3\epsilon - \eta_\phi) P_{\mathcal{R}}(k_1) P_{\mathcal{R}}(k_3) = (1 - n_{\rm s}) P_{\mathcal{R}}(k_1) P_{\mathcal{R}}(k_3)\,. \tag{2.41}$$

By taking the same limit in $B_{\mathcal{R}}^{\rm loc}$ given in Eq. 2.29 and comparing it with the above expression, we obtained the so-called *consistency relation* between $f_{\rm NL}^{\rm loc}$ and $n_{\rm s}$

$$f_{\rm NL}^{\rm loc} = \frac{5}{12}(1 - n_{\rm s})\,. \tag{2.42}$$

In general, this relation has been shown to hold for any single-filed inflation model, regardless of its potential and kinetic term (Creminelli & Zaldarriaga, 2004). Since current constraints on $n_{\rm s}$ indicate $1 - n_{\rm s} = \mathcal{O}(10^{-2})$, any detection of $f_{\rm NL}^{\rm loc} \gg 10^{-2}$ would rule out any single-field model for inflation. Today, the best constraints on PNGs come from the *Planck* data (Planck Collaboration et al., 2020b) and they indicate

$$f_{\rm NL}^{\rm loc} = -0.9 \pm 5.1\,, \quad f_{\rm NL}^{\rm equil} = -26 \pm 47\,, \quad f_{\rm NL}^{\rm ortho} = -38 \pm 24\,. \tag{2.43}$$

Chapter 3

Theory of Structure Formation

In the last chapter, we provided an explanation for why our Universe is homogeneous and isotropic on large scales. However, when we restrict our focus to smaller scales, specifically on the order of Megaparsecs (Mpc), it becomes evident that this approximation no longer holds. On these smaller scales, we observe significant fluctuations in matter density, with amplitudes reaching hundreds of times the mean density, indicating a highly nonlinear evolution. Thanks to CMB maps (Planck Collaboration et al., 2020c), it is possible to derive the order of magnitude of the perturbations at the time of recombination. Assuming adiabatic perturbations, we have

$$\frac{\delta T}{\bar{T}} \simeq \frac{\delta \rho}{\bar{\rho}} \simeq 10^{-5}\,, \tag{3.1}$$

where $\bar{T}$ and $\bar{\rho}$ indicates the mean black body temperature of the CMB and the mean density of the Universe, respectively. As a consequence of the small fluctuations we can conclude that, at this epoch, the Universe was nearly homogeneous.

From this evidence, it is clear that from recombination to the present day, perturbations have grown by several orders of magnitude due to gravitational instability. The evolution of fluid perturbations under the influence of gravity was first analytically described in Jeans (1902). The theory developed from these studies can be extended to the cosmological framework to predict the formation of large-scale structures (LSS) in the Universe. In the next section, we will specifically see how inhomogeneities in the primordial fluid are amplified as the Universe evolves. However, the results of Jeans' theory are accurate as long as we remain in a linear regime, where perturbations can be considered small. The evolution of perturbations in the nonlinear regime can be analytically addressed only for extremely simple models, such as spherical evolution (Sect. 3.2). Today, the study of structure evolution during the nonlinear phase is mostly carried out using N-body simulations (Sect. 3.3).

3.1 Linear Theory

The aim of Jeans' theory is to describe the rate at which initial density perturbations must grow to match the inhomogeneities observed today. This model is applicable to non-relativistic matter and on scales not exceeding the cosmological horizon, which represents the portion of the Universe that is in causal connection with the observer. The cosmological horizon is defined as in Sect. 2.1:

$$\eta(t) = \int_0^t \frac{\mathrm{d}t'}{a(t')} \,. \tag{3.2}$$

At this point we can separate perturbations according to their size:

- for scales $r > \eta$ the only force affecting the fluctuations is gravity and perturbations are always able to grow;
- for scales $r < \eta$ micro-physical processes become important and the perturbations evolves according to the prediction of the Jeans's theory at linear order.

To study the evolution of a perturbation with size larger than the cosmological horizon, we can treat the fluctuation as a closed Universe embedded in a flat and single-component background Universe. Let us write the first Friedmann equation (Eq. 1.26) for the perturbation and for the background:

$$H_\mathrm{b}^2 = \frac{8\pi G}{3}\rho_\mathrm{b} \,, \quad H_\mathrm{p}^2 = \frac{8\pi G}{3}\rho_\mathrm{p} - \frac{1}{a^2} \,, \tag{3.3}$$

where the subscripts "b" and "p" refer to the background and to the perturbation, respectively. Since the perturbation is entirely embedded within the background, their corresponding scale factors are initially the same, allowing us to set their Hubble parameters to be equal, yielding:

$$\frac{8\pi G}{3}\rho_\mathrm{b} = \frac{8\pi G}{3}\rho_\mathrm{p} - \frac{1}{a^2} \quad \Longrightarrow \quad \delta \equiv \frac{\rho_\mathrm{p} - \rho_\mathrm{b}}{\rho_\mathrm{b}} = \frac{3}{8\pi G \rho_\mathrm{b} a^2} \,. \tag{3.4}$$

Therefore, the evolution of the perturbation depends on the evolution of the density of the background Universe. Using the relations reported in Table 1.1 we obtain the following results:

$$\begin{cases} \delta = \delta_\mathrm{r} \propto a^2 \propto t & \text{for } z > z_\mathrm{eq} \\[2ex] \delta = \delta_\mathrm{m} \propto a \propto t^{2/3} & \text{for } z < z_\mathrm{eq} \,. \end{cases} \tag{3.5}$$

Consequently, for scales larger than the horizon, perturbations can continuously grow throughout cosmic time.

The analysis of perturbations on scales smaller than the horizon can be carried out in the Newtonian approximation using the equation governing a fluid embedded in an expanding Universe:

$$\begin{cases} \dfrac{\partial \rho}{\partial t} + \nabla \cdot (\rho \boldsymbol{v}) = 0 \\[2ex] \dfrac{\partial \boldsymbol{v}}{\partial t} + (\nabla \cdot \boldsymbol{v})\boldsymbol{v} = -\dfrac{1}{\rho}\nabla p - \nabla \Phi \\[2ex] \nabla^2 \Phi = 4\pi G \rho \\[2ex] \dfrac{\mathrm{d}S}{\mathrm{d}t} = 0 \\[2ex] p = p(\rho, S) = p(\rho) \end{cases} \tag{3.6}$$

where $\boldsymbol{v}$ is the velocity of a fluid element, Φ is the gravitational potential, S is the entropy and p is the pressure. This set of equations allows for a background solution that can be perturbed by introducing small fluctuations ($\delta q/q \ll 1$, with q any quantity of interest). Regarding velocity, it consists of two components: the first is the Hubble flow, and the second is the peculiar velocity

$$\boldsymbol{v} = H\boldsymbol{d} + \boldsymbol{v}_{\mathrm{p}}\,, \tag{3.7}$$

where $\boldsymbol{v}$ indicates the fluid position vector.

At this point it is useful to define the dimensionless *density contrast* through the relation

$$\delta(\boldsymbol{x}, t) = \frac{\delta\rho(\boldsymbol{x}, t)}{\rho_{\mathrm{b}}} \tag{3.8}$$

and solve the system of equation for it. In Fourier space, what results from this is the so-called *dispersion relation*

$$\ddot{\delta}_k + 2H\dot{\delta}_k + \delta_k\left(k^2 c_{\mathrm{s}}^2 - 4\pi G\rho_{\mathrm{b}}\right) = 0\,, \tag{3.9}$$

where δ_k is the Fourier transform of the density contrast and c_{s} is the sound speed. The term $2H(t)\dot{\delta}_k$ is the Hubble friction, while the term $k^2c_{\mathrm{s}}^2\delta_k$ accounts for the characteristic velocity field of the fluid. These terms tend to dissipate the fluctuations, acting against their growth. We can now separate the solutions of the dispersion relation based on their scale λ, in relation to a characteristic scale λ_{J}, known as the Jeans' scale, and given by

$$\lambda_{\mathrm{J}} = \frac{2\pi}{k_{\mathrm{J}}} = c_{\mathrm{s}}\left(\frac{\pi}{G\rho_{\mathrm{b}}}\right)^{1/2}\,. \tag{3.10}$$

In particular, for $\lambda < \lambda_{\mathrm{J}}$ the perturbation propagates as a wave with constant amplitude and velocity tending to c_{s} for $\lambda \ll \lambda_{\mathrm{J}}$. On the other hand, for $\lambda > \lambda_{\mathrm{J}}$, the dispersion relation leads to two solutions, referred to as the growing and decaying modes. For an EdS Universe we obtain

$$\begin{cases} \delta_+(t) \propto t^{2/3} \propto a \\ \delta_-(t) \propto t^{-1} \propto a^{-3/2}\,. \end{cases} \tag{3.11}$$

Since the decaying solution does not give rise to gravitational instability (i.e. collapsed structures), we are primarily interested in the growing mode. If the only relevant component apart from matter are a cosmological constant and curvature, we can obtain an integral form of the growing mode solution

$$\delta_+(a) \propto H(a)\int^a \frac{\mathrm{d}a'}{[a'H(a')]^3} \tag{3.12}$$

that, unfortunately, has no analytic solution. However, there exist an empirical relation for the growth rate f, i.e. the logarithmic derivative of δ_+ with respect to the scale factor:

$$f(a) \equiv \frac{\mathrm{d}\ln\delta_+(a)}{\mathrm{d}\ln a} \simeq \Omega_{\mathrm{m}}^{0.55} + \frac{\Omega_\Lambda}{70}\left(1 + \frac{\Omega_{\mathrm{m}}}{2}\right)\,. \tag{3.13}$$

This relation implies that the energy density of matter strongly influences the growth of perturbation, while dark energy has a minor impact. Moreover, the exponent 0.55 is a prediction of GR (Coles & Lucchin, 2002; Dodelson & Schmidt, 2020); thus, measuring it serves as a test for the theory.

3.2 Nonlinear Evolution: Spherical Model

The cosmic structures we observe in today's Universe, such as galaxies, galaxy clusters, and DM halos, are the result of gravitational instabilities that have occurred throughout cosmological history. Unfortunately, to describe the formation of these objects, characterized by a strongly nonlinear regime ($\delta \gg 1$), the small-perturbations approximation introduced in the previous section is no longer accurate. Once the linear regime breaks down, meaning δ close to unity, the weakly nonlinear regime begins. During this stage, the fluctuation distribution function already starts to deviate from a Gaussian shape. Moreover, we must consider that the evolution of the baryonic component differs from that of the DM. In fact, baryons are subject to hydrodynamical effects such as star formation, supernova explosions, and feedback from active galactic nuclei. These phenomena further complicate the description of the entire scenario with a comprehensive and solid theory. For this reason, the most viable way to understand what happens when we move out of the linear regime is to rely on N-body simulations. However, there are very few and specific cases in which it is possible to proceed analytically.

The model we present here, known as the spherical evolution model (Gunn & Gott, 1972), is sufficiently accurate for describing the isolated formation of spherical collapsed overdensities (i.e. DM halos). By considering an initially spherical perturbation, we can represent it as a closed or open Universe, respectively, evolving within a flat and single-component background Universe. We start at an initial time $t_\mathrm{i} \gg t_\mathrm{eq}$, which allows us to study the evolution of perturbations during the matter-dominated era. Assuming the validity of the CP, each perturbation can be treated as an independent Friedmann Universe as long as it evolves adiabatically. Therefore, the only interaction we need to consider is gravitational.

3.2.1 Overdensities

Let us study the evolution of an initially overdense shell embedded in an EdS Universe. In Sect. 3.1 we have derived the growing and decaying modes for a matter perturbation. Consequently, the density contrast can be written as the combination of these two modes:

$$\delta(t) = \delta_+(t_\mathrm{i}) \left(\frac{t}{t_\mathrm{i}}\right)^{2/3} + \delta_-(t_\mathrm{i}) \left(\frac{t}{t_\mathrm{i}}\right)^{-1} . \tag{3.14}$$

Assuming the perturbations have an initial zero velocity, we can take the derivative of the previous relation with respect to time and evaluate it at the initial time, yielding

$$\frac{2}{3}\delta_+(t_\mathrm{i}) = \delta_-(t_\mathrm{i}) \,. \tag{3.15}$$

Therefore, inserting this result in Eq. 3.14 leads to the following expression at $t = t_{\rm i}$:

$$\delta_{\rm i} \equiv \delta(t_{\rm i}) = \delta_+ + \delta_- = \frac{5}{3}\delta_+(t_{\rm i})\,. \tag{3.16}$$

Consequently, 3/5 of the initial perturbation is captured by the growing mode, while the remaining 2/5 decays over time, eventually becoming negligible.

Now, let us consider the density parameter of the perturbation, denoted as $\Omega_{\rm p}$. In the case of an overdense region, the perturbation behaves like a closed Universe undergoing collapse. Therefore, the overdensity must satisfy the condition $\Omega_{\rm p} > 1$, that can be written as

$$\Omega_{\rm p}(t_{\rm i}) \equiv \frac{\rho_{\rm p}(t_{\rm i})}{\rho_{\rm crit}(t_{\rm i})} = \frac{\rho_{\rm b}(t_{\rm i})(1+\delta_{\rm i})}{\rho_{\rm crit}(t_{\rm i})} = \Omega(t_{\rm i})(1+\delta_{\rm i}) > 1\,, \tag{3.17}$$

where $\Omega(t_{\rm i})$ denotes the initial density parameter of the EdS background Universe. It follows that the condition for collapse can be translated into a condition on the initial density contrast

$$\delta_{\rm i} > \frac{1-\Omega(t_{\rm i})}{\Omega(t_{\rm i})}\,. \tag{3.18}$$

At this point, we can make use of the tight relation that exists between the density parameter today and that at a given redshift z for a generic single-component Universe

$$\Omega(z) = \frac{\Omega_0(1+z)^{1+3w_{\rm s}}}{(1-\Omega_0) + \Omega_0(1+z)^{1+3w_{\rm s}}} \tag{3.19}$$

Since we are considering times deep into the matter-dominated era, we have $w_{\rm m} = 0$. Hence, the condition for collapse becomes

$$\delta_{\rm i} > \frac{1-\Omega_0}{\Omega_0(1+z_{\rm i})}\,. \tag{3.20}$$

Therefore, for a closed or flat Universe ($\Omega_0 \geq 1$), collapse occurs for any positive initial density contrast of the perturbation. However, in an open Universe ($\Omega_0 < 1$), expansion inhibits collapse, which can only occur if the initial density contrast exceeds a certain threshold.

The spherical overdense region begins to expand at a rate slower than the Hubble flow, thereby increasing its density contrast. The expansion gradually slows down until the overdensity reaches its maximum size, characterized by a radius $R_{\rm max}$. After this moment, known as *turn-around*, collapse begins, initiating structure formation. It is possible to show that, at the moment of maximum expansion $t = t_{\rm max}$, the density of the perturbation is

$$\rho_{\rm p}(t_{\rm max}) = \frac{3\pi}{32G}\frac{1}{t_{\rm max}^2}\,. \tag{3.21}$$

Recalling the fundamental relations for the EdS model in Table 1.1, we can easily compute the density contrast at the turn-around:

$$\delta(t_{\rm max}) = \frac{\rho_{\rm p}(t_{\rm max})}{\rho_{\rm b}(t_{\rm max})} - 1 = \left(\frac{3\pi}{4}\right)^2 \simeq 4.6\,. \tag{3.22}$$

Thus, even at the moment of the turn-around, the collapsing region is already within the nonlinear regime and exhibits a density more than 5 times larger than that of the background Universe. Alternatively, if computed within the framework of linear theory, the same quantity would be:

$$\delta^{\mathrm{L}}(t_{\mathrm{max}}) = \delta_{+}(t_{\mathrm{i}}) \left(\frac{t_{\mathrm{max}}}{t_{\mathrm{i}}}\right)^{2/3} = \frac{3}{5}\left(\frac{3\pi}{4}\right)^{2/3} \simeq 1.07\,. \tag{3.23}$$

In principle, a rigorous gravitational treatment of a closed Universe implies that after reaching the turn-around point, the size of the overdensity will shrink until it becomes a singularity at $t = 2t_{\mathrm{max}}$. However, due to hydrodynamical interactions of baryons or the increase of the dispersion velocity of DM particles, a virialized structure with a size given by the virial radius R_{vir} forms at $t_{\mathrm{vir}} = 3t_{\mathrm{max}}$.

Due to the virialization of the overdensity at the end of collapse, we can establish a relationship between its kinetic energy $\mathcal{T}$ and gravitational potential energy $\mathcal{V}$ using the *scalar virial theorem*, which states that

$$2\mathcal{T} + \mathcal{V} = 0\,. \tag{3.24}$$

Assuming a perfectly self-gravitating spherical overdensity with mass M, the gravational potential energy can be expressed as

$$\mathcal{V} = -\frac{3}{5}\frac{GM^2}{R}\,. \tag{3.25}$$

Using this and Eq. 3.24, we obtain that at the moment of virialization t_{vir}, the total energy of the system E is given by

$$E(t_{\mathrm{vir}}) = \mathcal{T} + \mathcal{V} = \frac{\mathcal{V}}{2} = -\frac{3}{10}\frac{GM^2}{R_{\mathrm{vir}}}\,. \tag{3.26}$$

In the absence of any energy loss during the collapse phase, we can assume that $E(t_{\mathrm{vir}}) = E(t_{\mathrm{max}})$, which leads to $2R_{\mathrm{vir}} = R_{\mathrm{max}}$. Since the density scales as R^{-3}, the density of the perturbation once it is virialized is 8 times higher than that at the turn-around. Therefore, the density contrasts at $t_{\mathrm{coll}} = 2t_{\mathrm{max}}$ and $t_{\mathrm{vir}} = 3t_{\mathrm{max}}$, assume the following values:

$$\delta(t_{\mathrm{coll}}) = \frac{\rho_{\mathrm{p}}(t_{\mathrm{coll}})}{\rho_{\mathrm{b}}(t_{\mathrm{coll}})} - 1 = \frac{8\rho_{\mathrm{p}}(t_{\mathrm{max}})}{\rho_{\mathrm{b}}(t_{\mathrm{max}})}\left(\frac{t_{\mathrm{coll}}}{t_{\mathrm{max}}}\right)^2 - 1 \simeq 178\,; \tag{3.27}$$

$$\delta(t_{\mathrm{vir}}) = \frac{\rho_{\mathrm{p}}(t_{\mathrm{vir}})}{\rho_{\mathrm{b}}(t_{\mathrm{vir}})} - 1 = \frac{8\rho_{\mathrm{p}}(t_{\mathrm{max}})}{\rho_{\mathrm{b}}(t_{\mathrm{max}})}\left(\frac{t_{\mathrm{vir}}}{t_{\mathrm{max}}}\right)^2 - 1 \simeq 402\,. \tag{3.28}$$

In linear theory, the same quantities would be

$$\delta^{\mathrm{L}}(t_{\mathrm{coll}}) = 1.06\left(\frac{t_{\mathrm{coll}}}{t_{\mathrm{max}}}\right)^{2/3} \simeq 1.686\,, \tag{3.29}$$

$$\delta^{\mathrm{L}}(t_{\mathrm{vir}}) = 1.06\left(\frac{t_{\mathrm{vir}}}{t_{\mathrm{max}}}\right)^{2/3} \simeq 2.2\,. \tag{3.30}$$

The values in Eq. 3.27 and Eq. 3.28 computed in nonlinear theory closely approximate those measured, but they are significantly influenced by the cosmological model, particularly by the curvature. In contrast, the corresponding linear values are less affected (Kitayama & Suto, 1996; Jenkins et al., 2001).

3.3 Numerical N-body Simulations

As we have seen, the analytical treatment of structure formation is very complex, especially when structures enter their nonlinear phase of evolution, for which only a few solutions exist. Additionally, the micro-physics related to the presence of baryons must also be considered, significantly complicating the overall model. To address these challenges, numerical simulations are used. Since the formation of cosmic structures is essentially the result of the dynamic evolution of a particle system, once the underlying cosmological scenario and initial conditions are established, the simulation is run to track the system's evolution.

The primary effect crucial for mimicking the evolution of density perturbations is the gravitational interaction, which dominates on large scales and affects the majority of the matter in the Universe (specifically, DM). Simulations that exclusively incorporate gravitational forces are termed *N-body simulations*. For a more accurate depiction of LSS, hydrodynamic effects arising from baryonic matter must also be included. Simulations that evolve both the baryonic and dark matter components are known as *hydrodynamic simulations*.

Focusing on simulations that accounts for only gravity, the particle system is evolved by solving the following set of equations:

$$\begin{cases} \boldsymbol{F}_i = G M_i \displaystyle\sum_{i \neq j}^{N} \frac{M_j}{r_{ij}^2} \widehat{\boldsymbol{r}}_{ij} \\ \ddot{\boldsymbol{x}}_i = \dfrac{\mathrm{d}\boldsymbol{v}_i}{\mathrm{d}t} = \dfrac{\boldsymbol{F}_i}{M_i} \\ \dot{\boldsymbol{x}}_i = \dfrac{\mathrm{d}\boldsymbol{x}_i}{\mathrm{d}t} = \boldsymbol{v}_i \,. \end{cases} \tag{3.31}$$

For each i-th particle, $\boldsymbol{F}_i$ represents the force acting on it, M_i denotes its mass, $\boldsymbol{x}_i$ is its comoving 3-dimensional position and $\boldsymbol{v}_i$ indicates its velocity components. Given this set of equations, the Euler equation can be written as

$$\frac{\mathrm{d}\boldsymbol{x}_i}{\mathrm{d}t} + 2\frac{\dot{a}}{a}\boldsymbol{v}_i = -\frac{1}{a^2}\nabla\Phi = -\frac{G}{a^3}\sum_{i\neq j}^{N} M_j \frac{\boldsymbol{x}_i - \boldsymbol{x}_j}{|\boldsymbol{x}_i - \boldsymbol{x}_j|^3} = \frac{\boldsymbol{F}_i}{a^3}\,, \tag{3.32}$$

where a is the scale factor. By utilizing the first Friedmann equation (Eq. 1.26), the Poisson equation becomes

$$\nabla^2\Phi = 4\pi G\bar{\rho}a^2\delta = \frac{3}{2}H_0^2\Omega_{0,\mathrm{m}}\frac{\delta}{a}\,, \tag{3.33}$$

with $\bar{\rho}$ and δ representing the mean matter density and the local density contrast, respectively.

N-body simulations integrate all the involved equations over discretized time steps δt, with varying efficiency depending on the method implemented. For instance, the simplest method to compute the gravitational force $\boldsymbol{F}_i$ acting on a particle is the *particle-particle method*. In each time step, the force is calculated by

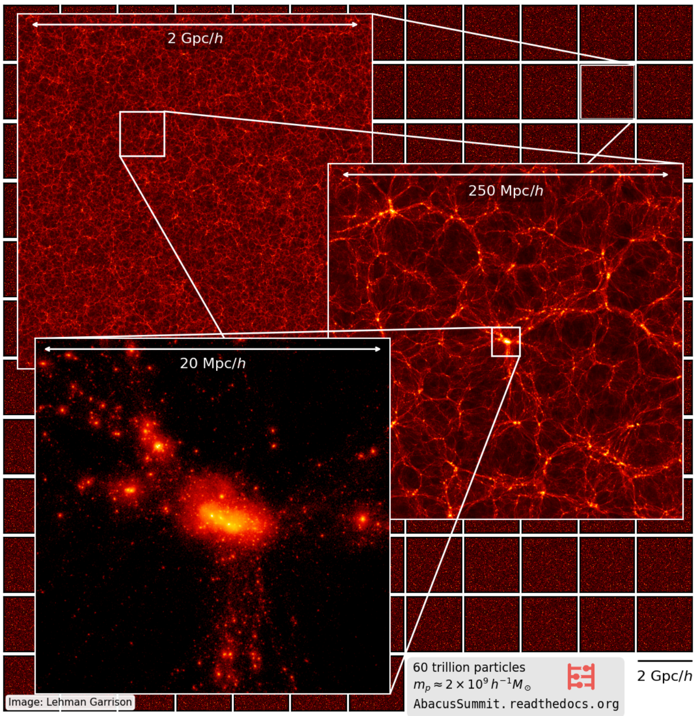

Figure 3.1: A visualization of the AbacusSummit boxes, showing progressive zoom-ins from the full box down to cluster scales. The 139 boxes that comprise the suite are shown as tiles in the background. Each box tracks the evolution of $6912^3 \simeq 330$ billions particles up to $z = 0.1$. Renderings display a snapshot at $z = 0.1$, and projections are 10 h^{-1} Mpc deep. Credits to Maksimova et al. (2021).

summing the contributions of all the particles. This method is the most precise, as it yields the exact values of the forces. However, it is also very computationally demanding, scaling as $\mathcal{O}(N^2)$ (where N is the number of particles). More efficient methods to compute the gravitational interaction are the *hierarchical tree* and the *particle mesh* (see Hockney & Eastwood, 1981; Barnes & Hut, 1986, for details).

The initial studies involving numerical simulations focused on solving the N-body problem for a few hundred particles (Aarseth, 1963; Peebles, 1970). Thanks to significant advancements in technology and computational techniques over recent decades, it is now possible to conduct simulations with particle counts reaching several billions (see Fig. 3.1 for an example). Despite the impressive achievements in this area of research, cosmological simulations still face significant limitations. Simulations with small volumes offer high resolution, which is essential for studying galaxy formation models and resolving detailed physical processes. In contrast, simulations with large volumes allow for an in-depth examination of the LSS of the Universe and facilitate statistical analyses of its properties. Achieving simulations that combine both high resolution and large volume remains difficult.

3.3.1 Halo finders

The final output of an N-body simulation is a set of snapshots that capture the configuration of the particle system (Fig. 3.1) at various redshifts, thereby illustrating the evolution of the total matter density field. In a snapshot, one immediately notices regions where there is an higher concentration of dark matter particles, known as DM halos. To identify these structures, two standard techniques are commonly applied: *spherical overdensity* (`SO`, Press & Schechter, 1974) and *Friends-of-Friends*, (`FOF`, Davis et al., 1985). These techniques have laid the foundation for the development of more advanced halo finding algorithms.

The `SO` method is based on defining spherical regions around density peaks, which are identified by sorting particles based on their local density. Once a density peak is found, a halo is identified by expanding a sphere around it. The expansion stops when the mean density within this sphere reaches the value $\Delta_\mathrm{c}\rho_\mathrm{crit}(z)$, where Δ_c is the chosen overdensity threshold and $\rho_\mathrm{crit}(z)$ is the critical density of the Universe at fixed redshift (Eq. 1.32). Then, the radius and mass of each halo are defined according to the following relation:

$$\frac{4\pi}{3} R^3_{\Delta_\mathrm{c}} \Delta_\mathrm{c} \rho_\mathrm{crit} = M_{\Delta_\mathrm{c}} \, . \tag{3.34}$$

Alternatively, the background density, defined as $\rho_\mathrm{b} \equiv \Omega_\mathrm{m}\rho_\mathrm{crit}$, can also be used as the reference density for determining the mass and radius of DM halos.

On the other hand, the `FOF` algorithm categorizes halos as groups of DM particles that are closer to each other than a specified linking length $\ell = b\,\bar{d}$, where $\bar{d}$ represents the average inter-particle separation in the DM particle data set, and b is a configurable parameter of the algorithm[1]. Closely related to the discussed halo finding techniques is `SUBFIND` (Springel et al., 2001), an algorithm designed to identify substructures within larger parent groups previously identified using a standard `FOF` finder. In this context, substructures are defined as locally overdense regions of particles that are self-bound. Another notable example of a more sophisticated halo finding algorithm is `ROCKSTAR` (Robust Overdensity Calculation using K-Space Topologically Adaptive Refinement, Behroozi et al., 2013), which will be employed to prepare most of the data sets used in the following chapters (see Sect. 5.3.1).

These halo-finding methods provide the foundation for constructing halo catalogs, which are essential for studying the LSS of the Universe. In particular, halos identified in N-body simulations serve as the counterparts of observed galaxy clusters. Therefore, predictions based on numerical simulations can be directly compared with real observational data, allowing to test and constrain cosmological models and parameters (see, e.g., Lesci et al., 2022; Ghirardini et al., 2024).

3.4 Statistical Characterization of the Universe's Density Field

Up to this point, we have examined the linear evolution of an individual perturbation in the density field, characterized by $\delta(\boldsymbol{x}, t) = \delta_+(t)\delta(\boldsymbol{x})$. Nevertheless, the true evolution of structures results from the superposition of density fluctuations across different scales. Utilizing the Fourier space description of these perturbations is particularly advantageous, as it enables us to depict this scenario as a superposition

[1] To construct halo catalogues, a value $b = 0.2$ is typically adopted (More et al., 2011).

of independent plane waves. Therefore, let us introduce the Fourier transform of the real-space density contrast $\delta(\boldsymbol{x})$:

$$\delta(\boldsymbol{k}) = \int \mathrm{d}^3x\, \delta(\boldsymbol{x})\mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{x}} \quad \Longleftrightarrow \quad \delta(\boldsymbol{x}) = \int \frac{\mathrm{d}^3k}{(2\pi)^3}\, \delta(\boldsymbol{k})\mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{x}} \,. \tag{3.35}$$

Given two perturbations characterized by $\delta(\boldsymbol{k})$ and $\delta(\boldsymbol{k}')$, the mean quadratic amplitude of the density fluctuations can be measured using the *matter power spectrum* $P_{\mathrm{m}}(k)$.

To find an expression for it, let us relate the potential on sub-horizon scales and during matter-domination, to the primordial curvature perturbation generated during inflation

$$\Phi(\boldsymbol{k}, a) = \frac{3}{5}\mathcal{R}(\boldsymbol{k})T(k)\frac{\delta_+(a)}{a} \,, \tag{3.36}$$

where $T(k)$ is the transfer function, which describes the evolution of perturbations through the epochs of horizon crossing and radiation/matter transition. Then, we can employ the Poisson equation to relate the matter overdensity $\delta(\boldsymbol{k})$ to the potential, leading to

$$\delta(\boldsymbol{k}, a) = \frac{2k^2 a}{3\Omega_{\mathrm{m}} H_0^2}\Phi(\boldsymbol{k}, a) = \frac{2}{5}\frac{k^2}{\Omega_{0,\mathrm{m}} H_0^2}\mathcal{R}(\boldsymbol{k})T(k)\delta_+(a) = \mathcal{M}(k, z)\mathcal{R}(\boldsymbol{k}) \,. \tag{3.37}$$

Employing the equation for the curvature power spectrum obtained in the previous chapter (Eq. 2.24), we can write the matter power spectrum as

$$P_{\mathrm{m}}(k, a) = \frac{8\pi^2}{25}\frac{A_{\mathrm{s}}}{\Omega_{0,\mathrm{m}}^2}\delta_+^2(a)T^2(k)\frac{k^{n_{\mathrm{s}}}}{H_0^4 k_{\mathrm{p}}^{n_{\mathrm{s}}-1}} \,. \tag{3.38}$$

Since the amplitudes of the fluctuations have a nearly Gaussian distribution in real space, their mean value is statistically zero by definition. Instead, the variance of the fluctuation amplitudes, σ^2, is defined by:

$$\sigma^2 = \langle \delta^2(\boldsymbol{x}) \rangle = \sum_{\boldsymbol{k}} \langle |\delta(\boldsymbol{k})|^2 \rangle = \frac{1}{V_{\mathrm{U}}}\sum_{\boldsymbol{k}} \delta_{\boldsymbol{k}}^2 \,, \tag{3.39}$$

where the average is taken over an ensemble of Universe realizations, each with volume V_{U}. In the limit $V_{\mathrm{U}} \to \infty$ and assuming the validity of the CP, the variance can be written as

$$\sigma^2 = \frac{1}{2\pi^2}\int_0^\infty P_{\mathrm{m}}(k)k^2 \mathrm{d}k \,. \tag{3.40}$$

computing σ^2 requires the evaluation of the density at each point in space, necessitating the reconstruction of the entire density field, which is impractical. A convenient method is to represent the fluctuation field by "filtering" on a scale R. With this approach, we can recover the density fluctuation from a discrete distribution of tracers as follows:

$$\delta_M = \frac{M - \bar{M}}{\bar{M}} \,. \tag{3.41}$$

Here, $\bar{M}$ is the mean mass present inside a spherical volume of radius R. The fluctuation δ_M is directly related to the "standard" density fluctuation. Specifically, it is the convolution of it with a window function (or filter) W:

$$\delta_M = \delta(\boldsymbol{x}) \otimes W(\boldsymbol{x}, R) \,. \tag{3.42}$$

Thanks to the last definitions and to Eq. 3.39, we can obtain the *mass variance* as

$$\sigma_M^2 = \langle \delta_M^2 \rangle = \frac{\langle (M - \bar{M})^2 \rangle}{\bar{M}^2} = \frac{1}{2\pi^2} \int_0^\infty P_{\rm m}(k) k^2 \widehat{W}^2(k, R) \mathrm{d}k \,, \tag{3.43}$$

where $\widehat{W}(k, R)$ is the Fourier transform of the window function and R is the scale associated to the mass M. Typically, a spherical top-hat window function is employed in cosmological analyses, and in Fourier space, it takes the form:

$$\widehat{W}(k, R) = \frac{3\left[\sin(kR) - kR\cos(kR)\right]}{(kR)^3} \,. \tag{3.44}$$

Although inflation theory does not predict the normalization of the power spectrum, a commonly used method is to set the value of the mass variance, computed using a filter with $R = 8\ h^{-1}$ Mpc at the present time

$$\sigma_8^2 = \frac{1}{2\pi^2} \int_0^\infty P_{\rm m}(k) k^2 \widehat{W}^2(k, R = 8\ h^{-1}\ \mathrm{Mpc}) \mathrm{d}k \,. \tag{3.45}$$

The square root of this quantity not only represents the mass fluctuation in spheres with a radius of $8\ h^{-1}$ Mpc, but also serves as a parameter to quantify the amplitude of the power spectrum. However, Sánchez (2020) showed that a more effective approach to characterize the power spectrum is to normalize it by a reference scale measured in physical units, i.e. Mpc. A suitable choice for this scale is 12 Mpc, which approximately corresponds to $8\ h^{-1}$ Mpc for $h \simeq 0.67$, as suggested by current CMB data (Planck Collaboration et al., 2020a). Referring to the corresponding square root of the variance as σ_{12}, it has been shown that this normalization not only better describes the degeneracy between h and $A_{\rm s}$ but also helps to mitigate the growth of structure tension (see Sect. 1.7).

3.4.1 Evolution of the matter power spectrum

Density perturbations entering the cosmological horizon before radiation-matter equality are damped by a phenomenon known as stagnation or Mészáros effect (Meszaros, 1974). This effect arises because the Hubble drag term during the radiation-dominated era is larger than during the matter-dominated era. Specifically, when comparing the free-fall time $\tau_{\rm ff}$, which is the characteristic time for a perturbation to collapse under its own gravitational force, with the Hubble time $\tau_{\rm H}$, which represents the characteristic time for the expansion of the Universe, we observe that

$$\frac{\tau_{\rm H}}{\tau_{\rm ff}} \propto \left(\frac{\rho_{\rm m}}{\rho_{\rm rad}}\right)^{1/2} \ll 1 \quad \text{for} \quad t < t_{\rm eq} \,. \tag{3.46}$$

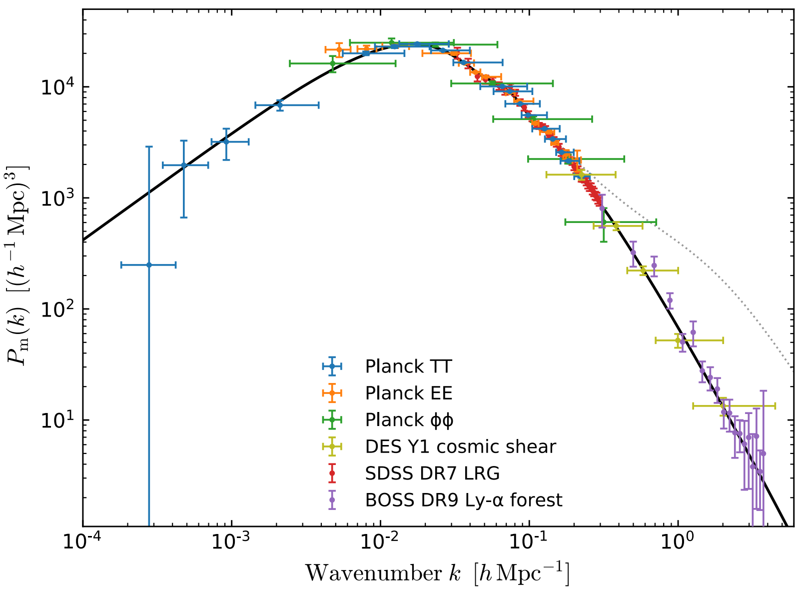

Figure 3.2: The observed linear matter power spectrum $P_{\mathrm{m}}(k)$ at $z = 0$ as a function of the wavenumber k, obtained from a combination of different cosmological probes (the dotted line shows the impact of nonlinear clustering at $z = 0$). The broad agreement of the model (black line) is an impressive testament to the explanatory power of ΛCDM. Credits to Planck Collaboration et al. (2020d).

Since the free-fall time is much longer than the expansion time, density perturbations cannot grow during this period. The main consequence of this effect is a change in the shape of the power spectrum.

As the cosmological horizon expands over time (Eq. 3.2), larger perturbations will enter the horizon at later times, experiencing less stagnation (or no stagnation if they enter the horizon only after the equality). In contrast, perturbations on scales larger than the horizon continue to grow at the same rate regardless of their scale, following the trends described in Eq. 3.5. Consequently, the power spectrum at the time of equivalence exhibits a peak at $k_{\mathrm{H,eq}}$, the wavenumber corresponding to the horizon at that epoch. This peak primarily depends on $\Omega_{0,\mathrm{m}}h^2$ and $\Omega_{0,\mathrm{r}}h^2$, which are related to the matter and radiation densities and the Hubble parameter.

The shape of the observed matter power spectrum $P_{\mathrm{m}}(k, z)$ depends on the amount and nature of matter in the Universe, providing crucial constraints for cosmology. HDM particles remain relativistic at decoupling, whereas CDM particles are non-relativistic before decoupling[2]. Consequently, CDM particles are generally more massive than HDM particles. If the matter component consisted entirely of HDM particles, the power spectrum would drop sharply to zero beyond the peak. Modern observations from the CMB, galaxy clusters, gravitational lensing, and the Lyα forest confirm that the Universe's matter component is predominantly cold, as shown in Fig. 3.2.

The shape of $P_{\mathrm{m}}(k)$ at the time of equivalence depends on the transfer function $T(k)$. In a scenario dominated by CDM, the transfer function assumes the following

[2]In principle, DM is non-collisional, meaning there is no decoupling. However, in scenarios where DM consists of WIMPs or axions, a (minimal) coupling could exist.

forms depending on the wavenumber:

$$T(k) = \begin{cases} 1 \quad \text{for} \quad k < k_{\rm H,eq} \\ \propto k^{-2} \quad \text{for} \quad k > k_{\rm H,eq} \,. \end{cases} \tag{3.47}$$

Therefore, on large scales, where $T(k) = 1$, the power spectrum is proportional to k^{n_s}, while in the large k regime it decays as k^{n_s-4}.

3.5 Bias Theory

DM halos trace the underlying matter distribution, as they are thought to form in regions of high-density peaks. However, the process of gravitational collapse leading to halo formation is highly non-linear and complex. On quasi-linear scales, these complexities can be effectively captured by a series of operators, whose corresponding coefficients are known as bias parameters (Bardeen et al., 1986; Desjacques et al., 2018). Within perturbation theory, the density contrast of halos, as well as other LSS tracers like galaxies or galaxy clusters, can be expressed as a function of position and redshift in the following way:

$$\delta_{\rm h}(\boldsymbol{x}, z) = \sum_O b_O(z) O(\boldsymbol{x}, z). \tag{3.48}$$

In this equation, the terms O represent various fields influencing the matter density field, while the coefficients b_O correspond to the bias parameters associated with each operator. The leading-order term in Eq. 3.48 is given by

$$\delta_{\rm h} = b_1 \, \delta_{\rm m} \,. \tag{3.49}$$

Here, we use b_1 to indicate the effective bias parameter. This relation is only valid within the purely linear regime, typically considered for wave numbers below $k_{\rm max} \lesssim 0.1 \; h \; {\rm Mpc}^{-1}$. Using the bias parameter b_1, the power spectrum of halos can be written in terms of the linear matter power spectrum $P_{\rm m}$ as:

$$P_{\rm h}(k, z) = b_1^2(z) P_{\rm m}(k, z) \,. \tag{3.50}$$

3.5.1 Scale-dependent bias

In the presence of local-type PNG, the bias expansion at leading order (Eq. 3.48) includes an additional term proportional to $f_{\rm NL}^{\rm loc}\Phi$, where Φ is the gravitational potential. Consequently, the halo density contrast can be expressed as

$$\delta_{\rm h}(k, z) = b_1(z)\delta_{\rm m}(k, z) + b_\Phi f_{\rm NL}^{\rm loc}\Phi = \left[b_1(z) + \frac{3 f_{\rm NL}^{\rm loc} b_\Phi(z)}{5\mathcal{M}(k, z)} \right] \delta_{\rm m}(k, z). \tag{3.51}$$

The second equality follows from Eq. 3.37, where $\delta_{\rm m} = \mathcal{M}\mathcal{R} = (5/3)\mathcal{M}\Phi$. Here, b_Φ denotes the bias parameter associated with the operator Φ, which only arises in the presence of PNG.

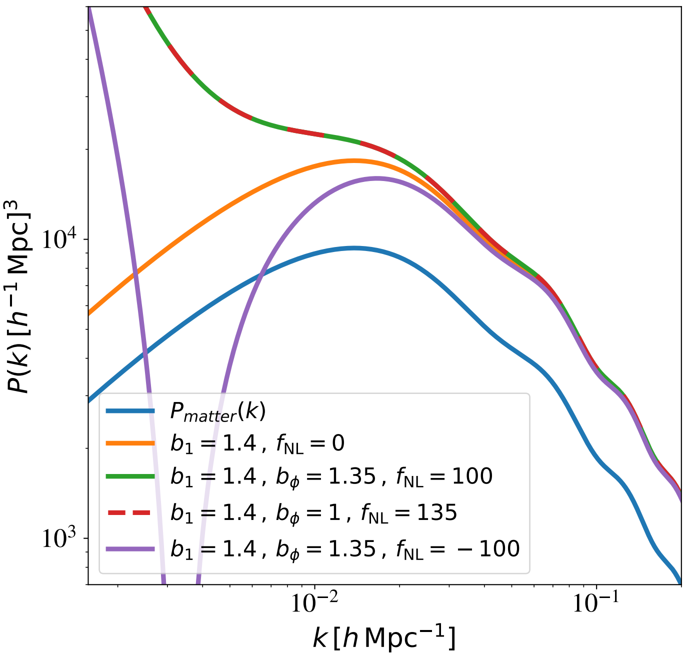

Figure 3.3: Theoretical power spectra obtained from Eq. 3.52. The blue line represents the linear matter power spectrum at $z = 1$. The orange line corresponds to a biased tracer with Gaussian initial conditions. The green line shows the impact of a non-Gaussian contribution with $f_{\rm NL}^{\rm loc} = 100$ and a PNG response parameter b_Φ given by Eq. 3.54 with $p = 1$. The red line demonstrates how a different value of $f_{\rm NL}^{\rm loc}$ can produce the same signal if the b_Φ parameter deviates from the universality relation. Since the product $b_\Phi f_{\rm NL}^{\rm loc}$ is the same, the red and green lines overlap. Finally, the purple line illustrates the expected signal for $b_\Phi f_{\rm NL}^{\rm loc} < 0$. Credits to Gutiérrez Adame et al. (2024).

With this new bias expansion, the power spectrum of DM halos — or any biased tracer of the total matter density — can be derived analogously to Eq. 3.50. The resulting halo power spectrum is given by

$$P_{\rm h}(k,z) = \left[b_1(z) + \frac{3 f_{\rm NL}^{\rm loc} b_\Phi(z)}{5\mathcal{M}(k,z)}\right]^2 P_{\rm m}(k,z). \tag{3.52}$$

The relation between $P_{\rm h}(k,z)$ and $P_{\rm m}(k,z)$ now exhibits a scale dependence due to the $\mathcal{M}(k,z)$ term, which scales as k^{-2}. This effect is commonly referred to as *scale-dependent bias*.

Figure 3.3 illustrates the expected behavior of the halo power spectrum for different combinations of $f_{\rm NL}^{\rm loc}$ and b_Φ. Notably, we observe that different values of $f_{\rm NL}^{\rm loc}$ can lead to identical signals depending on the value of the bias parameter b_Φ, highlighting a perfect degeneracy between $f_{\rm NL}^{\rm loc}$ and b_Φ (Barreira, 2022a).

The parameter b_Φ characterizes the response of halo abundance to perturbations in the primordial gravitational potential Φ. Since local-type PNG couples long- and short-wavelength perturbations, in the framework of the peak-background split theory (Kaiser, 1984; Bardeen et al., 1986), it can be shown that b_Φ also describes the response of the halo number density to variations in the amplitude of the primordial power spectrum (Desjacques et al., 2018; Barreira, 2022b):

$$b_\phi = \frac{1}{\bar{n}_{\rm h}} \frac{\partial \bar{n}_{\rm h}}{\partial \log\left(f_{\rm NL}^{\rm loc}\Phi\right)} = \frac{1}{\bar{n}_{\rm h}} \frac{\partial \bar{n}_{\rm h}}{\partial \log A_{\rm s}} \,, \tag{3.53}$$

where $\bar{n}_{\rm h}$ is the mean halo number density and $A_{\rm s}$ is the amplitude of the primordial scalar perturbations. Assuming a universal mass function, a theoretical prediction for b_Φ can be derived (Dalal et al., 2008; Slosar et al., 2008; Desjacques et al., 2018):

$$b_\Phi = 2\delta_{c,0}\left(b_1 - p\right) \,. \tag{3.54}$$

Here, $\delta_{c,0} = 1.686$ is the critical overdensity for spherical collapse (see Sect. 4.1), and p depends on the tracer population. Under the assumption of a universal mass function, $p = 1$ for all tracers of the matter distribution (Dalal et al., 2008; Slosar et al., 2008). This expression, with $p = 1$, is commonly referred to as the *universality relation* for b_Φ.

However, recent studies suggest that this relation does not always accurately describe the scale-dependent bias. For instance, Barreira et al. (2020) found that for galaxies selected by stellar mass, $p = 0.5$ provided a better fit for b_Φ. Similarly, Gutiérrez Adame et al. (2024) analyzed halos within a specific mass range and obtained $p = 0.955 \pm 0.013$, revealing a $\sim 3\sigma$ deviation from the universal relation. Additionally, b_Φ may depend on secondary properties beyond mass, such as halo concentration (Lazeyras et al., 2023; Fondi et al., 2024).

Chapter 4

Statistical Properties of Dark Matter Halos

In cosmological simulations, as well as in observation data, the LSS of the Universe reveals itself as a filamentary network of matter known as the *cosmic web*. This structure is characterized by galaxies that, through gravitational interactions, aggregate into dense clusters within the most overdense regions, which correspond to the most massive DM halos. By examining the statistical properties of galaxies and galaxy clusters as their number densities and spatial distributions, we can assess and refine cosmological models and constrain their key cosmological parameters.

In this chapter, we will provide an overview of the *Press-Schechter formalism* (PS, Press & Schechter, 1974) that will allow us to derive key properties of DM halos. Specifically, we will examine how this formalism aids in understanding the distribution of DM halos based on their mass. We will also address how these predictions do not align with data and introduce more advanced models to rectify these discrepancies. As these results are based on a Gaussian distribution of primordial density fluctuations, we will also introduce their modification induced by PNG.

4.1 Press-Schechter Formalism

According to linear theory, a generic density field evolved as $\delta(\boldsymbol{x}, t) = \delta_+(t)\delta_0(\boldsymbol{x})$, where $\delta_0(\boldsymbol{x})$ is the density field linearly extrapolated at $t = t_0$. As discussed in Sect. 3.2.1, regions with overdensities exceeding the threshold $\delta_\mathrm{c} \simeq 1.686$ have collapsed to produce DM halos by time t. By rearranging terms, we can also interpret this as follows: regions with $\delta_0(\boldsymbol{x}) > \delta_\mathrm{c}/\delta_+(t)$ have collapsed to produce halos by time t. In this case, we treat the density field as static (at the value extrapolated for the time t_0), while the *collapse barrier* evolves over time. Let δ_M be the linear density field smoothed on a mass scale M. In principle, locations where $\delta_M = \delta_\mathrm{c}(t)$ indicate the regions where, at time t, a halo of mass M condensed out of the evolving density field.

In this scenario, the *halo mass function* is determined by assessing the number density of peaks (Fig. 4.1) within the smoothed density field:

$$n(> M) = n_\mathrm{pk}(\delta_M)\,, \tag{4.1}$$

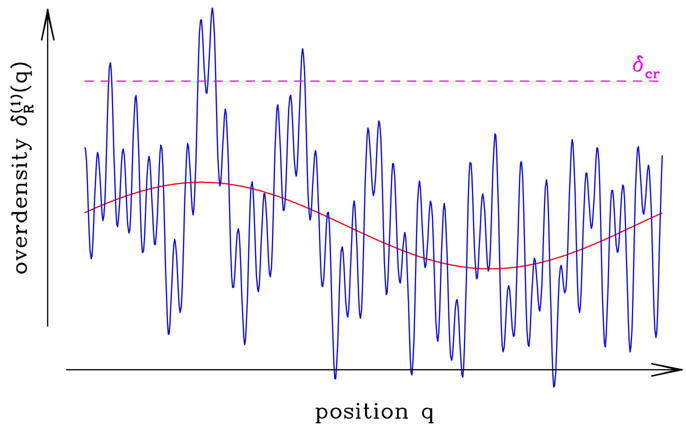

Figure 4.1: The solid blue line shows the 1D smoothed density field δ_M, while the red line indicates a long-wavelength perturbation. The dashed horizontal line marks the linear threshold for spherical collapse. Credits to Desjacques et al. (2018).

where $n(> M)$ is the number density of halos with mass above M and $n_{\rm pk}(\delta_M)$ is the number density of peaks above $\delta_{\rm c}$ in the smoothed density field. This idea was firstly investigated by Bardeen et al. (1986). However, it soon encountered the issue that some peaks — specifically, those that become part of a higher peak when smoothed with a larger filter — need to be excluded when associating peaks with halos. This issue is known as the *cloud-in-cloud problem*. As a result, the peak formalism was abandoned in favor of PS formalism, which, while less rigorous, proved to be more successful.

In this framework, it is assumed that the probability of $\delta_M > \delta_{\rm c}(t)$ corresponds to the mass fraction, at time t, contained in halos with mass greater than M. For a Gaussian random field, one finds that this probability is given by

$$\mathcal{P}(\delta_M > \delta_{\rm c}) = \int_{\delta_{\rm c}}^{+\infty} P(\delta_M)\mathrm{d}M = \frac{1}{\sqrt{2\pi\sigma_M^2}} \int_{\delta_{\rm c}}^{+\infty} \exp\left(-\frac{\delta_M^2}{2\sigma_M^2}\right) \mathrm{d}\delta_M \,, \tag{4.2}$$

with $P(\delta_M)\mathrm{d}\delta_M$ being the probability distribution function (PDF). In the limit $M \to 0$, the variance diverges, and thus, the probability is expected to approach unity, implying that all the matter should be part of collapsed structures. However, the PS formalism predict a value of $1/2$, suggesting that only half of it is locked-up in collapsed structures. This discrepancy is the manifestation of the cloud-in-cloud problem and it was addressed by introducing an adjustment factor of 2 into the probability, which corrected the predicted value to align with expectations:

$$F(> M) = 2\mathcal{P}(\delta_M > \delta_{\rm c})\,. \tag{4.3}$$

By utilizing the extended PS formalism, also known as *excursion-set*, it is possible to derive the factor of 2 directly from the theoretical framework, eliminating the need for any adjustment factors (see, Zentner, 2007, for a review of the topic).

4.1.1 Press-Schechter halo mass function

The halo mass function (HMF) can be formally defined as the number of halos per unit comoving volume, with masses in the range between M and $M + \mathrm{d}M$. Given that $(\partial F/\partial M)\mathrm{d}M$ represents the fraction of mass contained in halos with masses between M and $M + \mathrm{d}M$, the HMF can be expressed as

$$\frac{\mathrm{d}n(M,z)}{\mathrm{d}M}\mathrm{d}M = \frac{\bar{\rho}_\mathrm{m}}{M}\frac{\partial F(>M)}{\partial M}\mathrm{d}M\,, \tag{4.4}$$

where $\bar{\rho}_\mathrm{m}$ indicates the mean comoving matter density. Substituting Eq. 4.3 into this equation yields the general expression for the Press-Schechter HMF:

$$\frac{\mathrm{d}n(M,z)}{\mathrm{d}M} = \sqrt{\frac{2}{\pi}}\frac{\bar{\rho}_\mathrm{m}}{M^2}\frac{\delta_\mathrm{c}}{\sigma_M}\left|\frac{\mathrm{d}\ln\sigma_M}{\mathrm{d}\ln M}\right|\exp\left(-\frac{\delta_\mathrm{c}^2}{2\sigma_M^2}\right). \tag{4.5}$$

Alternatively, the HMF can be expressed more compactly as follows

$$\frac{\mathrm{d}n}{\mathrm{d}M} = f(\sigma_M)\frac{\bar{\rho}_\mathrm{m}}{M}\frac{\mathrm{d}\ln\sigma_M^{-1}}{\mathrm{d}M}\,, \tag{4.6}$$

where we have introduced the so-called *multiplicity function* $f(\sigma_M)$. It is common to define this function also in terms of the variable $\nu \equiv \delta_\mathrm{c}/\sigma_M$. In principle, it is dependent on the specific HMF model, as we will see in Sect. 4.2. For the PS scenario, it can be straightforwardly derived as

$$f(\sigma_M) = \sqrt{\frac{2}{\pi}}\frac{\delta_\mathrm{c}}{\sigma_M}\exp\left(-\frac{\delta_\mathrm{c}^2}{2\sigma_M^2}\right) = \sqrt{\frac{2}{\pi}}\nu\exp\left(-\frac{\nu^2}{2}\right). \tag{4.7}$$

4.2 Accurate Halo Mass Function Models

The PS approach combines the assumption of spherical collapse with the notion that initial density fluctuations follow a Gaussian distribution. However, this model does not accurately describe the data from cosmological simulations. In particular, the model tends to overestimate the abundance of high-mass halos while underestimating the number of low-mass halos (Sheth & Tormen, 1999).

Galaxy clusters have proven to be powerful probes of cosmology, leading to the construction of several large-scale surveys. To fully leverage the statistical power of recent large-scale surveys, we must be able to make accurate predictions for abundance evolution as a function of cosmological parameters.

In previous works, it has been proposed that the functional form of the HMF, when expressed in appropriate variables, should be universal across different redshifts and cosmologies (Bond et al., 1991; Sheth & Tormen, 1999). Although this universality is expected to be only approximate (Musso & Sheth, 2012; Paranjape et al., 2013), it greatly simplifies the task of constraining cosmological parameters from observational data. As a result, this idea has formed the foundation for fitting functions whose parameters are calibrated using N-body simulations (Sheth & Tormen, 1999; Jenkins et al., 2001; Warren et al., 2006). In this regard, Fig. 4.2 shows an example of a fitted universal HMF alongside some of the data points used to build the model. However, using a general fitting function, the collapse threshold δ_c is no longer directly tied to spherical collapse. Nevertheless, spherical collapse and the excursion-set offer a plausible physical explanation of the phenomenon. As

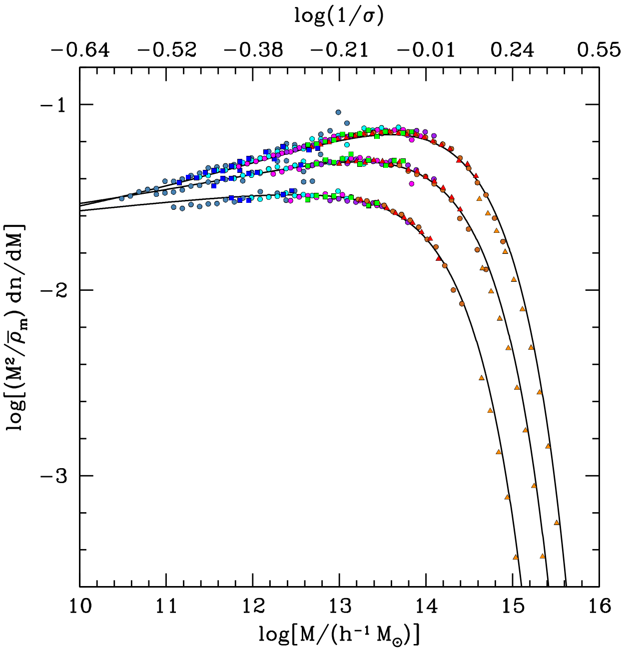

Figure 4.2: The halo mass function (multiplied by $M^2/\bar{\rho}_{\rm m}$) as measured in a large suite of N-body simulations. The three sets of points correspond to different mass definitions $M_{\Delta_{\rm b}}$, with $\Delta_{\rm b} = 200$, 800, and 3200 (from top to bottom). Markers and colors distinguish data points extracted from simulations with different resolution. Finally, the black solid lines represent the best-fit model for the mass function at the three different overdensities. Credits to Tinker et al. (2008).

the simulated data sets have grown, it has also become feasible to quantify small deviations from universality (Tinker et al., 2008; Crocce et al., 2010; Despali et al., 2016). Currently, modern models calibrated on numerical simulations achieve an accuracy of a few percent.

Starting from Eq. 4.6, the various HMFs are derived by adjusting the multiplicity function. In Table 4.1, we present the expressions of $f(\sigma_M)$ for some of the most well-known and widely-used models in literature.

4.3 The Effect of local-type PNG on the Halo Mass Function

In order to take into account the effect of local-type PNG on the smoothed density field, let us consider a general PDF $P(\delta_M)\mathrm{d}M$. The n-central moment for the PDF is defined as

$$\langle \delta_M^n \rangle = \int_{-\infty}^{+\infty} \delta_M^n P(\delta_M)\mathrm{d}\delta_M \,, \tag{4.8}$$

while each reduced p-th cumulant can be defined as

$$S_p(M) = \frac{\langle \delta_M^p \rangle}{\langle \delta_M^2 \rangle^{p-1}} \,. \tag{4.9}$$

Model	Multiplicity Function $f(\sigma_M)$
Press-Schechter	$f = \sqrt{\frac{2}{\pi}}\frac{\delta_c}{\sigma_M}\exp\left(-\frac{\delta_c^2}{2\sigma_M^2}\right)$
Sheth-Tormen	$f = A\sqrt{\frac{2a}{\pi}}\frac{\delta_c}{\sigma_M}\exp\left(-\frac{a\delta_c^2}{2\sigma_M^2}\right)\left[1+\left(\frac{\sigma_M^2}{a\delta_c^2}\right)^p\right]$
Warren et al.	$f = A\left(\sigma_M^{-a}+b\right)\exp\left(-\frac{c}{\sigma_M^2}\right)$
Tinker et al.	$f = A\left[\left(\frac{\sigma_M}{b}\right)^{-a}+1\right]\exp\left(-\frac{c}{\sigma_M^2}\right)$

Table 4.1: Multiplicity function for some of the various HMF models present in literature. The parameters entering the function in the last three cases are derived from fits to data obtained from various simulation sets (for details on the models, see Press & Schechter, 1974; Sheth & Tormen, 1999; Warren et al., 2006; Tinker et al., 2008).

In particular, $\langle\delta_M^2\rangle \equiv \sigma_M^2$, while S_3 and S_4 are referred to as normalized *skewness* and *kurtosis* of the distribution, respectively. For example, the cumulant S_3 can be expressed as follows:

$$S_3(M) = \frac{1}{\sigma_M^4}\int\frac{\mathrm{d}^3k_1}{(2\pi)^3}\frac{\mathrm{d}^3k_2}{(2\pi)^3}\frac{\mathrm{d}^3k_3}{(2\pi)^3}\widehat{\mathcal{M}}(k_1)\widehat{\mathcal{M}}(k_2)\widehat{\mathcal{M}}(k_3)\langle\mathcal{R}(k_1)\mathcal{R}(k_2)\mathcal{R}(k_3)\rangle\,, \tag{4.10}$$

with $\widehat{\mathcal{M}}(k,z) \equiv \mathcal{M}(k,z)\widehat{W}(kR)$, where $\mathcal{M}$ and $\widehat{W}$ are defined in Eqs. 3.37 and 3.44, respectively. Analogous formulae are valid for higher-order cumulants. Therefore, the normalized skewness is related to the bispectrum from Eq. 2.26 and is proportional to the parameter $f_{\mathrm{NL}}^{\mathrm{loc}}$ (hereafter f_{NL}), assuming a local shape. Similarly, we obtain that the kurtosis S_4 depends on the trispectrum of the distribution and is proportional to f_{NL}^2.

We know from Sect. 4.1 that, under the assumption of Gaussianity, from the PDF we can recover the standard PS mass function (Eq. 4.5). Therefore, let us consider a non-Gaussian PDF of the matter density fluctuations, based on the concept of the Edgeworth expansion. Specifically, we write the PDF $P(x)\mathrm{d}x$, with $x = \delta_M/\sigma_M$, in terms of the derivatives of the Gaussian one, $P_{\mathrm{G}}(x)$, as (Juszkiewicz et al., 1995; LoVerde et al., 2008)

$$P(x)\mathrm{d}x = \mathrm{d}x\left[c_0 P_{\mathrm{G}}(x) + \sum_{m=1}\frac{c_m}{m!}P_{\mathrm{G}}^{(m)}(x)\right]\,, \tag{4.11}$$

where we define P_{G} and $P_{\mathrm{G}}^{(m)}$ as

$$P_{\mathrm{G}}(x) \equiv \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)\,, \tag{4.12}$$

$$P_{\mathrm{G}}^{(m)}(x) \equiv \frac{\mathrm{d}^m}{\mathrm{d}x^m}P_{\mathrm{G}}(x) = (-1)^m H_m(x)P_{\mathrm{G}}(x)\,. \tag{4.13}$$

Here, H_m indicates the Hermite polynomials of order m. Therefore, we can regard Eq. 4.11 as a non-Gaussian PDF expanded in terms of the Hermite polynomials. Since they satisfy the orthogonal relations, we can evaluate the coefficients c_m as

$$c_m = (-1)^m \int_{-\infty}^{+\infty} H_m(x) P(x) \mathrm{d}x \,. \tag{4.14}$$

Then, we obtain the expressions for the coefficients c_m in terms of the mass variance and of the cumulants S_3 and S_4. As a result, the non-Gaussian PDF of the density field can be written as

$$\begin{aligned} P(x)\mathrm{d}x = \frac{\mathrm{d}x}{\sqrt{2\pi}} \exp\left(\frac{x^2}{2}\right) \bigg[1 + \frac{S_3\sigma_M}{6} H_3(x) + \frac{1}{2}\left(\frac{S_3\sigma_M}{6}\right) H_6(x) + \\ + \frac{S_4\sigma_M^2}{24} H_4(x) + \frac{1}{2}\left(\frac{S_4\sigma_M^2}{24}\right)^2 H_8(x) + ...\bigg] \,, \end{aligned} \tag{4.15}$$

up to the second-order terms in S_3 and S_4, and neglecting the contribution of higher-order cumulants, that are expected to be negligible.

By inserting this equation for the non-Gaussian PDF into the computation of the PS mass function, we obtain the fundamental relation

$$\frac{\mathrm{d}n_{\mathrm{PS}}^{\mathrm{NG}}}{\mathrm{d}M}(M,z) = \frac{\mathrm{d}n_{\mathrm{PS}}}{\mathrm{d}M}(M,z)\mathcal{C}(M,z) \,. \tag{4.16}$$

Thus, the non-Gaussian mass function can be simply written as the Gaussian one multiplied by a correction factor. It is typical to assume that the correction factor is independent of the model of Gaussian HMF, therefore, once it is computed, it can be applied to any known model, such as those introduced in Sect. 4.2.

At 2nd-order in S_3 and in terms of the variable $\nu \equiv \delta_\mathrm{c}/\sigma_M$, the correction factor found by LoVerde et al. (2008) (**LMSVQ** hereafter) can be written as

$$\begin{aligned} \mathcal{C}_{\mathrm{LMSVQ}}(M,z) = 1 + \frac{S_3\sigma_M}{6}\left[H_3(\nu) + \frac{1}{\nu}\frac{\mathrm{d}\ln(S_3\sigma_M)}{\mathrm{d}\ln\sigma_M} H_2(\nu)\right] + \frac{(S_3\sigma_M)^2}{72} \times \\ \times \left[H_6(\nu) + \frac{2}{\nu}\frac{\mathrm{d}\ln(S_3\sigma_M)}{\mathrm{d}\ln\sigma_M} H_5(\nu)\right] + \frac{S_4\sigma_M^2}{24}\left[H_4(\nu) + \frac{1}{\nu}\frac{\mathrm{d}\ln(S_4\sigma_M^2)}{\mathrm{d}\ln\sigma_M} H_3(\nu)\right] \,. \end{aligned} \tag{4.17}$$

In literature, other models have been proposed to account for the effect of local-type PNG on the HMF. For, example, in Matarrese et al. (2000) (**MVJ** hereafter), the authors employed the saddle-point approximation to calculate the level excursion probability, while in D'Amico et al. (2011a) (**DMNP** hereafter), they derived the non-Gaussian halo mass function by calculating the first-crossing rate of a random walk with non-Gaussian noise in the presence of an absorbing barrier. Their correction factors can be written respectively as

$$\mathcal{C}_{\mathrm{MVJ}}(M,z) = \exp\left(\delta_\mathrm{c}^3 \frac{S_3}{6\sigma_M^2}\right)\left[\frac{1}{6}\frac{\delta_\mathrm{c}}{\sqrt{1-\delta_\mathrm{c}S_3/3}}\frac{\mathrm{d}S_3}{\mathrm{d}\ln\sigma_M} + \sqrt{1-\frac{\delta_\mathrm{c}S_3}{3}}\right] \,, \tag{4.18}$$

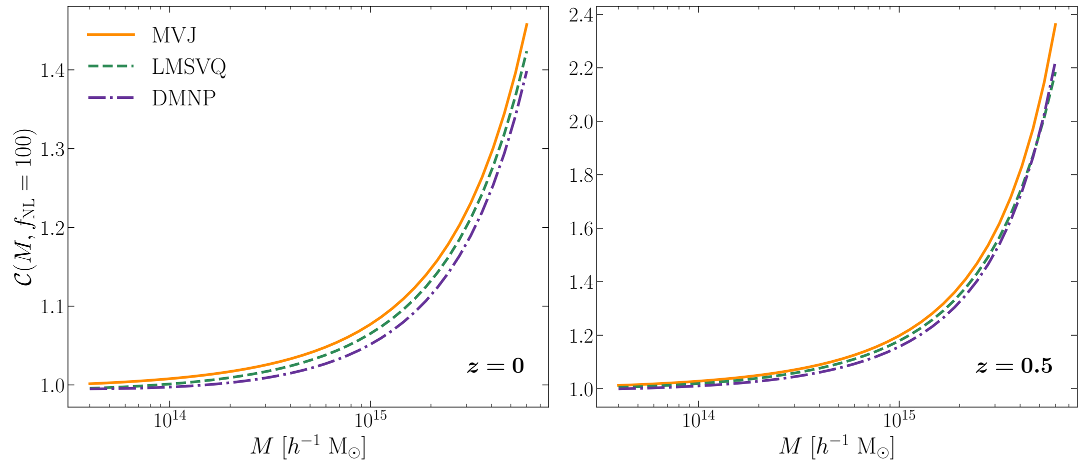

Figure 4.3: The non-Gaussian correction factors $\mathcal{C}$ for the halo mass function computed assuming $f_{\rm NL} = 100$ in the mass range from 4×10^{13} to 6×10^{15} h^{-1} M$_\odot$. Different colors and line styles are used to indicate the different theoretical models: solid orange line for MVJ (Eq. 4.18), green dashed line for LMSVQ (Eq. 4.17) and purple dotted line for DMNP (Eq. 4.19). Each panel refer to a different redshift, with the *left* one corresponding to $z = 0$, and the *right* one to $z = 0.5$.

$$\mathcal{C}_{\rm DMNP}(M, z) = \exp\left(\frac{\varepsilon_1 \nu^3}{6} - \frac{\nu^4}{8}\left(\varepsilon_1^2 - \frac{\varepsilon_2}{3}\right)\right)\left\{1 - \frac{\varepsilon_1 \nu}{4} \times \right. \\ \left. \times \left[(4 - c_1) + \frac{1}{\nu^2}\left(c_1 - \frac{c_2}{4} - 2\right)\right]\right\}, \tag{4.19}$$

where, in the second equation, the coefficients c_1 and c_2 are smoothly varying functions of the variance σ_M^2. Furthermore, we define the functions ε_n, known as the "equal-time" functions, where $\varepsilon_{n-2} \equiv \langle \delta_M^n \rangle / \langle \delta_M^2 \rangle^n$ for $n \geq 3$. It is straightforward to show that the equal-time functions ε_1 and ε_2 are related to the normalized skewness and kurtosis by $\varepsilon_1 = \sigma_M S_3$ and $\varepsilon_2 = \sigma_M^2 S_4$.

In this Thesis, we will focus on the aforementioned LMSVQ, MVJ and DMNP models, as they provide a prescription that is well-known and widely used in the literature to describe the deviations caused by PNG on the Gaussian HMF. As shown in Fig. 4.3, they predict that positive values of the parameter $f_{\rm NL}$ lead to an increase in the halo number counts at high masses and a decrease at low masses. The specific mass scale at which this change occurs varies with the model; for example, the MVJ model predicts that this turnover occurs at lower masses in comparison to the other models considered in this work. In contrast, negative values of $f_{\rm NL}$ are expected to produce the opposite effect. Finally, we observe that all the correction factors converge to unity in the limit $f_{\rm NL} \to 0$, naturally recovering the standard Gaussian model for the HMF.

4.4 Density Profiles of Dark Matter Halos

The density profile of a DM halo describes how the density of DM varies as a function of distance from the center of the halo. This profile can be calibrated from cosmological N-body simulations. A commonly used 2-parameter function that accurately describes the DM halo density profile is the Navarro-Frenk-White (NFW) profile (Navarro et al., 1997), in which the density is given by

$$\rho_{\rm NFW}(r) = \frac{\delta_{\rm c}\rho_{\rm crit}}{(r/r_{\rm s})(1+r/r_{\rm s})^2}\,, \tag{4.20}$$

where $r_{\rm s}$ is the scale radius and $\delta_{\rm c}$ is the characteristic overdensity of the halo. The former is related to the so-called *concentration* c, such that $r_{\rm s} = r_{\Delta_{\rm c}}/c$, where $r_{\Delta_{\rm c}}$ is the radius at which the average density of the halo is $\Delta_{\rm c}$ times the critical density of the Universe. Usually a reference value of $\Delta_{\rm c} = 200$ is used, but it is also common to employ the virial radius $r_{\rm vir}$.

Note that $r_{\rm 200c}$ determines the mass of the halos and that $\delta_{\rm c}$ and c are linked by the requirement that the mean density within $r_{\rm 200c}$ should be $200\rho_{\rm crit}$. That is

$$\delta_{\rm c} = \frac{200}{3}\frac{c^3}{[\ln(1+c) - c/(1+c)]}\,. \tag{4.21}$$

The concentration of DM halos has been found to correlate with the mass of the halo: at fixed redshift, the concentration decreases for increasing mass. This is the so-called *concentration-mass relation* (see, e.g., Diemer & Kravtsov, 2015; Klypin et al., 2016; Child et al., 2018).

This form of the density profile is seen to be universal for different masses and geometries of the Universe. From Eq. 4.20, we find that ρ is proportional to r^{-1} in the innermost regions of the halo, and proportional to r^{-3} in the outer regions.

Another model that has been widely used to describe the density profile of halos is the Einasto profile (Einasto, 1965):

$$\rho(r) = \rho_{\rm s}\exp\left\{-\frac{2}{\alpha}\left[\left(\frac{r}{r_{\rm s}}\right)^{\alpha} - 1\right]\right\}. \tag{4.22}$$

Here, $\rho_{\rm s} \equiv \rho(r_{\rm s})$ is the scale density and α is the Einasto index, which determines the shape of the profile. Thus, this profile is characterized by 3 parameters. Note that in this case we do not have a divergent solution in the limit $r \to 0$, contrary to the NFW profile. Moreover, this profile has the same functional form of the Sersic's law (Sérsic, 1963), which is used to describe the surface brightness profile of galaxies.

Finally, we emphasize that these functional forms for the density profiles of DM halos have been derived from simulations that feature a standard ΛCDM scenario. However, in the presence of PNGs, deviations in the density profiles may arise, as suggested in previous studies (Smith et al., 2011; Baldi et al., 2024). We will investigate these deviations in Sect. 6.3.2.

Chapter 5

Cosmological N-body Simulations and Numerical Tools

Nowadays, numerical simulations are a fundamental tool in cosmology, enabling the testing and validation of theoretical models by replicating the evolution of LSS under a wide range of physical conditions (Angulo & Hahn, 2022). These simulations offer a powerful framework for exploring the impact of various cosmological parameters — such as those characterizing PNGs — on the formation and evolution of cosmic structures. In this chapter, we provide an overview of the numerical N-body simulations used to analyze the effects of local-type PNG on DM halos. Specifically, we will describe a total of four simulation sets, highlighting their key characteristics, and comparing their similarities and differences. Finally, we will present the numerical tools employed for constructing halo catalogs, extracting and analyzing summary statistics, and performing a comprehensive Bayesian analysis.

5.1 The Quijote and Quijote-PNG Simulations

The most effective way to test the models discussed in Sect. 4.3 is by utilizing numerical N-body simulations that incorporate various levels of local-type PNG, described by specific values of the parameter $f_{\rm NL}$. For this purpose, we utilize the Quijote simulations[1] (Villaescusa-Navarro et al., 2020), a suite consisting of more than 82 000 full N-body simulations. Specifically, we focus on the set of simulations that features a standard ΛCDM model, as well as the set that includes PNG, known as Quijote-png (Coulton et al., 2023). These simulations were generated using the codes `2LPTIC` (Crocce et al., 2006) and `2LPTPNG`[2] (Scoccimarro et al., 2012) to produce the initial conditions at $z = 127$ and `GADGET-3` — an enhanced version of `GADGET-2` (Springel, 2005) — to follow their evolution up to $z = 0$.

Generating initial conditions with local non-Gaussianity follows a straightforward procedure. The process begins by generating the modes of a Gaussian primordial potential field, $\Phi(\boldsymbol{k})$, in Fourier space using the input power spectrum. This field is then transformed back to real space via an inverse Fourier transform. To introduce

[1] https://quijote-simulations.readthedocs.io

[2] https://github.com/dsjamieson/2LPTPNG

non-Gaussianity of the local type, the real-space field is squared, mean-subtracted, scaled by the chosen amplitude f_{NL}, and added to the original potential field. Finally, the modified potential field is transformed back to Fourier space, yielding the modes of the primordial potential that now include local-type PNG. In Fourier space, this procedure can be expressed as a convolution, leading to the following relation:

$$\begin{aligned}\Phi^{\mathrm{loc}}(\boldsymbol{k}) = \Phi(\boldsymbol{k}) + f_{\mathrm{NL}} \int \frac{\mathrm{d}^3 k_1}{(2\pi)^3} \frac{\mathrm{d}^3 k_2}{(2\pi)^3} \Phi(\boldsymbol{k}_1)\Phi(\boldsymbol{k}_2)(2\pi)^3 \delta_{\mathrm{D}}^{(3)}(\boldsymbol{k}_1 + \boldsymbol{k}_2 + \boldsymbol{k}) - \\ - f_{\mathrm{NL}} \int \frac{\mathrm{d}^3 k}{(2\pi)^3} \Phi(\boldsymbol{k})\Phi^*(\boldsymbol{k}) \,,\end{aligned} \tag{5.1}$$

where the last term removes the mean contribution $\langle \Phi^2 \rangle$, preventing an artificial modification of the background expansion. Once the modified potential is obtained, the initial conditions are generated by constructing primordial anisotropies from a power-law spectrum characterized by the amplitude A_{s} and spectral tilt n_{s}. These anisotropies are computed on a 1024^3 grid to reduce aliasing effects, and non-Gaussianity of the local type is introduced through Eq. 5.1. The perturbations are evolved to redshift $z = 0$ using the linear transfer function $T(k)$. Then, their amplitude is rescaled to $z = 127$ using the standard growth factor $\delta_+(z)$, using the following relation

$$P_{\mathrm{m}}(k, z = 127) = \frac{\delta_+^2(z = 127)}{\delta_+^2(z = 0)} P_{\mathrm{m}}(k, z = 0) \,, \tag{5.2}$$

which is derived from Eq. 3.38. Finally, the gridded field at $z = 127$ is used with the 2^{nd}-*order Lagrangian perturbation theory* (2LPT, Jenkins, 2010) to generate the initial particle displacements for the simulations.

For the purposes of our analysis, we utilize 8000 realizations for the ΛCDM cosmology and all 1000 realizations featuring local-type PNG (500 for $f_{\mathrm{NL}} = 100$ and 500 for $f_{\mathrm{NL}} = -100$). Each realization follows the evolution of 512^3 CDM particles in a periodic box with size $L_{\mathrm{box}} = 1\ h^{-1}$ Gpc, which translates into a mass resolution of $6.56 \times 10^{11}\ h^{-1}\ \mathrm{M}_\odot$. Both sets of simulations are characterized by the same cosmological parameters: $\Omega_{0,\mathrm{m}} = 0.3175$, $\Omega_{0,\mathrm{b}} = 0.049$, $h = 0.6711$, $n_{\mathrm{s}} = 0.9624$, and $\sigma_8 = 0.834$. The only difference is the parameter f_{NL}, which is set to 100 and -100 in the simulations that include PNG. With the exception of f_{NL}, the cosmological parameters are consistent with the constraints of *Planck* 2015 (Planck Collaboration et al., 2016). The main characteristics of the simulation suites are summarized in Table 5.1.

5.2 The PRINGLS Set

In addition to these publicly available suites, we ran another set of complementary simulations, maintaining a setup as similar as possible to that used in Quijote. Specifically, these new N-body simulations, referred to as Pringls (PRImordial Non-Gaussianity of Local-type Simulations), share most of the characteristics of the previously mentioned simulations, including cosmological parameters. The main

	IC	$f_{\rm NL}$	$N_{\rm real}$	$L_{\rm box}$ [h^{-1} Mpc]	$N_{\rm part}$	$M_{\rm DM}$ [h^{-1} $M_\odot$]
QUIJOTE	2LPTIC	0	8000	1000	512^3	6.56×10^{11}
QUIJOTE-PNG	2LPTPNG	± 100	2×500	=	=	=
PRINGLS	PNGRUN	$0, \pm 40, \pm 100$	5×10	=	=	=
PRINGLS-HR	PNGRUN	$0, \pm 100$	3×1	=	1024^3	8.21×10^{10}

Table 5.1: *From left to right*: Main characteristics of the N-body simulation sets employed in this work: the code used to generate the initial conditions, the values of the parameter $f_{\rm NL}$, the number of realizations, their box size, the number of particles, and the DM particle mass.

differences lie in the code used to generate the initial conditions, the number of realizations, and the values of $f_{\rm NL}$. Local-type PNG are introduced in the initial conditions using the `PNGRUN` code (Wagner et al., 2010), with 10 realizations for each cosmological scenario. Furthermore, we include two additional values of $f_{\rm NL}$ (-40 and 40) compared to QUIJOTE-PNG.

The code `PNGRUN` employed to determine initial conditions for these simulations is based on the computation of the non-Gaussian contribution $\Phi^{\rm NG}$ to the gravitational potential $\Phi = \Phi^{\rm G} + \Phi^{\rm NG}$ starting from the desired bispectrum $B(k_1, k_2, k_3)$ of the potential field, which is defined in Fourier space as

$$\langle \Phi(k_1)\Phi(k_2)\Phi(k_3)\rangle = (2\pi)^3 \delta^D(\boldsymbol{k}_1 + \boldsymbol{k}_2 + \boldsymbol{k}_3) B(k_1, k_2, k_3)\,, \tag{5.3}$$

where Φ_k is the Fourier transform of the real-space potential Φ. As we have seen in Sect. 2.4.1, in the local approximation and in terms of the potential, the bispectrum takes the form:

$$B(k_1, k_2, k_3) = 2 f_{\rm NL} \left[P(k_1)P(k_2) + P(k_2)P(k_3) + P(k_1)P(k_3)\right] . \tag{5.4}$$

The non-Gaussian component of the gravitational potential can be computed as

$$\Phi^{\rm NG}_{\boldsymbol{k}} = \frac{1}{6} \sum_{\boldsymbol{k}'} B(k, k', |\boldsymbol{k} + \boldsymbol{k}'|) \frac{\Phi^{*\rm G}_{\boldsymbol{k}'}}{P(k')} \frac{\Phi^{\rm G}_{\boldsymbol{k}+\boldsymbol{k}'}}{P(|\boldsymbol{k} + \boldsymbol{k}'|)}\,, \tag{5.5}$$

where $\Phi^{\rm G}_{\boldsymbol{k}}$ is a random realization of a Gaussian field with the power spectrum given by $P(k) \propto k^{n_s - 4}$. Once the Fourier-space non-Gaussian potential $\Phi^{\rm NG}_{\boldsymbol{k}}$ has been computed, the linear density field $\delta_{\boldsymbol{k}}$ at redshift z is derived from the potential $\Phi_{\boldsymbol{k}} = \Phi^{\rm G}_{\boldsymbol{k}} + \Phi^{\rm NG}_{\boldsymbol{k}}$ through the transfer function $T(k)$ and the Poisson equation. Finally, this density field is used to displace N-body particles from a regular cartesian grid according to 2LPT at the desired starting redshift of the simulations.

Finally, we also realized a higher-resolution version of this suite, known as PRINGLS-HR, which follows the same setup as PRINGLS but with a number of particles equal to 1024^3, allowing the simulations to have a mass resolution of 8.21×10^{10} h^{-1} $M_\odot$. This suite comprises only one realization for each value of the parameter $f_{\rm NL}$, which is set to 0, 100, and -100. We report in Table 5.1 also the properties of these simulations.

In the next section, we will present the numerical methods employed to build and analyze halo catalogs, extract summary statistics, and carry out Bayesian inference.

5.3 Numerical Tools

5.3.1 ROCKSTAR

Our analysis relies on the usage of DM halo catalogs, which were obtained using the publicly available halo finder `ROCKSTAR`[3]. As a first step, the `ROCKSTAR` code employs a rapid variant of the 3D `FOF` method to identify overdense regions (see Sect. 3.3.1 for details). It then constructs a hierarchy of `FOF` subgroups in phase space by progressively and adaptively reducing the 6D linking length. This process ensures that a configurable fraction of particles is included in each subgroup relative to their immediate parent group. Subsequently, it translates this hierarchy of `FOF` groups into a list detailing which particles belong to each halo. The algorithm proceeds by establishing relationships between host halos and subhalos, utilizing information from previous time steps when available. Finally, it removes particles that are not gravitationally bound from halos and computes various halo properties.

For determining halo masses, `ROCKSTAR` calculates spherical overdensities using various user-specified density thresholds, Δ. These thresholds can include the virial threshold (Bryan & Norman, 1998) or overdensities relative to the critical or background density, $\rho_{\rm crit}(z)$ and $\rho_{\rm b} \equiv \rho_{\rm crit}(z)\Omega_{\rm m}(z)$ respectively. For Quijote and Quijote-png simulations, we have used the publicly available catalogs at $z = 0,\ 0.5$ and 1, in which masses are characterized by $\Delta_{\rm b} = 200$, and $\Delta_{\rm c} = 200$, 500, and 2500, where the subscripts "b" and "c" denote overdensities relative to $\rho_{\rm b}$ and $\rho_{\rm crit}$, respectively. In this regard, Fig. 5.1 shows an example of a halo identified by `ROCKSTAR` along with two spherical overdensities defined by $R_{200\rm b}$ and $R_{200\rm c}$ that are used to define the corresponding masses. These catalogs were realized assuming a parameter $b = 0.27$ for the linking length of `FOF` groups (see Sect. 3.3.1) and considering halos as collections of at least 10 particles. For Pringls and Pringls-hr, we generated catalogs using the same setup as for Quijote, defining halo masses according to $\Delta_{\rm b} = 200$, and $\Delta_{\rm c} = 200$, 300, 500, 800, 1000 and 1200 at the same redshifts considered for Quijote.

In Fig. 5.2, we show four snapshots at $z = 0$, 0.5, 1 and 2 of one realization of the Quijote suite, including also the DM halos identified by `ROCKSTAR` and characterized by a mass $M_{200\rm b}$ greater than $5 \times 10^{13}\ h^{-1}\ \mathrm{M}_{\odot}$. This mass threshold, which we adopt for all halos considered in this work regardless of the mass definition, corresponds to halos composed of at least 76 particles for the low-resolution simulations (Quijote, Quijote-png and Pringls) and 609 for the high-resolution ones (Pringls-hr).

5.3.2 CosmoBolognaLib

All the cosmological computations and data manipulations in this work are conducted within the framework of `CosmoBolognaLib`[4] (`CBL`, Marulli et al., 2016), a large set of *free-software* C++/Python libraries designed for cosmological research. This software is continuously updated, offering a range of tools for analyzing cosmo-

[3] `https://bitbucket.org/gfcstanford/rockstar/src/main/`
[4] `https://gitlab.com/federicomarulli/CosmoBolognaLib`

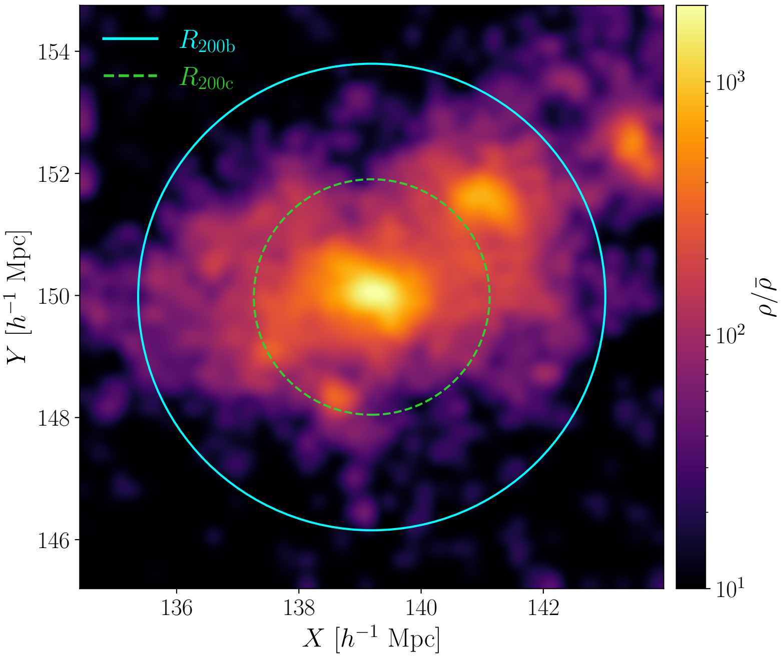

Figure 5.1: An example of a DM halo identified by ROCKSTAR in one of the PRINGLS realizations. The colors represent the magnitude of the density field, expressed as the ratio $\rho/\bar{\rho}$, where $\bar{\rho}$ is the mean matter density of the Universe. Higher-density regions appear in yellow, while lower-density areas are darker. The blue and green circles indicate the radii R_{200b} and R_{200c}, with the former represented by a solid line and the latter by a dashed line. These radii define the spherical overdensities used to compute the corresponding halo masses, M_{200b} and M_{200c}, according to Eq. 3.34.

logical data, including functions for managing catalogs, customizing cosmological models, and performing statistical analysis. Additionally, it provides capabilities for calculating correlation functions, and power spectra, with built-in wrappers for `CAMB`[5] (Lewis et al., 2000) and `CLASS`[6] (Lesgourgues, 2011), or accurate fitting models (see e.g. Eisenstein & Hu, 1999).

The core of this software is represented by the `cbl::cosmology::Cosmology` class, which allows users to set various cosmological parameters, such as the density parameters for different cosmological components ($\Omega_{\rm m}$, $\Omega_{\rm b}$, $\Omega_{\rm r}$, Ω_Λ), the spectral index $n_{\rm s}$, the scalar amplitude $A_{\rm s}$, and the non-Gaussian parameter $f_{\rm NL}$. It also includes pre-configured cosmologies such as *Planck* 2015 (Planck Collaboration et al., 2016) and *Planck* 2018 (Planck Collaboration et al., 2020a). This class supports the calculation of DM halo properties, including their mass function and their linear effective bias. As our goal is to analyze the effects of PNG on numerical simulations, we optimized existing functions for non-Gaussian cosmologies to enhance computational efficiency and incorporate all the theoretical prescriptions discussed in the

[5]`CAMB` (Code for Anisotropies in the Microwave Background) is a cosmology code for calculating CMB anisotropies, lensing, galaxy clustering, dark-age 21 cm power spectra, matter power spectra, and transfer functions. For further details, see `https://camb.readthedocs.io/en/latest/`.

[6]`CLASS` (Cosmic Linear Anisotropy Solving System) is a code used to simulate the evolution of linear perturbations in the Universe and to compute CMB and LSS observables. For details, see `http://class-code.net/class.html`.

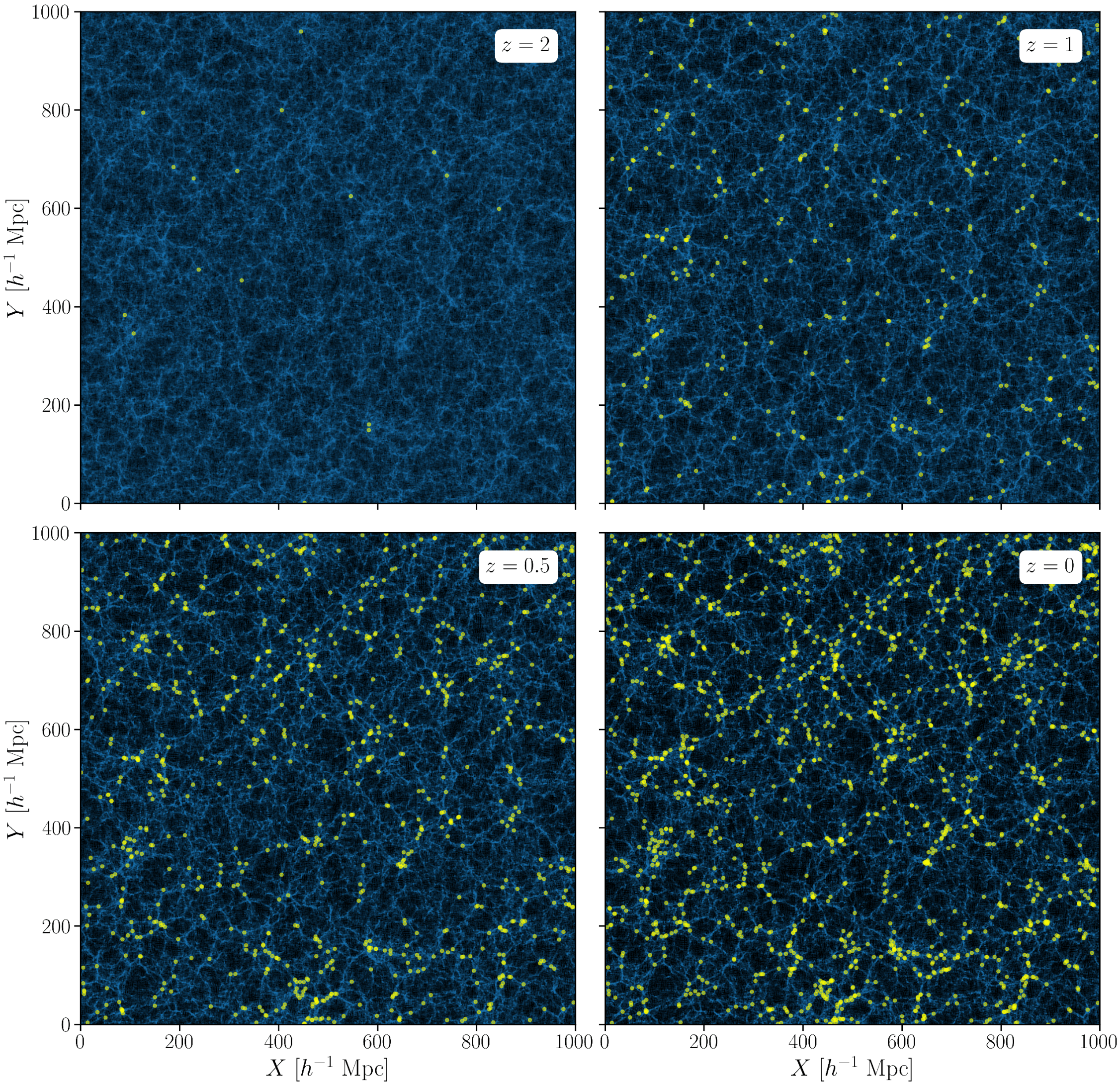

Figure 5.2: Slices with a depth of 20 h^{-1} Mpc of one realization of the Quijote simulations. The panels show the evolution of 512^3 CDM particles from $z = 2$ (*top left panel*) to $z = 0$ (*bottom right panel*). Each particle is represented by a blue dot, while the yellow circles indicate `ROCKSTAR` halos with masses $M_{200\mathrm{b}} > 5 \times 10^{13}\ h^{-1}\ \mathrm{M}_{\odot}$. Finally, we can appreciate how, as time passes, the number of collapsed structures increases and the cosmic web becomes more pronounced, with filaments and nodes growing denser and more interconnected, shaping the structure of the Universe.

previous chapters, such as those for the mass function correction (see Sect. 4.3).

The `cbl::catalogue::Catalogue` class is designed to manage collections of various astrophysical objects, including halos, galaxies, galaxy clusters, and voids. Its characterized by private attributes to store object properties such as positions (in both comoving and observed coordinates), masses, velocities, magnitudes, and other relevant attributes. Additionally, this class allows for the creation of custom catalogs, random catalogs, and sub-catalogs, offering the flexibility to apply user-defined filters to selectively include or exclude objects.

Finally, functions and classes within the `cbl::statistics` namespace were used to perform Bayesian analysis using a Markov Chain Monte Carlo (MCMC) algorithm (Hogg & Foreman-Mackey, 2018; Speagle, 2019), a widely adopted method in cosmology. Bayesian inference provides a framework for estimating the parameters $\boldsymbol{\theta}$ of a model $\mathcal{M}$ given a data set $\mathcal{D}$, allowing us to determine the full probability distribution of the parameters, known as the posterior distribution $P(\boldsymbol{\theta}|\mathcal{D})$. According to Bayes' theorem, the posterior is given by

$$P(\boldsymbol{\theta}|\mathcal{D}) = \frac{\mathcal{L}(\mathcal{D}|\boldsymbol{\theta})P(\boldsymbol{\theta})}{P(\mathcal{D})}\,, \tag{5.6}$$

where $\mathcal{L}(\mathcal{D}|\boldsymbol{\theta})$ is the likelihood function, representing the probability of obtaining the data given a set of parameters, $P(\boldsymbol{\theta})$ is the prior distribution, reflecting prior knowledge about the parameters, and $P(\mathcal{D})$ is the evidence, a normalization constant. Since the posterior is often analytically intractable, numerical methods like MCMC are widely used. This algorithm generates a sequence of parameter samples that follow a Markovian trajectory, meaning that each state in the chain depends only on the previous one. This technique allows for an efficient exploration of the posterior distribution, ensuring accurate parameter estimation.

5.3.3 Pylians

We rely on the Python library `Pylians`[7] (Villaescusa-Navarro, 2018) to compute the matter and halo power spectra of the various realizations of our simulations sets.

The matter power spectrum $P_{\rm m}(k)$ is obtained by first interpolating the distribution of DM particles onto a 3D grid using a mass assignment scheme (MAS), such as cloud-in-cell (Cui et al., 2008). This step constructs the density field, which is then transformed into Fourier space via a fast Fourier transform, yielding the matter density contrast $\delta_{\rm m}(\boldsymbol{k})$. The power spectrum is then computed as

$$P_{\rm m}(k) = \frac{1}{V_{\rm box}}\langle|\delta_{\rm m}(\boldsymbol{k})|^2\rangle\,, \tag{5.7}$$

where $V_{\rm box}$ is the simulation volume, and the average is taken over modes with the same wavenumber k. Since the interpolation onto the grid modifies the Fourier-space representation of the density field, `Pylians` applies a correction by deconvolving the power spectrum with the square of the MAS window function $W_{\rm MAS}(k)$, ensuring an accurate estimate of $P_{\rm m}(k)$. The same procedure is applied to compute the halo power spectrum $P_{\rm h}(k)$, replacing DM particles with halo positions to construct the halo density contrast $\delta_{\rm h}(\boldsymbol{k})$.

[7] https://pylians3.readthedocs.io/en/master/

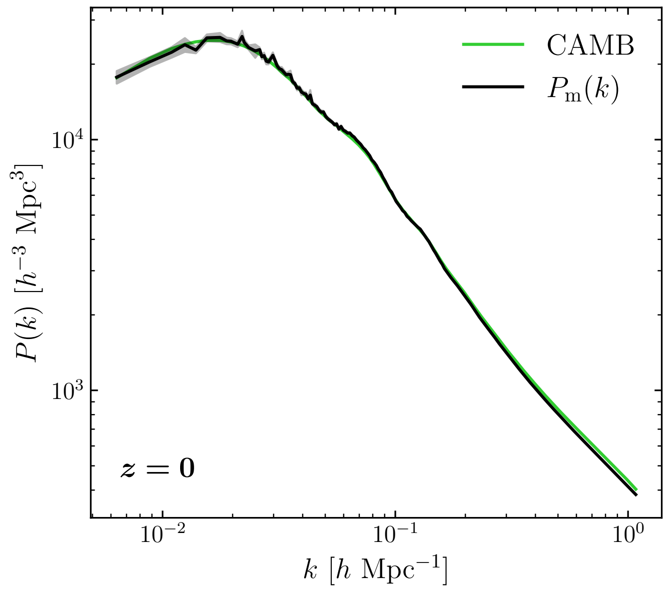

Figure 5.3: The matter power spectrum (black line) at $z = 0$ as a function of the wavenumber k measured by averaging the individual spectra from 100 realizations of the Quijote simulations. The shaded black region indicates the error. In green, we show the nonlinear theoretical prediction computed with `CAMB`, assuming the same cosmological parameters of the simulation set.

For the purposes of our analysis, we modified the power spectrum computation function of `Pylians` to enable a logarithmic binning of the wavenumber k. In Fig. 5.3, we present the matter power spectrum at $z = 0$ evaluated from the Quijote simulations alongside the nonlinear theoretical prediction from `CAMB`. The power spectrum was obtained by averaging the 100 individual power spectra, with an error estimated as the standard deviation across the spectra divided by the square root of the number of realizations. The figure shows good agreement between our measurement and the theoretical model, highlighting how the numerical method employed by our version of `Pylians` is robust. Noticeable deviations emerge only at large k, corresponding to small scales where nonlinear effects become significant. Analogous results are also obtained at higher redshifts.

Chapter 6

Measuring the Effects of local-type PNG on LSS in Simulations

In the previous chapters, we discussed how PNG can alter measurable properties of the LSS. For example, we showed how PNG generates a scale-dependent bias (Sect. 3.5.1) and affects the abundance and mass distribution of DM halos (Sect. 4.3). Currently, the best constraints on the parameter f_{NL} from LSS come from modeling the 3D power spectrum (D'Amico et al., 2022; Cabass et al., 2022; Castorina et al., 2019; Mueller et al., 2022), yielding $\sigma(f_{\mathrm{NL}}) = \mathcal{O}(20)$. However, future LSS observations may improve this to $\sigma(f_{\mathrm{NL}}) < 1$ (Seljak, 2009; Doré et al., 2014; Karagiannis et al., 2018; Ferraro & Wilson, 2019), enabling us to distinguish between inflationary models. These analyses assume that the scale-dependent bias follows a universal relation. However, recent studies have shown this assumption may not always hold (Gutiérrez Adame et al., 2024), suggesting the need for further tests. Regarding the halo mass function, previous works on N-body simulations have shown discrepancies from theoretical models, often attributed to differences in halo mass definitions and halo-finding methods (Grossi et al., 2007; Dalal et al., 2008; Desjacques et al., 2009).

We aim now to investigate on the aforementioned discrepancies with the goal, eventually, to provide solutions to account for them. Hence, in the following sections, we use the data extracted from the simulations presented in Chapter 5 to assess the impact of the local shape of PNG on the bias parameter and mass function, comparing these measurements with the theoretical predictions.

6.1 Scale-dependent Bias in the Quijote-PNG

To compute the scale-dependent bias, the first step is to measure the effective linear bias b_1 in the Gaussian simulations. According to Sect. 3.5, one way to estimate the bias is through the square root of the ratio of the halo power spectrum to the matter power spectrum (Eq. 3.50).

Using our modified version of `Pylians` (see Sect. 5.3.3), we analyzed 100 realizations of the Quijote suites at $z = 0,\ 0.5$, and 1. For each realization, we computed both the matter and halo power spectra, with the halo power spectrum estimated using `ROCKSTAR` halos and mass definitions $M_{200\mathrm{b}}$, $M_{200\mathrm{c}}$, and $M_{500\mathrm{c}}$. Additionally, we limited our analysis to halos with masses above $5 \times 10^{13}\ h^{-1}\ \mathrm{M}_{\odot}$, regardless of the mass definition.

At this point, estimating b_1 ideally involves computing the square root of the ratio $P_\mathrm{h}/P_\mathrm{m}$ for each of the 100 realizations, then taking the mean value at each k bin and fit through an MCMC for the best-fit value of b_1. However, for the halo power spectrum, it is crucial to account for shot noise, which can be assumed to follow a Poissonian distribution. In this case, the shot noise contribution is given by $V_\mathrm{box}/N_\mathrm{halo}$, where N_halo is the number of halos in the simulation box.

To mitigate the impact of shot noise, we estimate the effective bias using the cross power spectrum between halos and DM, as for it the shot noise is expected to be significantly reduces and thus negligible. The computation follows the same steps of the auto power spectrum, with the key difference that we now take the average product of the two density fields, δ_m and δ_h, as described in Eq. 5.7. For each realization, the effective bias is then simply given by

$$b_1(z) = \frac{P_\mathrm{h,m}(k,z)}{P_\mathrm{m}(k,z)}\,, \tag{6.1}$$

where $P_\mathrm{h,m}(k)$ is the cross power spectrum between halos and DM. Then, we average the values of these ratios and for each k bin and we derive the error as the standard deviation between the ratios divided by the square root of the number of realizations. Finally, we determine the best-fit value of b_1 using an MCMC fitting procedure, assuming a Gaussian likelihood and a flat prior on b_1.

Figure 6.1 shows the measured effective bias and the corresponding best-fit b_1 values, in the k range between 6×10^{-3} and 3×10^{-2} h Mpc^{-1}, for different mass definitions and redshifts. Additionally, we compare our fitted values with the theoretical effective bias of DM halos, computed as

$$b_1(z) = \frac{\displaystyle\int_{M_\mathrm{min}}^{M_\mathrm{max}} \mathrm{d}M\, b(M,z)\phi(M,z)}{\displaystyle\int_{M_\mathrm{min}}^{M_\mathrm{max}} \mathrm{d}M\,\phi(M,z)}\,, \tag{6.2}$$

where $\phi(M,z)$ is the halo mass function from Tinker et al. (2008) and the integration limits, M_min and M_max, correspond to the minimum and maximum halo masses in our sample. The linear bias $b(M,z)$ is provided by the model proposed by Tinker et al. (2010), in which b is computed according to the following expression:

$$b(\nu) = 1 - A\frac{\nu^a}{\nu^a + \delta_\mathrm{c}^a} + B\nu^b + C\nu^c\,. \tag{6.3}$$

Here, $\nu \equiv \delta_\mathrm{c}/\sigma_M$ while the six parameters A, a, B, b, C, c are fitted from numerous N-body simulation sets.

In general, we find deviations of up to 4% between our measured effective bias and the predicted values. However, since our primary focus is on relative deviations from the Gaussian case, we decided to rely on the measured values of b_1 as the best estimate of the effective bias for the cosmological simulations considered.

To test the non-Gaussian bias model presented in Sect. 3.5.1, we use the corresponding 100 realizations for each cosmological scenario of the QUIJOTE-PNG simulations (f_NL = 100 and −100). Following the same approach of the Gaussian

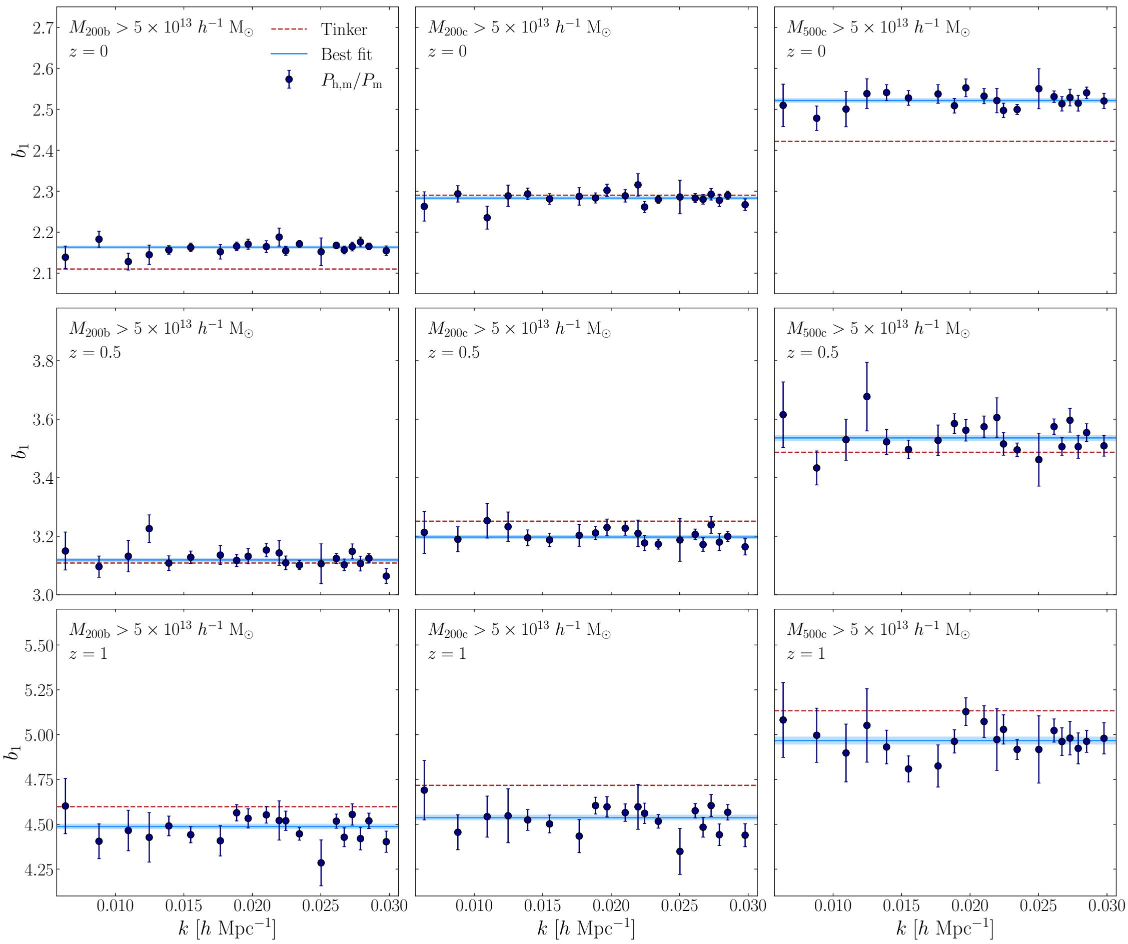

Figure 6.1: Comparison between the measured and predicted values of the effective halo bias b_1 for different z and mass definitions. Rows correspond to data sets with the same redshift (from *top* to *bottom* z = 0, 0.5, and 1), while columns refer to different halo mass definitions (from *left* to *right* M_{200b}, M_{200c}, M_{500c}). All the halos have masses above 5×10^{13} h^{-1} $M_\odot$, regardless of the chosen halo identification threshold. The blue dots represent the measured bias at any k bin, while the best fit value for b_1 and its error are indicated by the light blue line and shaded region, respectively. Finally, the red dashed line refers to the prediction (Eq. 6.2) proposed by Tinker et al. (2010).

simulations, we restrict the analysis to halos with masses $M_{200\mathrm{b}}$, $M_{200\mathrm{c}}$ and $M_{500\mathrm{c}}$ greater than 5×10^{13} h^{-1} $\mathrm{M}_\odot$. For each realization, we measure the halo-matter power spectrum $P_{\mathrm{h,m}}(k, z)$ and the matter power spectrum $P_{\mathrm{m}}(k, z)$ in the same k interval of the Gaussian case, from 6×10^{-3} to 3×10^{-2} h Mpc^{-1}. This interval was chosen due to the dependence k^{-2} on the non-Gaussian bias, as we expect the signal induced by PNG to be stronger at very large scale, in the low k regime (see Fig. 3.3).

At this point, we average the scale-dependent bias measured in the 100 realizations for each scenario and assign an error to each k bin, calculated as the standard deviation divided by the square root of the number of realizations employed. Finally, we compute the correction Δb to the effective bias by subtracting the mean bias measured in the PNG simulations by the one measured in the Gaussian ΛCDM simulations (Fig. 6.1), propagating the error. The results of this analysis are presented in Fig. 6.2, which shows a direct comparison of the measured Δb with the predictions from Eq. 3.52, assuming the universal relation (i.e. $p = 1$). In this context, the theoretical behavior of the correction Δb is given by

$$\Delta b(k, z) = \frac{3 f_{\mathrm{NL}} b_\Phi(z)}{5 \mathcal{M}(k, z)} = \frac{3 f_{\mathrm{NL}}}{5 \mathcal{M}(k, z)} 2 \delta_{\mathrm{c},0}(b_1(z) - p) \,. \tag{6.4}$$

The plot displays the measured deviations at different redshifts and for various mass definitions. We observe that the overall trend of the data is well reproduced by the non-Gaussian bias model, despite the fact that small discrepancies are observed for $M_{200\mathrm{b}}$ halos at $z = 0.5$ and $z = 1$. We tested the effect of leaving p as a free parameter of the model to reveal possible deviations from universality. However, the additional degree of freedom appears to not significantly improve the agreement of the model with the observed trend, except in a few cases. Moreover, the best-fit values obtained for p range from 0.64 to 1.47 without a clear pattern. Thus, we conclude that no strong evidence is found for deviations from universality, at least with the precision achieved with the data employed in our analysis.

6.2 Deviations from Gaussianity in the Halo Mass Function

We begin this section by examining the Gaussian mass function derived from the standard ΛCDM simulations. Figure 6.3 presents the Gaussian halo mass function as measured in the Quijote simulations, focusing on ROCKSTAR halos with masses defined by $M_{200\mathrm{b}}$, $M_{200\mathrm{c}}$, and $M_{500\mathrm{c}}$ at redshift $z = 0$. For each definition, the mass function is calculated as the average of the 8000 individual mass functions obtained from each realization. The error for each mass bin is estimated as $\sigma/\sqrt{N_{\mathrm{real}}}$, where σ represents the standard deviation between all realizations, and N_{real} is the total number of realizations. Given the large number of realizations in Quijote, the statistical error on the mean is significantly reduced compared to that resulting from the analysis of an individual halo mass function.

We also compare the estimated mass functions with the phenomenological model

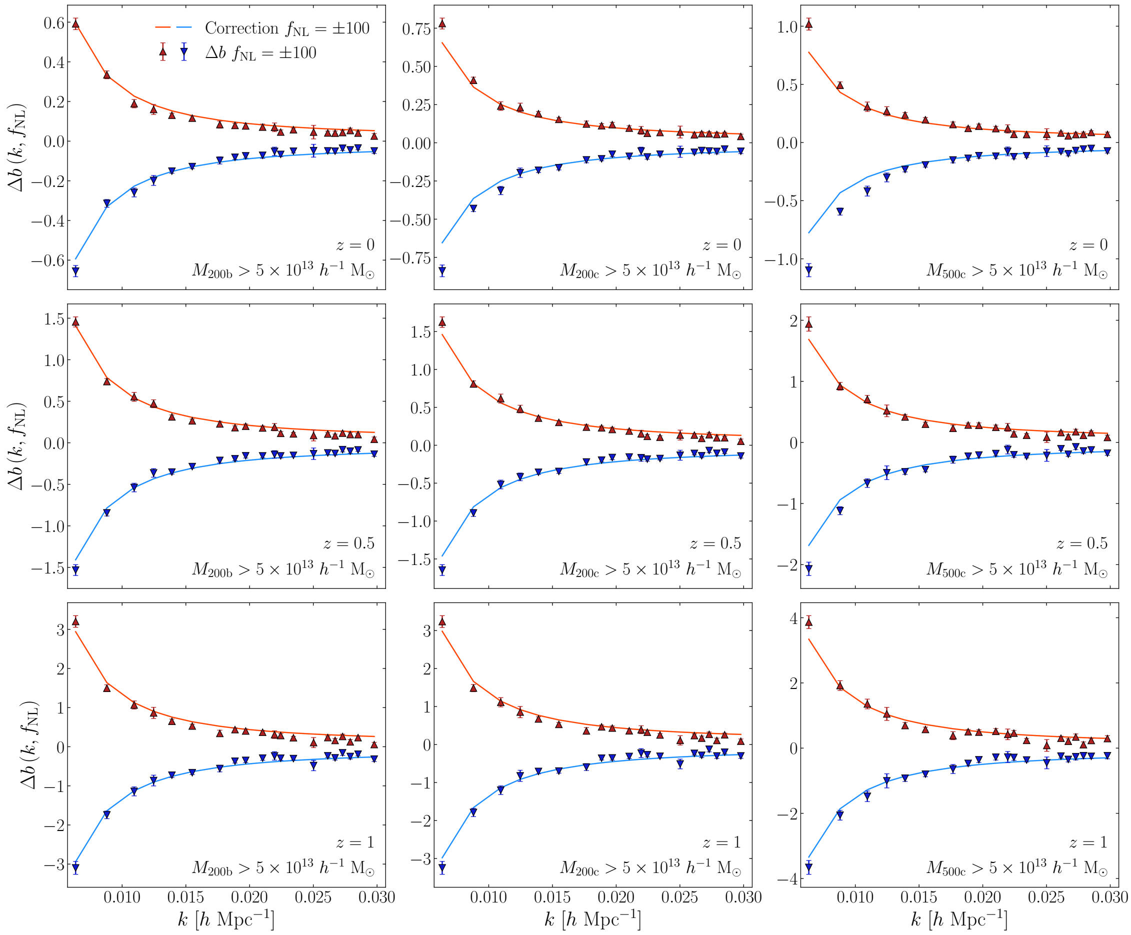

Figure 6.2: Deviations of the halo bias from b_1 caused by the presence of PNG. The quantity Δb is the difference between the non-Gaussian and the Gaussian bias and is reported as a function of scale k, for different z and mass definitions. Rows correspond to data sets with the same redshift (from top to bottom $z = 0$, 0.5, and 1), while columns refer to different halo mass definitions (from left to right $M_{200\mathrm{b}}$, $M_{200\mathrm{c}}$, $M_{500\mathrm{c}}$). All the halos have masses above $5 \times 10^{13}\ h^{-1}$ M$_\odot$, regardless of the chosen halo identification threshold. The red upward-pointing triangles and the blue downward-pointing triangles represent the data extracted from simulations with $f_{\mathrm{NL}} = 100$ and $f_{\mathrm{NL}} = -100$, respectively. The orange and light blue solid lines are instead the predictions computed for the non-Gaussian halo bias assuming $p = 1$ in Eq. 3.54, again for $f_{\mathrm{NL}} = 100$ and $f_{\mathrm{NL}} = -100$ respectively.

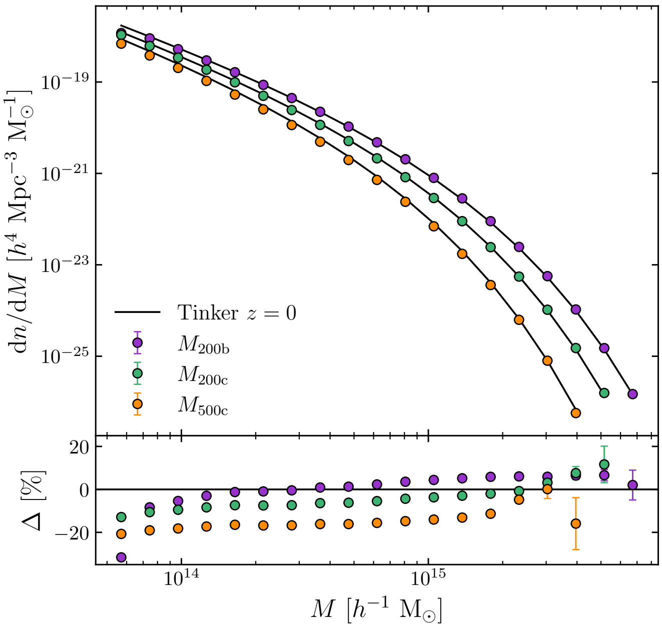

Figure 6.3: The average Gaussian halo mass function measured from 8000 realizations of the QUIJOTE simulations at redshift $z = 0$, within the mass range of 5×10^{13} to $7 \times 10^{15}\ h^{-1}\ \mathrm{M}_\odot$. The purple, green and orange circles correspond to different mass definitions: $M_{200\mathrm{b}}$, $M_{200\mathrm{c}}$, and $M_{500\mathrm{c}}$, respectively. The solid lines indicate the model proposed by Tinker et al. (2008) for each definition of mass. The subpanel displays the percentage residuals of the data relative to the model.

proposed by Tinker et al. (2008), which can be expressed as (Table 4.1):

$$\frac{\mathrm{d}n}{\mathrm{d}M} = f\left(\sigma_M\right) \frac{\bar{\rho}_\mathrm{m}}{M} \frac{\mathrm{d} \ln \sigma_M^{-1}}{\mathrm{d}M}\,, \quad \text{with} \quad f\left(\sigma_M\right) = A \left[\left(\frac{\sigma_M}{b}\right)^{-a} + 1\right] \mathrm{e}^{-c/\sigma_M^2}\,. \tag{6.5}$$

The four parameters characterizing the multiplicity function are determined by fitting the data obtained from various simulation sets. It is important to note that the model is calibrated for halo masses defined by overdensities relative to the background density. To align our halo definitions with this model, we convert the Δ_c values characterizing our halos into Δ_b by imposing

$$\Delta_\mathrm{c} \rho_\mathrm{crit} = \Delta_\mathrm{b} \rho_\mathrm{b} \quad \Longrightarrow \quad \Delta_\mathrm{b} = \frac{\Delta_\mathrm{c}}{\Omega_\mathrm{m}(z)}\,. \tag{6.6}$$

Here, the matter density parameter evolves with redshift according to the relation

$$\Omega_\mathrm{m}(z) = \frac{\Omega_{0,\mathrm{m}}(1+z)^3}{E^2(z)}\,, \tag{6.7}$$

which is obtained by combining Eqs. 1.32 and 1.37. Therefore, each value Δ_c (e.g., 200, 500, 2500) corresponds to a different value Δ_b at different z. We report in Table 6.1, the converted values of our Δ_c at the three redshifts of our snapshots.

As illustrated in the subplot of Fig. 6.3, the phenomenological model does not reproduce the data with high accuracy. For halos defined by $M_{200\mathrm{b}}$, the agreement

Δ_c	Δ_b		
	$z = 0$	$z = 0.5$	$z = 1$
200	630	327	254
300	945	491	381
500	1575	818	634
800	2520	1310	1015
1000	3150	1637	1269
1200	3780	1964	1522
2500	7874	4092	3172

Table 6.1: Conversion of the overdensity threshold Δ_c into Δ_b for different redshifts. The values of Δ_b are computed according to Eqs. 6.6 and 6.7. Each column shows the corresponding Δ_b at $z = 0,\ 0.5$ and 1.

is relatively good, with residuals around 5% (excluding the first two data points). However, as we move to other mass definitions, the agreement deteriorates, with residuals increasing up to 10% and 20% (in absolute value) for M_{200c} and M_{500c}, respectively. Similar results are also observed when the mass functions are examined at higher redshifts. We repeated the analysis on the halo mass functions extracted from PRINGLS and PRINGLS-HR and verified their consistency with the QUIJOTE data in all definitions of mass and redshifts. In this case as well, the model again showed poor agreement with the PRINGLS data, despite the larger error bars due to the smaller number of available realizations (only 10 per cosmological model). For PRINGLS-HR, we only have one simulation for the ΛCDM scenario, therefore we computed the associated uncertainty as the standard Poissonian error, equal to the square root of the counts in each bin. For this simulation we found good agreement for the M_{200b} case, with residuals below 5%. However, the situation remains unchanged for the other mass definitions.

Because of the statistically relevant discrepancies found between the model and the Gaussian data, we chose to use the measured Gaussian mass function as a reference to calculate the deviation of the non-Gaussian halo mass function. In other words, we substituted the halo number counts predicted by the Tinker et al. (2008) model with the data extracted from the QUIJOTE simulations, which have an associated statistical uncertainty that is almost negligible. Although this approach cannot be applied to real data, it eliminates systematic errors related to poor agreement between the data and the mass function model and allows us to directly measure the expected deviations from Gaussianity.

In this context, we focus now on the modifications induced by PNG on the halo mass function, comparing the predictions of the theoretical models introduced in Sect. 4.3 and data extracted from the simulations presented in Chapter 5. As we have seen in Sect. 4.3, the corrections to the halo mass function model require the computation of the normalized skewness and kurtosis of the distribution (S_3 and S_4, respectively) at each value M of the mass. However, the numerical integration required for these computations is time consuming. To accelerate the process, we

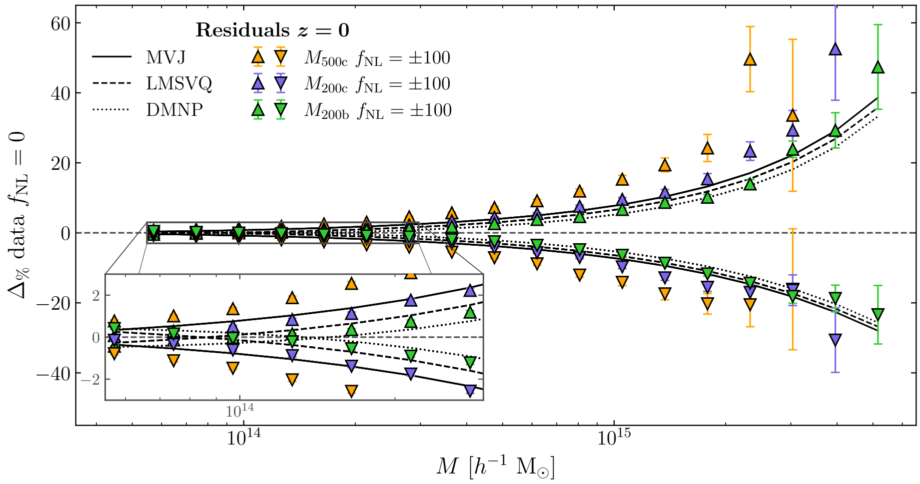

Figure 6.4: Deviations between halo mass functions in PNG scenarios with respect to their Gaussian counterpart. We compare different theoretical models and data sets at $z = 0$, for the two scenarios $f_{\rm NL} = 100$ and -100. On the y-axis we show the percentage residuals (Eq. 6.9) between the mean halo number counts measured using Quijote-png and those extracted from Quijote (representing the reference ΛCDM number counts). Different marker shapes are used for the two $f_{\rm NL}$ values: upward-pointing triangles for $+100$ and downward-pointing triangles for -100. Colors denote different mass definitions: green for $M_{\rm 200b}$, blue for $M_{\rm 200c}$, and orange for $M_{\rm 500c}$. The theoretical models for mass function corrections are indicated with different line styles: solid for MVJ, dashed for LMSVQ, and dotted for DMNP (Eqs. 4.17 to 4.19, respectively). The subplot displays a zoomed region at low masses to allow a clearer understanding of the behavior of both model and data points.

employ empirical relations for S_3 and S_4 of the form (Enqvist et al., 2011; Yokoyama et al., 2011):

$$S_3 \simeq f_{\rm NL}\frac{\alpha}{\sigma_M^{2\beta}}\,, \qquad S_4 \simeq f_{\rm NL}^2\frac{\gamma}{\sigma_M^{2\theta}}\,, \tag{6.8}$$

with $\alpha = 2.16 \times 10^{-4}$, $\beta = 0.4$, $\gamma = 8.43 \times 10^{-8}$ and $\theta = 0.99$. It is important to emphasize that these empirical formulae were calibrated specifically for the cosmology used in our simulations. Consequently, if different cosmological parameters are adopted, the empirical functions will need to be re-calibrated accordingly. We will discuss this limitation in the final chapter of this Thesis.

In Fig. 6.4 we show the residuals at $z = 0$ between the mean halo mass functions measured in Quijote-png (one for $f_{\rm NL} = 100$ and one for $f_{\rm NL} = -100$) and the reference Gaussian halo mass function derived from Quijote. The residuals are expressed as percentages through the following relation:

$$\Delta_{\%}\left(M_i\right) = 100\left(\frac{\mathrm{d}\bar{n}/\mathrm{d}M(M_i, f_{\rm NL} \neq 0)}{\mathrm{d}\bar{n}/\mathrm{d}M(M_i, f_{\rm NL} = 0)} - 1\right)\,, \tag{6.9}$$

where $\mathrm{d}\bar{n}/\mathrm{d}M$ represents the mean halo mass function and M_i indicates its various mass bins. We report in the same figure also the deviations predicted by the MVJ, LMSVQ, and DMNP models (Eqs. 4.17 to 4.19, respectively).

It is easy to note how these theoretical models fail to accurately reproduce the data, with the agreement worsening as the overdensity used to define halo masses increases. LMSVQ and DMNP models provide a reasonable fit for M_{200b} halos, while MVJ aligns more closely with the M_{200c} case. However, all models exhibit slight over- or underpredictions in the amplitude of the deviation from the Gaussian case (the standard ΛCDM scenario) and the behavior of halos characterized by other mass definitions, such as M_{500c} and M_{2500c} — the latter not displayed for clarity reasons — is not well reproduced by any of the theoretical models. We also found analogous trends at $z = 0.5$ and 1.

Testing the models against the data extracted from Pringls and Pringls-hr simulations with the same f_{NL} values led to very similar results. In fact, despite larger error bars, the observed deviations for the same mass definitions are fully consistent with those shown in Fig. 6.4, as expected.

Previous studies (Grossi et al., 2009; Pillepich et al., 2010; Wagner & Verde, 2012) found that similar discrepancies also arise for halos identified by a `FOF` algorithm. In particular, both the LMSVQ and the MVJ corrections have been evidenced to overpredict the effect of PNG in this scenario. However, these differences could be mitigated by introducing an external factor q into the calculations for the corrections of the halo mass function. This factor modifies the present linear collapse threshold as $\delta_{c,0} = 1.686\sqrt{q}$, with fits to numerical simulations indicating a value $q \simeq 0.75-0.8$. This adjustment is usually attributed to the effects of ellipsoidal collapse, which are reflected in a deviation from sphericity in the case of `FOF` halos. However, this modification does not improve the description of halos identified using a spherical overdensity algorithm, since in this case halos are spherical by construction.

Consequently, in Sect. 7.1, we will introduce a new parameterization that incorporates the dependency on the halo mass definition into the theoretical framework, without interpreting this correction as due to the shape of the overdensities. The following section will focus on exploring the physical motivation behind the necessity of incorporating a dependency on the density contrast threshold into the theoretical model of the mass function in PNG scenarios.

6.3 Density Profiles in PNG Cosmologies

To have hints about the behavior observed in Fig. 6.4, we compute the stacked halo density profiles using Pringls simulations. In this case, we do not use the Quijote-png data sets because the total number of halos in Pringls is sufficient to provide good enough statistics and we are interested in exploring multiple values of f_{NL}. A similar density profile analysis was performed by Smith et al. (2011) using `FOF` halos in simulations with $f_{NL} = \pm 100$. In contrast, we decided to focus on halos with masses M_{200b} identified by `ROCKSTAR`, extending the analysis also to smaller values of $|f_{NL}|$.

6.3.1 Evaluating the density profiles

Density profiles are derived by counting DM particles within concentric and logarithmically spaced spheres of radii R_i, which range from $0.05R_{200b}$ to $3R_{200b}$. Here,

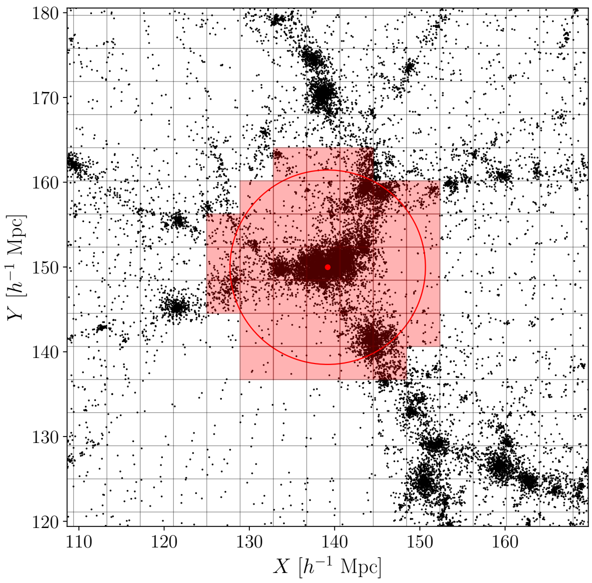

Figure 6.5: Illustration of the `ChainMesh3D` algorithm applied to a slice of the simulation box. The DM particles are displayed in black, while the grid used for spatial partitioning is overlaid in thin black lines. The red circle represents the search region centered on a `ROCKSTAR` halo (marked by the red dot), while the red-shaded cells highlights the grid elements that intersect with this search region.

R_{200b} is defined as the radius where the density is 200 times the background density and is provided directly by the halo finder. The profile estimate is then represented in terms of the dimensionless density contrast:

$$\delta(R_i) = \frac{1}{\bar{n}_{\rm DM}} \frac{N(R_i)}{V(R_i)} - 1 \,, \tag{6.10}$$

where $\bar{n}_{\rm DM}$, $N(R_i)$, and $V(R_i)$ denote the mean number density of DM particles in the simulation box, the number of particles within the i-th sphere and its volume, respectively. This method was applied to all halos within the mass range of 10^{14} to 4×10^{15} h^{-1} M$_\odot$, using snapshots at $z = 0$ for all available realizations.

To efficiently identify DM particles within a given searching area, we utilize the `ChainMesh3D` algorithm from `CBL`, a spatial partitioning technique specifically designed for fast neighbor searches in cosmological simulations. This method organizes particles into a cubic grid with cells of side length $\ell_{\rm cell}$, reducing the computational cost by limiting the search to cells that intersect the target region.

In our implementation, we set the cell size to $\ell_{\rm cell} = 2\,\bar{d}$, where $\bar{d}$ denotes the mean inter-particle separation, evaluated as

$$\bar{d} = \left(\frac{V_{\rm box}}{N_{\rm DM}}\right)^{1/3} = \left(\frac{1}{\bar{n}_{\rm DM}}\right)^{1/3} \simeq 1.95\ h^{-1}\ {\rm Mpc}\,, \tag{6.11}$$

where $N_{\rm DM}$ and $V_{\rm box}$ is the total number of DM particles in the simulated box and the volume of the latter, respectively. For each halo, particles are selected from grid cells overlapping with a spherical search region of radius $R = 3R_{200b}$ and centered on the halo center position provided by the catalog. Figure 6.5 illustrates a 2D

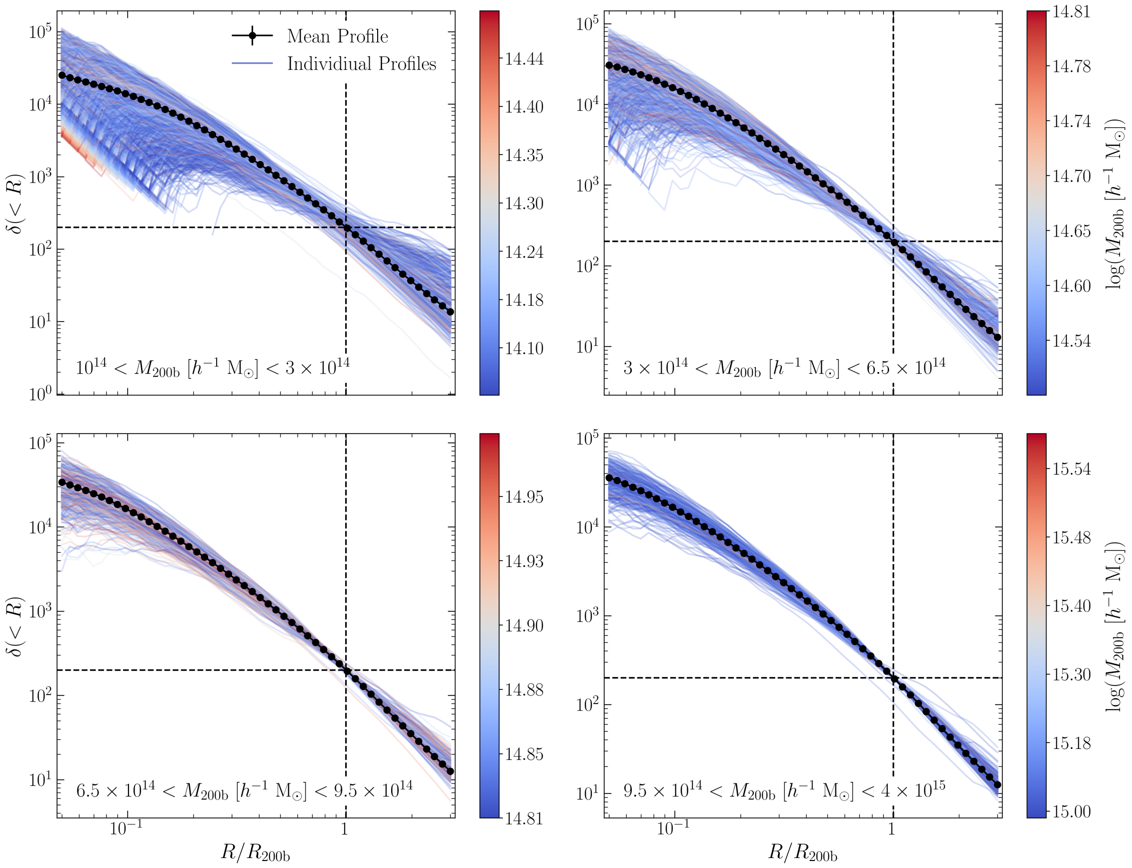

Figure 6.6: Stacked DM halo density profiles at $z = 0$ from one realization of the PRINGLS ΛCDM. The individual profiles are colored according to their mass: the higher is the mass the redder is the color. The mean profile in each radial bin is indicated with black dots while the vertical and horizontal black dashed lines represent the radius R_{200b} and the corresponding value of the density contrast, respectively. The panels refer to the different M_{200b} bins, covering in total the range between 10^{14} and $4 \times 10^{15}\ h^{-1}\ M_\odot$.

projection of the grid created by the `ChainMesh3D` algorithm, and the region around a halo from which particles are recovered.

To avoid stacking density profiles of halos having significantly different masses, we divide the individual profiles into four bins according to the mass of the halos. Finally, for each mass bin we average the individual profiles of realizations sharing the same value of $f_{\rm NL}$ $(0, \pm 40, \pm 100)$, assigning an error to each radial bin equal to $\sigma/\sqrt{N_{\rm halos}}$, where σ is the standard deviation between individual profiles and $N_{\rm halos}$ is the total number of halos in each mass bin.

Figure 6.6 shows an example of the mean density profiles obtained by averaging the individual profiles measured in one realization of the PRINGLS ΛCDM simulations. Notably, due to the mass definition used, every halo is characterized by the same density contrast at a distance $R = R_{200b}$ from the center, therefore we expect all individual profiles, as well as the mean ones, to converge to that value as R approaches R_{200b}. However, as evidenced by some profiles, particularly in the *top right* panel, certain profiles violate this condition. This is likely due to the sophisticated approach employed by the `ROCKSTAR` halo finder to compute halo properties (see

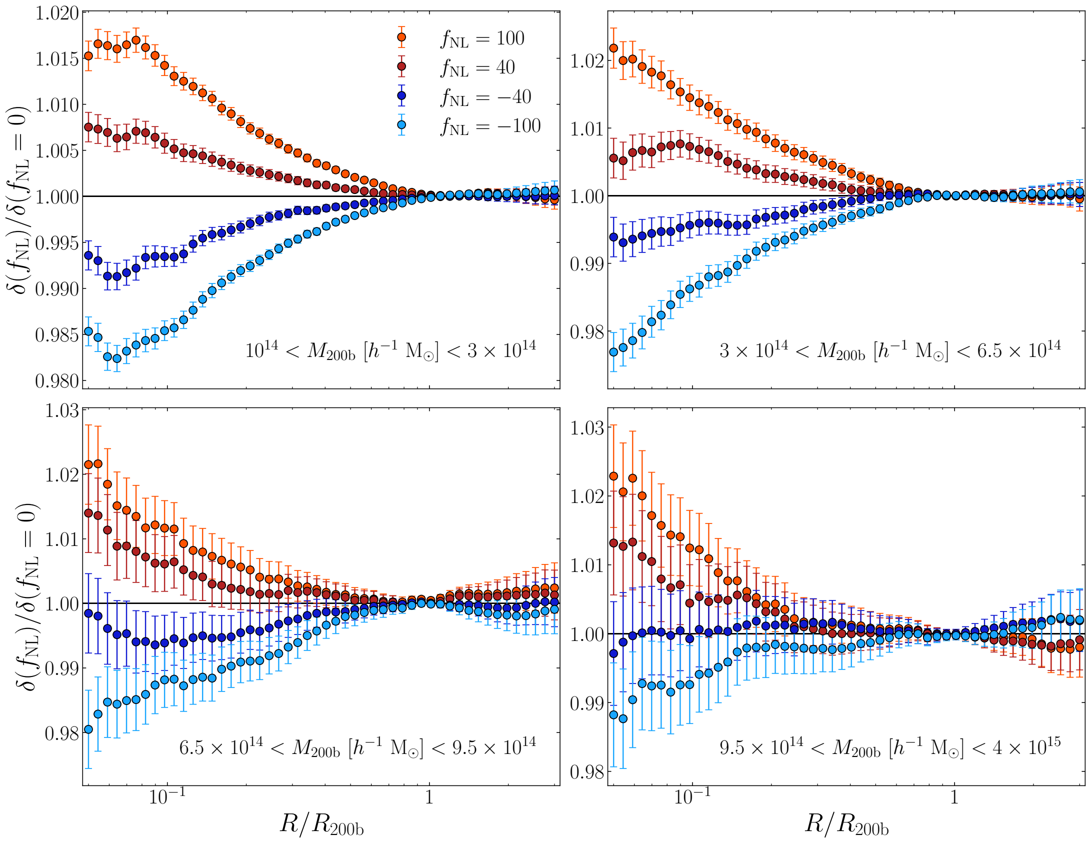

Figure 6.7: Ratio at $z = 0$ of stacked halo density profiles from PRINGLS with different levels of PNG to those measured in the Gaussian ΛCDM scenario. In light blue, blue, red, and orange we show the results for $f_{\mathrm{NL}} = -100,\ -40,\ 40$, and 100, respectively. The panels correspond to different M_{200b} bins, covering in total the range between 10^{14} and $4 \times 10^{15}\ h^{-1}\ \mathrm{M}_\odot$.

Sect. 3.3.1), which may introduce small deviations compared to simpler methods, such as the direct particle count that we employed for this analysis.

Finally, the mean profile in the first mass bin appears to be flatter in the innermost region compared to the others. This is because, due to the limited resolution of the simulations, lower-mass halos tend to have poorly resolved inner regions, leading to artificially low central densities. This effect is evident when inspecting the individual profiles, which reveal that many halos exhibit very low inner densities. We will investigate the effects of unresolved density profiles on our re-parameterization of the model for the mass function correction in Sect. 7.1.3.

6.3.2 The impact of PNG

To assess deviations from Gaussianity, we normalized the mean profiles measured in simulations with $f_{\mathrm{NL}} \neq 0$ by profile obtained from the Gaussian ΛCDM simulations, and then propagated the error of the latter on the ratio. The results of this analysis are shown in Fig. 6.7.

DM halos exhibit an increase in inner density for positive values of f_{NL}, while a decrease is observed when f_{NL} is negative. We note that all profiles become

M_{200b} [10^{14} h^{-1} M$_\odot$]	Number of halos				
	$f_{\rm NL} = 100$	$f_{\rm NL} = 40$	$f_{\rm NL} = 0$	$f_{\rm NL} = -40$	$f_{\rm NL} = -100$
$1 - 3$	285 787	285 216	284 678	284 332	283 663
$3 - 6.5$	48 059	47 471	47 088	46 580	45 983
$6.5 - 9.5$	7085	6913	6775	6666	6488
$9.5 - 40$	3795	3602	3492	3375	3206

Table 6.2: Total number of DM halos across each set of 10 PRINGLS simulations characterized by a specific $f_{\rm NL}$ value at $z = 0$: from left to right 100, 40, 0, -40 and -100. Each row corresponds to a specific mass range, with the mass defined by the overdensity $\Delta_{\rm b} = 200$.

almost identical when the x-axis reaches unity, as these halos have the same mass definition (M_{200b}) and the distance from their center is rescaled by their radius (R_{200b}). Furthermore, the measured mean deviations from the Gaussian profile appear symmetric, with variations of at most 2% in the innermost regions of the halos. Within each mass bin, the number of halos increases as $f_{\rm NL}$ increases, with the opposite behavior occurring as $f_{\rm NL}$ decreases, as shown in Table 6.2.

The change we observe in the slope of the density profiles, i.e. in the compactness of DM halos, is reflected in the deviations found in the halo mass function for different values of the identification threshold (see Fig. 6.4). In fact, given a threshold Δ, the radius at which that density contrast is reached depends on the value of $f_{\rm NL}$, and the mass of the halo changes as a consequence. For example, for positive values of $f_{\rm NL}$, the threshold is met at a greater distance from the halo center, resulting in a larger mass within the corresponding sphere. Additionally, higher Δ values cause the identified halo radii to be closer to the center, leading to greater deviations from the Gaussian case.

We believe that the effects induced by PNG on the halo density profile were not included in the original development of the theoretical framework for the halo mass function, as very complex to predict without relying on numerical simulations. Thus, our goal is now to incorporate these effects into the theoretical modeling by adding a semi-analytical parameterization dependent on the halo mass definition.

Chapter 7

A Revisited Correction to the Mass Function for PNG

Given the results presented in Chapter 6, it is essential to develop a method that explicitly accounts for the dependence on mass definitions in the PNG correction factor for the mass function. In the following sections, we will take one of the previously discussed models as a starting point and introduce a simple modification to incorporate this dependence. We will then test the proposed approach, examining its potential systematic errors and assessing possible resolution effects. Finally, we will evaluate the improvements offered by our method, particularly in the perspective of constraining f_{NL} using cluster number counts.

7.1 A New Parameterization for the Mass Function Correction

7.1.1 Calibration of the correction factor κ

Based on the results presented in the Chapter 6, it is clear that the theoretical models, without adjustments for specific mass definitions, do not adequately agree with data from simulations, especially at higher overdensities. To address this, we sought a method that incorporates the effects of varying mass definitions. In the literature, it is common to introduce a correction factor to align theoretical models with `FOF` data sets when correcting the halo mass function for PNG (as discussed in Sect. 6.2). We adopted a similar approach by introducing a modification of the linear critical density for collapse, parameterized by a factor κ, which varies with the chosen overdensity criterion used to identify the halos. Among all the theoretical non-Gaussian mass function models analyzed, we chose to rely on the LMSVQ model since it showed the best agreement with our data (see e.g. the $M_{200\mathrm{b}}$ case in Fig. 6.4). Consequently, we modify Eq. 4.17 by substituting δ_{c} with $\kappa \times \delta_{\mathrm{c}}$.

We start by focusing on Quijote-png data sets, as the high number of realizations available for these simulations is expected to provide a more accurate and precise calibration of this correction factor. In order to extend the LMSVQ model to halos with any mass definitions, we perform a Bayesian analysis using a MCMC

Mass	$\kappa(z=0)$		$\kappa(z=0.5)$		$\kappa(z=1)$	
	$f_{\rm NL}=100$	$f_{\rm NL}=-100$	$f_{\rm NL}=100$	$f_{\rm NL}=-100$	$f_{\rm NL}=100$	$f_{\rm NL}=-100$
$M_{200\rm b}$	0.971 ± 0.003	0.976 ± 0.003	0.978 ± 0.002	0.980 ± 0.002	0.982 ± 0.002	0.979 ± 0.002
$M_{200\rm c}$	1.091 ± 0.003	1.097 ± 0.003	1.021 ± 0.002	1.022 ± 0.002	0.998 ± 0.002	0.995 ± 0.002
$M_{500\rm c}$	1.243 ± 0.003	1.245 ± 0.003	1.139 ± 0.002	1.136 ± 0.003	1.089 ± 0.002	1.081 ± 0.002
$M_{2500\rm c}$	1.670 ± 0.004	1.665 ± 0.004	1.509 ± 0.004	1.508 ± 0.004	1.415 ± 0.006	1.396 ± 0.007

Table 7.1: The correction factor κ as a function of redshift and mass definition for the two $f_{\rm NL}$ values from the QUIJOTE-PNG simulations. Parameter values and their associated errors are estimated through MCMC fitting.

algorithm (see Sect. 5.3.2) to estimate the factor required to correct the model for data sets created with different overdensity thresholds.

Since we are interested in modeling the deviations from the Gaussian mass function, our data are given by the residuals between the halo number counts extracted from the simulations with $f_{\rm NL} \neq 0$ and those with $f_{\rm NL} = 0$, as defined in Eq. 6.9. For this fit we set a flat prior on the only free parameter of the model (i.e. κ) and a Gaussian likelihood function expressed as:

$$\mathcal{L}\left(\{x_i\}|\{\sigma_i\},\{\mu_i\}\right) = \prod_{i=1}^{N_{\rm d}} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(x_i-\mu_i)^2}{2\sigma_i^2}\right) , \tag{7.1}$$

where $N_{\rm d}$ is the total number of data points, and x_i, σ_i, and μ_i are the data, the error, and the model at the mass bin i, respectively. We repeat the procedure to cover all the available redshifts and for both cosmological scenarios $f_{\rm NL} = \pm 100$. Table 7.1 shows the resulting values of κ derived for different redshifts and mass definitions. For a fixed mass definition, we found no significant differences between the fits of the $f_{\rm NL} = 100$ and -100 data sets, as almost all estimated parameters are consistent within the uncertainties. This suggests that κ is independent of the sign of $f_{\rm NL}$ (see Sect. 7.1.2 for further details). In contrast, the values of κ derived for halos defined by a critical overdensity threshold, $\Delta_{\rm c}$, exhibit a slight dependence on redshift, with κ decreasing as redshift increases.

To find a relation that can be used for halos identified at any redshift, we convert all critical overdensities $\Delta_{\rm c}$ (200, 500, and 2500) into background overdensities $\Delta_{\rm b}$ by dividing by $\Omega_{\rm m}(z)$, according to Eq. 6.6. In our case, each value of $\Delta_{\rm c}$ corresponds to three values of $\Delta_{\rm b}$, one for each redshift of 0, 0.5 and 1 (see Table 6.1). The advantage of converting these thresholds in terms of background densities lies also in removing the dependence on h: given the current uncertainty on the estimation of the Hubble constant, it is important to calibrate a relation that is independent of the value of h used to build the cosmological simulations[1].

Now our goal is to search for a relation between the different values of the factor κ and the corresponding background density contrasts used to define the halos. For this purpose, we rely on a second-degree polynomial of the form:

$$\kappa(x) = a + bx + cx^2 \quad \text{with} \quad x = \log(\Delta_{\rm b}/200) , \tag{7.2}$$

[1]For a review of the complications associated with quantities that depend explicitly on h see, e.g., Sánchez (2020).

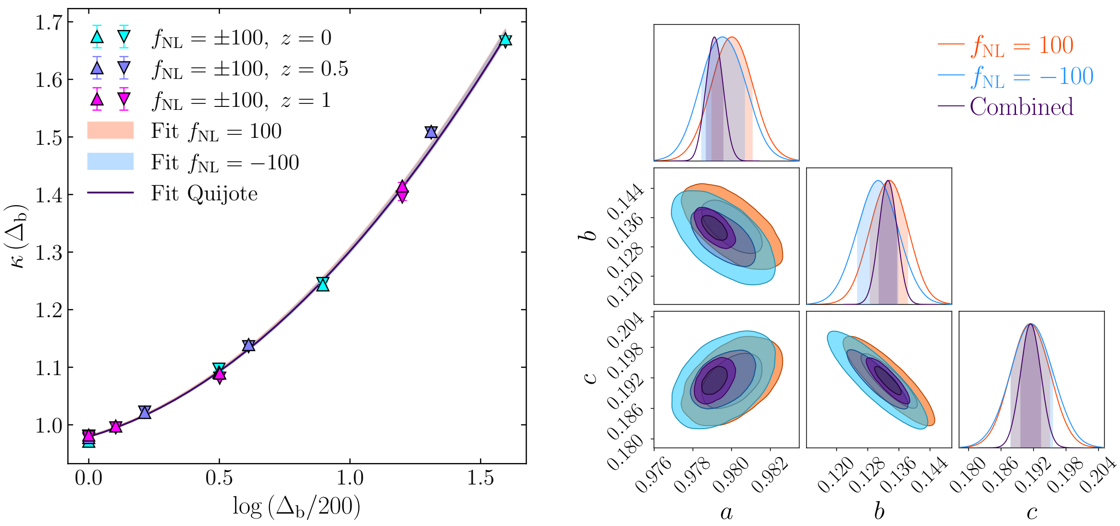

Figure 7.1: *Left*: factor κ as a function of the overdensity $\Delta_{\rm b}$. Different markers indicate the two opposite values of the $f_{\rm NL}$ parameter in the Quijote-png simulations (upward-pointing triangles for $+100$ and downward-pointing triangles for -100), while colors correspond to redshifts $z = 0$, 0.5, and 1 (cyan, light purple, and magenta, respectively). In red and blue we indicate the 1σ confidence region around the best fit of the $f_{\rm NL} = 100$ and -100 data points, respectively (note that these regions are almost perfectly overlapping). The dark purple curve represents the simultaneous fit of both data sets, including the effect of covariance. *Right:* 68% and 95% 2D confidence contours for the parameters a, b, and c characterizing the second-degree polynomials (Eq. 7.2) for each fitting procedure. The same colors are used to indicate the results for $f_{\rm NL} = 100$, -100 and their combination. The projected 1D marginalized posterior distributions are shown at the top of each column along with the 68% uncertainty (shaded bars).

and fit the coefficients a, b and c for the two scenarios $f_{\rm NL} = \pm100$. The best-fit values of these coefficients are $a = 0.980 \pm 0.001$ (0.979 ± 0.001), $b = 0.134 \pm 0.005$ (0.131 ± 0.005), $c = 0.191 \pm 0.004$ (0.191 ± 0.004) for $f_{\rm NL} = 100$ (-100). The results of this fit are shown in Fig. 7.1. In the *left panel* all the values of κ are displayed as a function of $\Delta_{\rm b}$. From this plot, we can appreciate how the data points closely follow the shape of the second-degree polynomial and how the results for $f_{\rm NL} = 100$ and $f_{\rm NL} = -100$ are highly compatible. In the *right panel* of Fig. 7.1 we instead show the posterior distributions of the polynomial coefficients derived from the fit of $\kappa(\Delta_{\rm b})$. As a further confirmation, here we can see that the confidence contours for $f_{\rm NL} = \pm100$ cosmologies are consistent within 1σ. We conclude therefore that our re-parameterization of the halo mass function model can be considered independent of the sign of $f_{\rm NL}$.

Given the latter result, a simple improvement of the analysis just presented consists in modeling simultaneously the deviations measured for $f_{\rm NL} = 100$ and $f_{\rm NL} = -100$ to increase the precision in the calibration of the parameters a, b and c. However, since realizations with opposite sign of $f_{\rm NL}$ share the same random seed for initial conditions, these data sets cannot be considered as independent. Therefore, to follow this strategy we need to consider the covariance between the mass bins in the two different cosmological scenarios and across all mass definitions used.

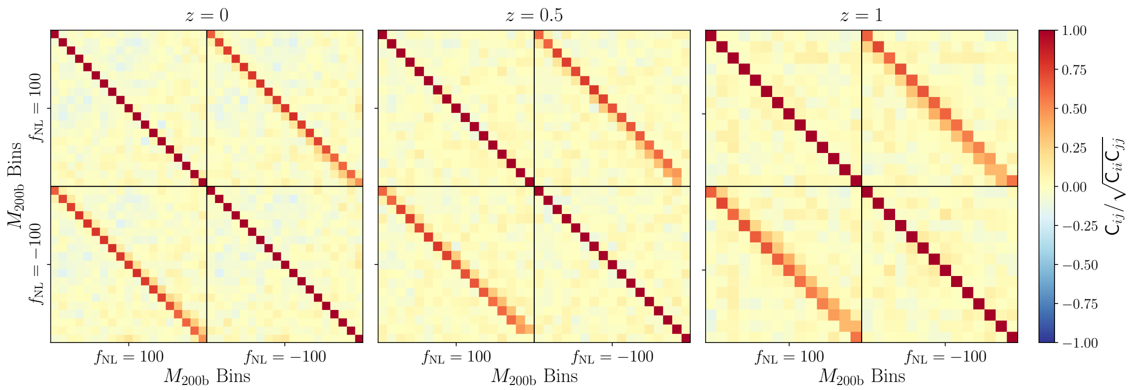

Figure 7.2: Covariance matrices (normalized by their diagonal elements) of the residuals between the 2×500 individual halo mass functions measured in QUIJOTE-PNG for $f_{\rm NL} = 100$ and -100 and the mean one measured in the QUIJOTE standard ΛCDM simulations. Each subplot corresponds to the M_{200b} covariance matrix at different redshifts: 0 (*left*), 0.5 (*center*), and 1 (*right*). Each matrix is divided into four blocks separated by vertical and horizontal black lines to isolate the two different cosmological scenarios.

Figure 7.2 presents the covariance matrix C normalized by the diagonal elements, estimated from the residuals between the individual 2×500 halo mass functions M_{200b} measured in each realization of QUIJOTE-PNG (500 for $f_{\rm NL} = 100$ and 500 for $f_{\rm NL} = -100$) and the mean halo mass function extracted from QUIJOTE ($f_{\rm NL} = 0$). Similar covariance matrices are also obtained for the other mass definitions.

Analytically, for a given mass definition and redshift, the full covariance matrix C is formed by 4 blocks, such that

$$\mathsf{C} = \begin{pmatrix} \mathsf{C}^{++} & \mathsf{C}^{+-} \\ \mathsf{C}^{-+} & \mathsf{C}^{--} \end{pmatrix} , \tag{7.3}$$

where C^{++} and C^{--} are the autocovariances of the residuals for $f_{\rm NL} = 100$ and $f_{\rm NL} = -100$, respectively, while C^{+-} and C^{-+} account for the cross-covariances between the two cosmological scenarios. Each block is computed as

$$\mathsf{C}_{ij}^{\alpha\beta} = \frac{1}{N-1} \sum_{k=1}^{N} \left(R_i^{\alpha,k} - \bar{R}_i^{\alpha} \right) \left(R_j^{\beta,k} - \bar{R}_j^{\beta} \right) , \tag{7.4}$$

where $R_i^{\alpha,k}$ is the residual in the i-th mass bin of the k-th realization of the scenario α (with $\alpha, \beta = +$ or $-$), $\bar{R}_i^{\alpha}$ is the mean residual in that bin over all realizations, and $N = 500$ is the number of realizations per scenario. Finally, each element of the matrix is rescaled by a factor N to account for the number of realizations; in this way, the diagonal elements will correctly represent the error on the mean of the residuals.

Figure 7.2 also shows a strong correlation between the same mass bins in the $f_{\rm NL} = \pm 100$ data sets, whereas for the other bins, the correlation is consistent with zero, except for minor fluctuations due to noise. To avoid possible inaccuracies due to the latter, we set to zero all elements that do not belong to the diagonals of the four blocks in each matrix. The resulting covariance matrices were then used as input for the MCMC analysis to constrain κ for each redshift and mass definition, generalizing Eq. 7.1 as follows:

$$\mathcal{L}\left(\boldsymbol{x}|\mathsf{C},\boldsymbol{\mu}\right) = \frac{1}{(2\pi)^{N_{\mathrm{d}}/2}\det(\mathsf{C})^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\mathsf{C}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right), \tag{7.5}$$

where N_{d} is the total number of data points, $\boldsymbol{x}$ is the data vector, and $\boldsymbol{\mu}$ is the vector of expected values. The results for this case are also shown in Fig. 7.1. As expected, the new best fit of $\kappa(x)$ is consistent with that obtained from the analysis of the individual cases $f_{\mathrm{NL}} = 100$ and $f_{\mathrm{NL}} = -100$. This is evident in the *right panel*, where the posterior distribution of the second-degree polynomial parameters is shown to be consistent with those of the earlier cases. The best fit we obtained for this case is:

$$\kappa(x) = 0.9792 + 0.132x + 0.190x^2\,, \tag{7.6}$$

with x defined in Eq. 7.2. The errors associated with each polynomial parameter are $\Delta a = 0.0004$, $\Delta b = 0.002$, and $\Delta c = 0.002$. Consequently, we observe a reduction in the errors of approximately 50% compared to those obtained from individual fits.

Finally, Fig. 7.3 compares data with the new parameterization for the LMSVQ model across redshifts and mass definitions $M_{200\mathrm{b}}$, $M_{200\mathrm{c}}$, and $M_{500\mathrm{c}}$. With the addition of the factor $\kappa\left(\Delta_{\mathrm{b}}\right)$, the model accurately reproduces the data in all considered cases (including $M_{2500\mathrm{c}}$). Notably, as shown in the zoomed region of the *top panel*, now we accurately capture the point at low masses where the effect of f_{NL} on halo counts reverses, which was not well constrained by the original implementation.

To evaluate the goodness of this fit, we calculate the reduced chi-square, $\tilde{\chi}^2$, using the total data vector that includes the two scenarios $f_{\mathrm{NL}} = \pm 100$ and accounting for covariance. In this case, the $\tilde{\chi}^2$ is calculated from the likelihood (Eq. 7.5) as

$$\tilde{\chi}^2 \equiv \frac{1}{N_{\mathrm{d}}-1}\left[(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\mathsf{C}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]. \tag{7.7}$$

When considering the LMSVQ model reparameterized by means of κ, the agreement with the data is remarkable, as $\tilde{\chi}^2$ is typically of order $2-3$. For example, at $z=0$ we observe $\tilde{\chi}^2 = 3,\ 3.2$ and 1.2 for the definitions of mass $M_{200\mathrm{c}}$, $M_{500\mathrm{c}}$ and $M_{2500\mathrm{c}}$, respectively. The improvement is particularly significant, as $\tilde{\chi}^2$ evaluated with the original model is 155, 553 and 1265, respectively for the same mass definitions. Similar results are also obtained at higher redshifts.

7.1.2 Testing different f_{NL} values

In the previous section, we demonstrated how the correction factor κ appears to be independent of the sign of f_{NL}. To further test the non-Gaussian halo mass function re-parameterization, we compare it with the data extracted from Pringls, which also includes scenarios with $f_{\mathrm{NL}} = \pm 40$. Thus, the same analysis performed with Quijote-png is now repeated using the halos identified in Pringls. The Pringls sets with $f_{\mathrm{NL}} = \pm 100$ will be used to verify the consistency with the results obtained with Quijote-png.

Now, because of the limited number of realizations, we expect the derived constraints to be less precise. Additionally, the reduced number of realizations prevents

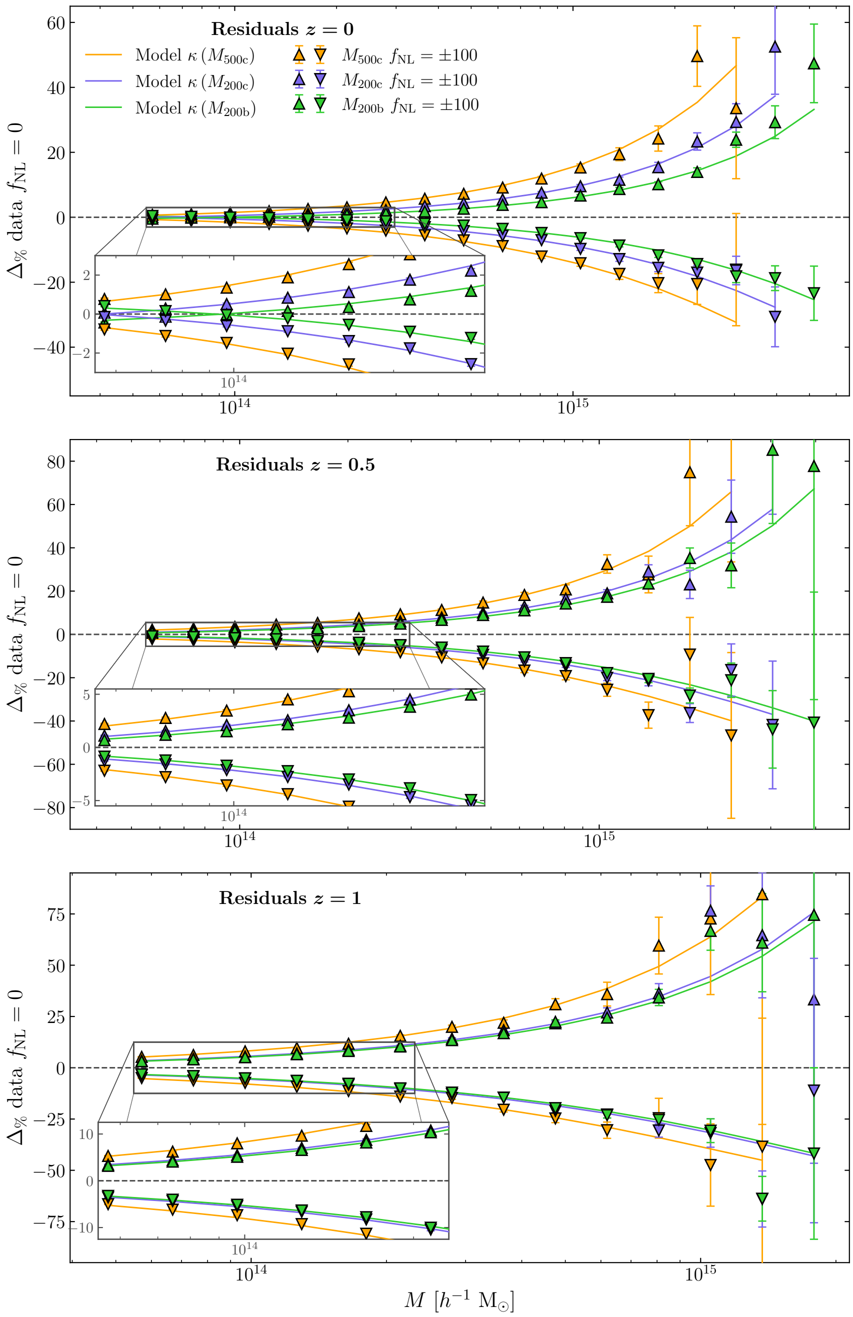

Figure 7.3: *From top to bottom*: residuals in percentage (Eq. 6.9) of the halo mass function correction as a function of mass for various mass definitions across different redshifts. Markers indicate opposite $f_{\rm NL}$ values, with upward-pointing triangles for $f_{\rm NL} = 100$ and downward-pointing triangles for $f_{\rm NL} = -100$. Green, blue and orange refers to the mass definitions $M_{200\rm b}$, $M_{200\rm c}$, $M_{500\rm c}$, respectively. Finally, with solid lines we indicate the LMSVQ model corrected by the κ factor. The colors of the lines matches the colors of the corresponding mass definition. In each panel the subplot displays a zoomed region at low masses to allow a clearer understanding of the behavior of both model and data points.

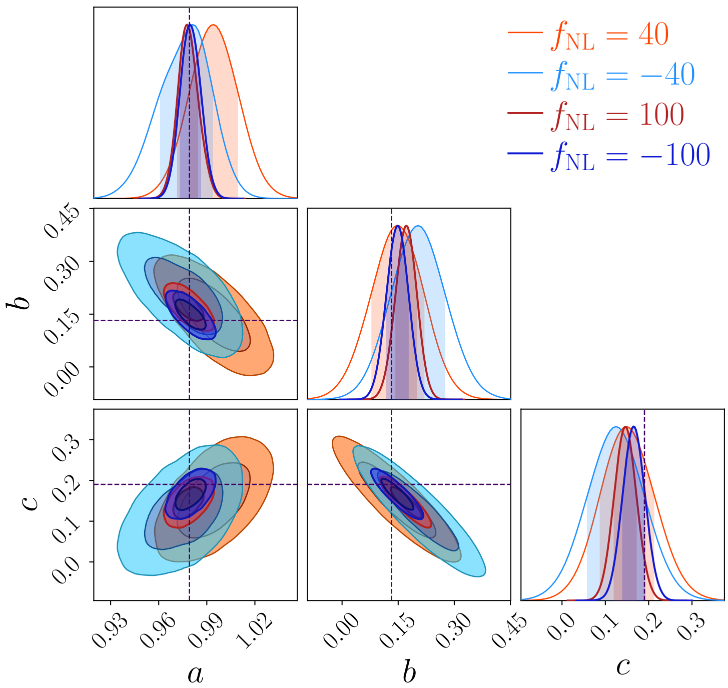

Figure 7.4: 68% and 95% confidence levels for the coefficient a, b, and c of the second-degree polynomial used to fit the values of $\kappa(\Delta_{\rm b})$ derived with PRINGLS simulations. Blue, light blue, orange and red correspond to the different values of $f_{\rm NL}$: -100, -40, 40, 100, respectively. Black dashed lines are the best fit values of the parameters reported in Eq. 7.6, which we obtained using QUIJOTE-PNG. The projected 1D marginalized posterior distributions are shown at the top of each column along with the 68% uncertainty (shaded bars).

us from simultaneously modeling halo number counts with different values of $f_{\rm NL}$: in this case the covariance matrix is dominated by noise and does not allow us to accurately estimate the correlation between the mass bins of different sets of simulations. Consequently, we will continue testing the calibration of κ using individual fits as in Sect. 7.1.

Also in this case, we measure the mass function of halos identified with ROCKSTAR, transforming the thresholds $\Delta_{\rm c}$ (200, 300, 500, 800 1000 and 1200) in their corresponding value $\Delta_{\rm b}$ at $z = 0, 0,5$ and 1. Then, for each threshold value we find the correction factor κ required to match the measured deviations from the Gaussian halo mass function (Eq. 6.9). Consequently, we fit the coefficients a, b, and c of the second-degree polynomial used to parameterize κ as a function of $\Delta_{\rm b}$ (see Eq. 7.2).

We present the main outcome of this analysis in Fig. 7.4, where we show the confidence contours relative to a, b and c for the four scenarios $f_{\rm NL} = \pm 40$ and $f_{\rm NL} = \pm 100$. We note that the posterior distributions of these parameters are in good agreement with the best-fit values obtained with QUIJOTE-PNG (vertical and horizontal dashed lines in the plot, also reported in Eq. 7.6). A minor deviation is observed for the $f_{\rm NL} = 100$ case, which is, however, compatible at the 2σ level with our reference values. We consider this minor discrepancy negligible for our analysis, attributing it to statistical fluctuations due to the presence of noise in the data. Despite using the same number of realizations, we observe in Fig. 7.4 that the contours for the $f_{\rm NL} = \pm 40$ scenarios are broader. This is because for $f_{\rm NL}$ values closer to zero, the differences from the ΛCDM case are smaller, and the relatively large errors of the residuals limit the constraining power.

Most importantly, given the compatibility of the constraints obtained for all the scenarios analyzed, we can conclude that the calibrated relation for $\kappa(\Delta_{\rm b})$ is not only independent of the sign of $f_{\rm NL}$, but also of its absolute value. We attribute this result to the fact that, in the expression for the halo mass function correction, all the information on the magnitude of $f_{\rm NL}$ is already well encapsulated in the higher-order cumulants of the distribution, S_3 and S_4. Having no statistical evidence to the contrary, we do not need to extend the reparameterization of the collapse threshold $\delta_{\rm c}$ to include this additional dependency. This outcome is clearly very important: given a sample of DM halos identified with any threshold Δ, we now have a unique and simple recipe to predict the mass function in scenarios featuring whatever level of local-type PNG.

As a final consideration, we note that the proposed re-parametrization of $\delta_{\rm c}$ by means of the factor $\kappa(\Delta_{\rm b})$ can be applied also to other non-Gaussian halo mass function models, for example MVJ and DMNP. We confirm that this modification is indeed effective in correcting the model predictions according to the selected halo density thresholds, although the best agreement with the data is achieved using the LSMVQ model. Since we used the latter to calibrate the $\kappa(\Delta_{\rm b})$ relation, we recommend applying Eq. 7.6 solely in combination with the LMSVQ model for an accurate prediction of the non-Gaussian halo mass function.

7.1.3 Calibration against high resolution simulations

Defining halo masses within a radius enclosing a given overdensity Δ implies that the halo mass can be derived from the integral of the density profile within a fixed radius. Consequently, the mass depends on the internal density distribution of the halo and is therefore more sensitive to resolution effects, especially at high overdensities. As discussed in Sect. 6.3.1, the inner regions of DM halos, particularly at low masses, are not well resolved (*top panels* of Fig. 6.6). To evaluate the impact of resolution on our parameterization, we repeated the analysis performed with the PRINGLS suite using its higher resolution counterpart, PRINGLS-HR.

This set consists of only three realizations: one for the Gaussian case and two for $f_{\rm NL} = 100$ and $f_{\rm NL} = -100$. Despite the increase in the number of particles, the number of realizations implies that the confidence contours will be significantly wider than those obtained from QUIJOTE. We compute the halo mass functions from the halos identified with `ROCKSTAR`, adopting the same overdensity thresholds used in PRINGLS. After converting the critical overdensity threshold $\Delta_{\rm c}$ into the corresponding $\Delta_{\rm b}$ values at $z = 0,\ 0.5$, and 1, we determine the correction factor κ required to match the residuals across all mass definitions and redshifts. Finally, we extract through a MCMC analysis the three parameters defining the second-order polynomial used to model the behavior of κ.

The results of this analysis are shown in Fig. 7.5. We notice how the contours derived from the residuals for $f_{\rm NL} = 100$ and -100 are fully compatible, high-

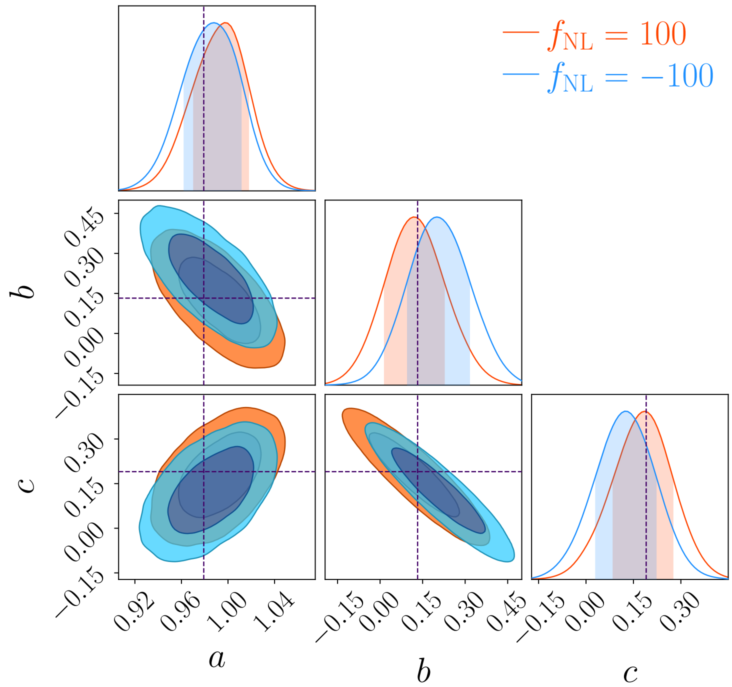

Figure 7.5: 68% and 95% confidence levels for the coefficients a, b, and c of the second-degree polynomial used to fit $\kappa(\Delta_{\rm b})$ from PRINGLS-HR simulations. Light blue and orange contours correspond to $f_{\rm NL} = -100$ and $f_{\rm NL} = 100$, respectively. Black dashed lines indicate the best-fit values of the parameters reported in Eq. 7.6, obtained using QUIJOTE-PNG. The projected 1D marginalized posterior distributions are shown at the top of each column, with shaded bars representing the 68% uncertainty.

lighting that the independence of this new parameterization of the sign of $f_{\rm NL}$ is robust even when considering high-resolution simulations. Moreover, we observe from the posterior distributions, that the results obtained with the QUIJOTE-PNG data sets (vertical and horizontal dashed lines) are compatible within 1σ with those of PRINGLS-HR. Therefore, we conclude that our new parameterization seems to be not significantly affected by resolution effects.

7.2 Impact on Forecasted Constraints

Equipped with a new prescription to correct the non-Gaussian halo mass function model for different halo identification thresholds, we now investigate its impact in forecasting constraints on $f_{\rm NL}$. We set $f_{\rm NL}$ as a free parameter with uniform prior and we compare the outcomes of two models: the original LMSVQ halo mass function correction (Eq. 4.17) and its re-parametrized version, in which the halo linear collapse threshold is multiplied by the function $\kappa(\Delta_{\rm b})$ expressed in Eq. 7.6.

With the goal of building up a more realistic setting, we use a subset of QUIJOTE-PNG simulations that cover a volume of $50\ h^{-3}\ \mathrm{Gpc}^3$, similar to those anticipated for stage IV spectroscopic surveys like *Euclid* (Euclid Collaboration et al., 2024) and *Dark Energy Spectroscopic Instrument* (DESI, DESI Collaboration et al., 2016). In practice, we use the mean halo number counts in the mass range from 5×10^{13} to

Mass	f_{NL} from LMSVQ			f_{NL} from LMSVQ with $\kappa(\Delta_{\mathrm{b}})$		
	$z=0$	$z=0.5$	$z=1$	$z=0$	$z=0.5$	$z=1$
$M_{200\mathrm{b}}$	$91.5^{+4.4}_{-4.0}$	$90.8^{+2.7}_{-3.0}$	$92.1^{+2.1}_{-1.8}$	$105.1^{+4.9}_{-4.7}$	$101.2^{+3.2}_{-3.2}$	$98.9^{+2.3}_{-1.8}$
$M_{200\mathrm{c}}$	$142.6^{+6.5}_{-7.1}$	$111.6^{+3.4}_{-2.9}$	$98.1^{+2.1}_{-1.9}$	$99.7^{+4.2}_{-4.5}$	$101.2^{+2.7}_{-3.0}$	$98.9^{+2.2}_{-1.9}$
$M_{500\mathrm{c}}$	$237.3^{+11.8}_{-11.5}$	$182.4.8^{+4.9}_{-5.5}$	$139.8^{+3.4}_{-3.5}$	$99.9^{+3.6}_{-3.9}$	$98.5^{+2.6}_{-3.0}$	$98.9^{+2.7}_{-2.0}$
$M_{2500\mathrm{c}}$	$305.0^{+33.9}_{-34.8}$	$633.1^{+24.8}_{-25.3}$	$364.1^{+22.5}_{-16.3}$	$100.4^{+3.2}_{-3.2}$	$96.7^{+3.4}_{-3.1}$	$94.7^{+4.4}_{-5.0}$

Table 7.2: Constraints on f_{NL} obtained by fitting the mean non-Gaussian mass functions extracted from 50 realizations of the QUIJOTE-PNG simulations with $f_{\mathrm{NL}} = 100$, for different halo mass definitions and redshifts. The left part of the table shows the values obtained using the original LMSVQ correction (Eq. 4.17), while the right part refers to its re-parametrized version, that includes the dependency on the halo mass definition trough the previously calibrated factor $\kappa(\Delta_{\mathrm{b}})$.

$7 \times 10^{15}\ h^{-1}\ \mathrm{M}_{\odot}$ derived from the 50 realizations of QUIJOTE-PNG characterized by $f_{\mathrm{NL}} = 100$, and we employ the mean Gaussian halo mass function derived from QUIJOTE to compute the residuals as in Eq. 6.9. We then perform a MCMC analysis, to determine the best-fit value of the parameter f_{NL} and the corresponding uncertainty. The analysis is repeated for different halo mass definitions ($M_{200\mathrm{b}}$, $M_{200\mathrm{c}}$, $M_{500\mathrm{c}}$, $M_{2500\mathrm{c}}$) and redshifts (0, 0.5, 1) to check for possible systematic errors.

The best-fit values of f_{NL} and their 1σ uncertainty are presented in Table 7.2. For the sake of clarity, we remind the reader that these results should be compared with the true cosmological parameters of the simulations, which feature $f_{\mathrm{NL}} = 100$. We can notice that employing the original LMSVQ model leads to biased constraints, with a deviation from the true value that increases with the threshold used to identify the halos. This offset is present at all the analyzed redshifts with a similar intensity. As expected from the results presented in Sect. 7.1, the case with $M_{200\mathrm{b}}$ is the least affected by systematic errors.

In contrast, including $\kappa(\Delta_{\mathrm{b}})$ in the model significantly improves the accuracy of the constraints, which are almost always consistent with the true simulation value at the 1σ level. Minor systematic errors emerge only for extreme mass definitions (i.e. $M_{2500\mathrm{c}}$) and at $z \geq 0.5$.

This result provides further confirmation that, without accounting for the halo identification threshold in the halo mass function correction for PNG, significant deviations from the theory arise. This systematic error, if not corrected, could potentially compromise upcoming cosmological constraints on PNG derived with cluster counts. It is therefore essential to incorporate this dependency on the halo definition into the model in order to avoid biased estimates of the parameters f_{NL}.

Chapter 8

Conclusions and Future Perspectives

PNG refer to deviations from the Gaussian distribution of primordial density fluctuations. These deviations can provide crucial insights into the physics governing inflation, the brief period of rapid expansion in the very early Universe that gave rise to these fluctuations. Studying PNG is particularly important in the era of *precision cosmology*, where highly accurate measurements of the cosmic microwave background and LSS will allow us to probe cosmological models with unprecedented precision. In this context, LSS offer a promising route for detecting potential signatures of PNG, as such non-Gaussianities should influence halo statistics. This can help refine our understanding of the evolution of the Universe and the fundamental physics that governs it.

The goal of this Thesis was to use numerical simulations to study the effects caused by local-type of PNG, which is quantified by the parameter $f_{\rm NL}$. The data were extracted from halo catalogs generated by applying the halo finder `ROCKSTAR` (Behroozi et al., 2013) to the four sets of N-body cosmological simulations Quijote, Quijote-png, Pringls and Pringls-hr (see Sects. 5.1 and 5.2). These simulations share the same standard ΛCDM cosmological parameters and together cover different levels of PNG, i.e. $f_{\rm NL} = 0$, $f_{\rm NL} = \pm 40$ and $f_{\rm NL} = \pm 100$. A key point in our analysis consisted in the creation of catalogs with different thresholds Δ for halo identification. In particular, we focused on different threshold values, defined both with respect to $\rho_{\rm b}$ (e.g. $\Delta_{\rm b} = 200$) and $\rho_{\rm crit}$ (e.g. $\Delta_{\rm c} = 200$, 500, 2500).

We started our analysis in Sect. 6.1 by measuring the scale-dependent bias of DM halos defined with different mass definitions at different redshifts, employing the Quijote-png simulations. We then compared the measured behavior of the bias with the theoretical prescription presented in Dalal et al. (2008) and Desjacques et al. (2018), showing that no significant deviations emerge.

Then, in Sect. 6.2, we focused on testing the accuracy of already existing halo mass function models developed for cosmological scenarios featuring PNG. We selected three well-known models from the literature, that is, Matarrese et al. (2000), LoVerde et al. (2008), and D'Amico et al. (2011a), which provide a theoretical prescription to predict the deviation from the Gaussian halo mass function produced by local type of PNG. We compared their predictions with the halo number counts extracted from the halo catalogs: we found that none of them was able to perfectly capture the deviations caused by PNG on the halo mass function. We showed how this discrepancy depends on the halo mass definition, as it increases significantly when higher density thresholds are used.

To investigate the reason for this systematic error, in Sect. 6.3.1 we compared the stacked density profiles of DM halos measured in simulations with different values of $f_{\rm NL}$. We showed that PNG modify the DM matter distribution around the halo center such that for $f_{\rm NL} > 0$ halos exhibit a steeper inner density profile, whereas the opposite occurs for $f_{\rm NL} < 0$. Moreover, the change in the slope of the halo density profile becomes more significant as the value of $|f_{\rm NL}|$ increases. This behavior explains why using different identification thresholds produces a deviation in the observed halo mass function. In fact, halos become more (less) compact in a $f_{\rm NL} > 0$ (< 0) cosmology, and the higher the identification threshold, i.e. closer we get to the halo center, the larger (smaller) their mass will result with respect to a standard ΛCDM scenario.

At this point, the main part of our work consisted of developing a correction to the theoretical model for the non-Gaussian halo mass function that incorporates the dependence on the halo mass definition. We took as a reference the quadratic model of LoVerde et al. (2008) since it was the one showing the best agreement with the measured halo mass function (see $\Delta_{\rm b} = 200$ data in Fig. 6.4). The prescription we propose is simple: we correct the linear density threshold for halo collapse, $\delta_{\rm c}$, by means of a factor $\kappa(\Delta_{\rm b})$. The latter is calibrated as a function of the halo identification threshold, which we require to be expressed in terms of the background density of the Universe, $\rho_{\rm b}(z)$.

In Sect. 7.1.1 we calibrated this correction factor by fitting the deviations on the measured halo mass function caused by PNG at different redshifts ($z = 0$, 0.5, 1), and for halos identified with different thresholds (converting those overdensities relative to $\rho_{\rm crit}$ into the corresponding background quantity). We found that a second-degree polynomial can accurately represent the variation of κ as a function of $\Delta_{\rm b}(z)$, for all analyzed redshifts. We estimated the best-fit values for the three polynomial coefficients and their relative uncertainty, finding that a unique relation can be used to incorporate the dependence of the halo identification threshold into the non-Gaussian halo mass function model. In other terms, we proposed a re-parametrization of the model that holds for cosmologies with different levels of local-type PNG, independent of the sign and the amplitude of the parameter $f_{\rm NL}$ (see Sect. 7.1.2) and that is robust against resolution effects (see Sect. 7.1.3). The best fit of the function $\kappa(\Delta_{\rm b})$ is presented in Eq. 7.6, where we combined the halo number counts extracted from simulations with $f_{\rm NL} = -100$ and $f_{\rm NL} = 100$, including the covariance between these data.

Furthermore, in Sect. 7.2, we presented a simple example of cosmological forecasts focused on constraining $f_{\rm NL}$. We considered a hypothetical survey volume of $50\ h^{-3}\ \mathrm{Gpc}^3$ and compared the constraints derived with the original model and its re-parameterized version (see Table 7.2). We demonstrated how the original implementation leads to extremely biased cosmological constraints for halos identified with thresholds as $\Delta_{\rm c} = 200$ or 500, with a systematic error increasing for higher values of Δ. In contrast, we found that the model re-parametrized with $\Delta_{\rm b}(z)$ allows us to accurately constrain $f_{\rm NL}$ without statistically relevant systematic errors.

It is important to emphasize that these results are particularly relevant in the perspective of the application to real data. Indeed, wide field surveys allow us to extract the number of galaxy clusters as a function of their mass, but their definition of mass can vary depending on the wavelength analyzed. For example, optical and X-ray observations generally identify clusters with masses around $M_{\rm 200c}$ and $M_{\rm 500c}$, respectively, making it crucial to have a model that consistently adapts to any identification threshold. The re-parametrized model offers a natural solution to this need. However, a real application to survey data will require all cosmological parameters to be left free, not just $f_{\rm NL}$. Therefore, it is important to consider that non-Gaussian halo mass function models are computationally very expensive to calculate for different cosmologies, as the approximated relations in Eq. 6.8 cannot be used. For this reason, in the next follow-ups of this work, we aim at using machine learning techniques to develop emulators of the non-Gaussian halo mass function to speed up model computation, making it feasible for use in MCMC analyses.

Furthermore, we underline that the methodology proposed in this work is not meant to be limited only to local-type PNG: a natural extension of this work will consist of testing the models present in the literature also for other shapes of the potential bispectrum (i.e. equilateral and orthogonal). Moreover, one of our goals is to extend our analysis to the underdense counterpart of DM halos, namely cosmic voids. In fact, the same formalism used to predict the non-Gaussian halo mass function can be applied to void counts, allowing us to model the non-Gaussian void size function (Kamionkowski et al., 2009; D'Amico et al., 2011b). These two models can ultimately be used in combination to extract the overall information from halo and void number counts. In fact, several studies have demonstrated the strong complementarity of these two statistics (see e.g. Bayer et al., 2021; Kreisch et al., 2022; Contarini et al., 2022; Pelliciari et al., 2023), making them excellent probes for breaking key degeneracies between cosmological parameters.

Bibliography

Aarseth, S. J. 1963, MNRAS, 126, 223, doi: 10.1093/mnras/126.3.223

Abbott, T. M. C., Abdalla, F. B., Alarcon, A., et al. 2018, Phys. Rev. D, 98, 043526, doi: 10.1103/PhysRevD.98.043526

Albrecht, A., & Steinhardt, P. J. 1982, Phys. Rev. Lett., 48, 1220, doi: 10.1103/PhysRevLett.48.1220

Alishahiha, M., Silverstein, E., & Tong, D. 2004, Phys. Rev. D, 70, 123505, doi: 10.1103/PhysRevD.70.123505

Angulo, R. E., & Hahn, O. 2022, Living Reviews in Computational Astrophysics, 8, 1, doi: 10.1007/s41115-021-00013-z

Arkani-Hamed, N., Creminelli, P., Mukohyama, S., & Zaldarriaga, M. 2004, J. Cosmology Astropart. Phys., 2004, 001, doi: 10.1088/1475-7516/2004/04/001

Armendáriz-Picón, C., Damour, T., & Mukhanov, V. 1999, Physics Letters B, 458, 209, doi: 10.1016/S0370-2693(99)00603-6

Baldi, M., Fondi, E., Karagiannis, D., et al. 2024, J. Cosmology Astropart. Phys., 2024, 053, doi: 10.1088/1475-7516/2024/11/053

Bardeen, J. M., Bond, J. R., Kaiser, N., & Szalay, A. S. 1986, ApJ, 304, 15, doi: 10.1086/164143

Barnes, J., & Hut, P. 1986, Nature, 324, 446, doi: 10.1038/324446a0

Barreira, A. 2022a, J. Cosmology Astropart. Phys., 2022, 013, doi: 10.1088/1475-7516/2022/11/013

—. 2022b, J. Cosmology Astropart. Phys., 2022, 033, doi: 10.1088/1475-7516/2022/01/033

Barreira, A., Cabass, G., Schmidt, F., Pillepich, A., & Nelson, D. 2020, J. Cosmology Astropart. Phys., 2020, 013, doi: 10.1088/1475-7516/2020/12/013

Bayer, A. E., Villaescusa-Navarro, F., Massara, E., et al. 2021, ApJ, 919, 24, doi: 10.3847/1538-4357/ac0e91

Behroozi, P. S., Wechsler, R. H., & Wu, H.-Y. 2013, ApJ, 762, 109, doi: 10.1088/0004-637X/762/2/109

Bond, H. E., Nelan, E. P., VandenBerg, D. A., Schaefer, G. H., & Harmer, D. 2013, ApJ, 765, L12, doi: 10.1088/2041-8205/765/1/L12

Bond, J. R., Cole, S., Efstathiou, G., & Kaiser, N. 1991, ApJ, 379, 440, doi: 10.1086/170520

Bryan, G. L., & Norman, M. L. 1998, ApJ, 495, 80, doi: 10.1086/305262

Byrnes, C. T., & Choi, K.-Y. 2010, Advances in Astronomy, 2010, 724525, doi: 10.1155/2010/724525

Byrnes, C. T., Choi, K.-Y., & Hall, L. M. H. 2008, J. Cosmology Astropart. Phys., 2008, 008, doi: 10.1088/1475-7516/2008/10/008

Byrnes, C. T., & Wands, D. 2006, Phys. Rev. D, 74, 043529, doi: 10.1103/PhysRevD.74.043529

Cabass, G., Ivanov, M. M., Philcox, O. H. E., Simonović, M., & Zaldarriaga, M. 2022, Phys. Rev. D, 106, 043506, doi: 10.1103/PhysRevD.106.043506

Calabrese, E., Slosar, A., Melchiorri, A., Smoot, G. F., & Zahn, O. 2008, Phys. Rev. D, 77, 123531, doi: 10.1103/PhysRevD.77.123531

Castorina, E., Hand, N., Seljak, U., et al. 2019, J. Cosmology Astropart. Phys., 2019, 010, doi: 10.1088/1475-7516/2019/09/010

Chen, X., Huang, M.-x., Kachru, S., & Shiu, G. 2007, J. Cosmology Astropart. Phys., 2007, 002, doi: 10.1088/1475-7516/2007/01/002

Child, H. L., Habib, S., Heitmann, K., et al. 2018, ApJ, 859, 55, doi: 10.3847/1538-4357/aabf95

Coles, P., & Lucchin, F. 2002, Cosmology: The Origin and Evolution of Cosmic Structure, Second Edition (John Wiley & Sons Inc)

Contarini, S., Verza, G., Pisani, A., et al. 2022, A&A, 667, A162, doi: 10.1051/0004-6361/202244095

Coulton, W. R., Villaescusa-Navarro, F., Jamieson, D., et al. 2023, ApJ, 943, 64, doi: 10.3847/1538-4357/aca8a7

Creminelli, P., Nicolis, A., Senatore, L., Tegmark, M., & Zaldarriaga, M. 2006, J. Cosmology Astropart. Phys., 2006, 004, doi: 10.1088/1475-7516/2006/05/004

Creminelli, P., & Zaldarriaga, M. 2004, J. Cosmology Astropart. Phys., 2004, 006, doi: 10.1088/1475-7516/2004/10/006

Crocce, M., Fosalba, P., Castander, F. J., & Gaztañaga, E. 2010, MNRAS, 403, 1353, doi: 10.1111/j.1365-2966.2009.16194.x

Crocce, M., Pueblas, S., & Scoccimarro, R. 2006, MNRAS, 373, 369, doi: 10.1111/j.1365-2966.2006.11040.x

Cui, W., Liu, L., Yang, X., et al. 2008, ApJ, 687, 738, doi: 10.1086/592079

Dalal, N., Doré, O., Huterer, D., & Shirokov, A. 2008, Phys. Rev. D, 77, 123514, doi: 10.1103/PhysRevD.77.123514

D'Amico, G., Lewandowski, M., Senatore, L., & Zhang, P. 2022, arXiv e-prints, arXiv:2201.11518, doi: 10.48550/arXiv.2201.11518

D'Amico, G., Musso, M., Noreña, J., & Paranjape, A. 2011a, J. Cosmology Astropart. Phys., 2011, 001, doi: 10.1088/1475-7516/2011/02/001

—. 2011b, Phys. Rev. D, 83, 023521, doi: 10.1103/PhysRevD.83.023521

Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. 1985, ApJ, 292, 371, doi: 10.1086/163168

DESI Collaboration, Aghamousa, A., Aguilar, J., et al. 2016, arXiv e-prints, arXiv:1611.00036, doi: 10.48550/arXiv.1611.00036

Desjacques, V., Jeong, D., & Schmidt, F. 2018, Phys. Rep., 733, 1, doi: 10.1016/j.physrep.2017.12.002

Desjacques, V., Seljak, U., & Iliev, I. T. 2009, MNRAS, 396, 85, doi: 10.1111/j.1365-2966.2009.14721.x

Despali, G., Giocoli, C., Angulo, R. E., et al. 2016, MNRAS, 456, 2486, doi: 10.1093/mnras/stv2842

Di Valentino, E., Melchiorri, A., & Silk, J. 2020, Nature Astronomy, 4, 196, doi: 10.1038/s41550-019-0906-9

Di Valentino, E., Anchordoqui, L. A., Akarsu, Ö., et al. 2021a, Astroparticle Physics, 131, 102605, doi: 10.1016/j.astropartphys.2021.102605

—. 2021b, Astroparticle Physics, 131, 102604, doi: 10.1016/j.astropartphys.2021.102604

—. 2021c, Astroparticle Physics, 131, 102607, doi: 10.1016/j.astropartphys.2021.102607

Diemer, B., & Kravtsov, A. V. 2015, ApJ, 799, 108, doi: 10.1088/0004-637X/799/1/108

Dodelson, S., & Schmidt, F. 2020, Modern Cosmology (Academic Press), doi: 10.1016/C2017-0-01943-2

Doré, O., Bock, J., Ashby, M., et al. 2014, arXiv e-prints, arXiv:1412.4872, doi: 10.48550/arXiv.1412.4872

Dvali, G., Gruzinov, A., & Zaldarriaga, M. 2004, Phys. Rev. D, 69, 023505, doi: 10.1103/PhysRevD.69.023505

Einasto, J. 1965, Trudy Astrofizicheskogo Instituta Alma-Ata, 5, 87

Einstein, A. 1915, Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften, 844

Eisenstein, D. J., & Hu, W. 1999, ApJ, 511, 5, doi: 10.1086/306640

Enqvist, K., Hotchkiss, S., & Taanila, O. 2011, J. Cosmology Astropart. Phys., 2011, 017, doi: 10.1088/1475-7516/2011/04/017

Enqvist, K., & Sloth, M. S. 2002, Nuclear Physics B, 626, 395, doi: 10.1016/S0550-3213(02)00043-3

Erben, T., Hildebrandt, H., Miller, L., et al. 2013, MNRAS, 433, 2545, doi: 10.1093/mnras/stt928

Euclid Collaboration, Mellier, Y., Abdurro'uf, et al. 2024, arXiv e-prints, arXiv:2405.13491, doi: 10.48550/arXiv.2405.13491

Ferraro, S., & Wilson, M. J. 2019, BAAS, 51, 72, doi: 10.48550/arXiv.1903.09208

Fiorino, L., Contarini, S., Marulli, F., et al. 2024, arXiv e-prints, arXiv:2410.21457, doi: 10.48550/arXiv.2410.21457

Fondi, E., Verde, L., Villaescusa-Navarro, F., et al. 2024, J. Cosmology Astropart. Phys., 2024, 048, doi: 10.1088/1475-7516/2024/02/048

Freedman, W. L., Madore, B. F., Hatt, D., et al. 2019, ApJ, 882, 34, doi: 10.3847/1538-4357/ab2f73

Garriga, J., & Mukhanov, V. F. 1999, Physics Letters B, 458, 219, doi: 10.1016/S0370-2693(99)00602-4

Ghirardini, V., Bulbul, E., Artis, E., et al. 2024, A&A, 689, A298, doi: 10.1051/0004-6361/202348852

Gong, J.-O. 2017, International Journal of Modern Physics D, 26, 1740003, doi: 10.1142/S021827181740003X

Grossi, M., Dolag, K., Branchini, E., Matarrese, S., & Moscardini, L. 2007, MNRAS, 382, 1261, doi: 10.1111/j.1365-2966.2007.12458.x

Grossi, M., Verde, L., Carbone, C., et al. 2009, MNRAS, 398, 321, doi: 10.1111/j.1365-2966.2009.15150.x

Gunn, J. E., & Gott, J. Richard, I. 1972, ApJ, 176, 1, doi: 10.1086/151605

Guth, A. H. 1981, Phys. Rev. D, 23, 347, doi: 10.1103/PhysRevD.23.347

Guth, A. H., & Pi, S. Y. 1982, Phys. Rev. Lett., 49, 1110, doi: 10.1103/PhysRevLett.49.1110

Gutiérrez Adame, A., Avila, S., Gonzalez-Perez, V., et al. 2024, A&A, 689, A69, doi: 10.1051/0004-6361/202349037

Handley, W. 2021, Phys. Rev. D, 103, L041301, doi: 10.1103/PhysRevD.103.L041301

Hildebrandt, H., Köhlinger, F., van den Busch, J. L., et al. 2020, A&A, 633, A69, doi: 10.1051/0004-6361/201834878

Hockney, R. W., & Eastwood, J. W. 1981, Computer Simulation Using Particles (CRC press)

Hogg, D. W., & Foreman-Mackey, D. 2018, ApJS, 236, 11, doi: 10.3847/1538-4365/aab76e

Hu, W., & Dodelson, S. 2002, ARA&A, 40, 171, doi: 10.1146/annurev.astro.40.060401.093926

Izumi, K., & Mukohyama, S. 2010, J. Cosmology Astropart. Phys., 2010, 016, doi: 10.1088/1475-7516/2010/06/016

Jeans, J. H. 1902, Philosophical Transactions of the Royal Society of London Series A, 199, 1, doi: 10.1098/rsta.1902.0012

Jenkins, A. 2010, MNRAS, 403, 1859, doi: 10.1111/j.1365-2966.2010.16259.x

Jenkins, A., Frenk, C. S., White, S. D. M., et al. 2001, MNRAS, 321, 372, doi: 10.1046/j.1365-8711.2001.04029.x

Jimenez, R., Cimatti, A., Verde, L., Moresco, M., & Wandelt, B. 2019, J. Cosmology Astropart. Phys., 2019, 043, doi: 10.1088/1475-7516/2019/03/043

Joyce, A., Lombriser, L., & Schmidt, F. 2016, Annual Review of Nuclear and Particle Science, 66, 95–122, doi: 10.1146/annurev-nucl-102115-044553

Juszkiewicz, R., Weinberg, D. H., Amsterdamski, P., Chodorowski, M., & Bouchet, F. 1995, ApJ, 442, 39, doi: 10.1086/175420

Kaiser, N. 1984, ApJ, 284, L9, doi: 10.1086/184341

Kamionkowski, M., Verde, L., & Jimenez, R. 2009, J. Cosmology Astropart. Phys., 2009, 010, doi: 10.1088/1475-7516/2009/01/010

Karagiannis, D., Lazanu, A., Liguori, M., et al. 2018, MNRAS, 478, 1341, doi: 10.1093/mnras/sty1029

Kitayama, T., & Suto, Y. 1996, ApJ, 469, 480, doi: 10.1086/177797

Klypin, A., Yepes, G., Gottlöber, S., Prada, F., & Heß, S. 2016, MNRAS, 457, 4340, doi: 10.1093/mnras/stw248

Kofman, L. 2003, arXiv e-prints, astro, doi: 10.48550/arXiv.astro-ph/0303614

Komatsu, E., & Spergel, D. N. 2001, Phys. Rev. D, 63, 063002, doi: 10.1103/PhysRevD.63.063002

Kreisch, C. D., Pisani, A., Villaescusa-Navarro, F., et al. 2022, ApJ, 935, 100, doi: 10.3847/1538-4357/ac7d4b

Langlois, D., & Vernizzi, F. 2004, Phys. Rev. D, 70, 063522, doi: 10.1103/PhysRevD.70.063522

Lazarides, G., Ruiz de Austri, R., & Trotta, R. 2004, Phys. Rev. D, 70, 123527, doi: 10.1103/PhysRevD.70.123527

Lazeyras, T., Barreira, A., Schmidt, F., & Desjacques, V. 2023, J. Cosmology Astropart. Phys., 2023, 023, doi: 10.1088/1475-7516/2023/01/023

Lesci, G. F., Marulli, F., Moscardini, L., et al. 2022, A&A, 659, A88, doi: 10.1051/0004-6361/202040194

Lesgourgues, J. 2011, arXiv e-prints, arXiv:1104.2932, doi: 10.48550/arXiv.1104.2932

Lewis, A., Challinor, A., & Lasenby, A. 2000, ApJ, 538, 473, doi: 10.1086/309179

Liguori, M., Matarrese, S., & Moscardini, L. 2003, ApJ, 597, 57, doi: 10.1086/378394

Linde, A. D. 1982, Physics Letters B, 108, 389, doi: 10.1016/0370-2693(82)91219-9

LoVerde, M., Miller, A., Shandera, S., & Verde, L. 2008, J. Cosmology Astropart. Phys., 2008, 014, doi: 10.1088/1475-7516/2008/04/014

Lyth, D. H., & Wands, D. 2002, Physics Letters B, 524, 5, doi: 10.1016/S0370-2693(01)01366-1

Maksimova, N. A., Garrison, L. H., Eisenstein, D. J., et al. 2021, MNRAS, 508, 4017, doi: 10.1093/mnras/stab2484

Maldacena, J. 2003, Journal of High Energy Physics, 2003, 013, doi: 10.1088/1126-6708/2003/05/013

Marulli, F., Veropalumbo, A., & Moresco, M. 2016, Astronomy and Computing, 14, 35, doi: 10.1016/j.ascom.2016.01.005

Matarrese, S., Verde, L., & Jimenez, R. 2000, The Astrophysical Journal, 541, 10–24, doi: 10.1086/309412

Meszaros, P. 1974, A&A, 37, 225

More, S., Kravtsov, A. V., Dalal, N., & Gottlöber, S. 2011, ApJS, 195, 4, doi: 10.1088/0067-0049/195/1/4

Moroi, T., & Takahashi, T. 2001, Physics Letters B, 522, 215, doi: 10.1016/S0370-2693(01)01295-3

Moroi, T., Takahashi, T., & Toyoda, Y. 2005, Phys. Rev. D, 72, 023502, doi: 10.1103/PhysRevD.72.023502

Mueller, E.-M., Rezaie, M., Percival, W. J., et al. 2022, MNRAS, 514, 3396, doi: 10.1093/mnras/stac812

Mukhanov, V. F., & Chibisov, G. V. 1981, Soviet Journal of Experimental and Theoretical Physics Letters, 33, 532

Mukherjee, S., Ghosh, A., Graham, M. J., et al. 2020, arXiv e-prints, arXiv:2009.14199. https://arxiv.org/abs/2009.14199

Musso, M., & Sheth, R. K. 2012, MNRAS, 423, L102, doi: 10.1111/j.1745-3933.2012.01266.x

Nandi, D. 2017, arXiv e-prints, arXiv:1707.02976, doi: 10.48550/arXiv.1707.02976

Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, ApJ, 490, 493, doi: 10.1086/304888

Paranjape, A., Sheth, R. K., & Desjacques, V. 2013, MNRAS, 431, 1503, doi: 10.1093/mnras/stt267

Peebles, P. J. E. 1970, AJ, 75, 13, doi: 10.1086/110933

Pelliciari, D., Contarini, S., Marulli, F., et al. 2023, MNRAS, 522, 152, doi: 10.1093/mnras/stad956

Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, ApJ, 517, 565, doi: 10.1086/307221

Pillepich, A., Porciani, C., & Hahn, O. 2010, MNRAS, 402, 191, doi: 10.1111/j.1365-2966.2009.15914.x

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, A&A, 594, A13, doi: 10.1051/0004-6361/201525830

Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020a, A&A, 641, A6, doi: 10.1051/0004-6361/201833910

Planck Collaboration, Akrami, Y., Arroja, F., et al. 2020b, A&A, 641, A9, doi: 10.1051/0004-6361/201935891

Planck Collaboration, Akrami, Y., Ashdown, M., et al. 2020c, A&A, 641, A4, doi: 10.1051/0004-6361/201833881

Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020d, A&A, 641, A1, doi: 10.1051/0004-6361/201833880

Press, W. H., & Schechter, P. 1974, ApJ, 187, 425, doi: 10.1086/152650

Riess, A. G., Casertano, S., Yuan, W., Macri, L. M., & Scolnic, D. 2019, ApJ, 876, 85, doi: 10.3847/1538-4357/ab1422

Riess, A. G., et al. 1998, AJ, 116, 1009, doi: 10.1086/300499

Riess, A. G., Macri, L. M., Hoffmann, S. L., et al. 2016, ApJ, 826, 56, doi: 10.3847/0004-637X/826/1/56

Sánchez, A. G. 2020, Phys. Rev. D, 102, 123511, doi: 10.1103/PhysRevD.102.123511

Schlaufman, K. C., Thompson, I. B., & Casey, A. R. 2018, ApJ, 867, 98, doi: 10.3847/1538-4357/aadd97

Scoccimarro, R., Hui, L., Manera, M., & Chan, K. C. 2012, Physical Review D, 85, doi: 10.1103/physrevd.85.083002

Seery, D., & Lidsey, J. E. 2005a, J. Cosmology Astropart. Phys., 2005, 003, doi: 10.1088/1475-7516/2005/06/003

—. 2005b, J. Cosmology Astropart. Phys., 2005, 011, doi: 10.1088/1475-7516/2005/09/011

Seljak, U. 2009, Phys. Rev. Lett., 102, 021302, doi: 10.1103/PhysRevLett.102.021302

Senatore, L., Smith, K. M., & Zaldarriaga, M. 2010, J. Cosmology Astropart. Phys., 2010, 028, doi: 10.1088/1475-7516/2010/01/028

Sérsic, J. L. 1963, Boletin de la Asociacion Argentina de Astronomia La Plata Argentina, 6, 41

Sheth, R. K., & Tormen, G. 1999, MNRAS, 308, 119, doi: 10.1046/j.1365-8711.1999.02692.x

Slosar, A., Hirata, C., Seljak, U., Ho, S., & Padmanabhan, N. 2008, J. Cosmology Astropart. Phys., 2008, 031, doi: 10.1088/1475-7516/2008/08/031

Smith, R. E., Desjacques, V., & Marian, L. 2011, Phys. Rev. D, 83, 043526, doi: 10.1103/PhysRevD.83.043526

Speagle, J. S. 2019, arXiv e-prints, arXiv:1909.12313, doi: 10.48550/arXiv.1909.12313

Springel, V. 2005, MNRAS, 364, 1105, doi: 10.1111/j.1365-2966.2005.09655.x

Springel, V., White, S. D. M., Tormen, G., & Kauffmann, G. 2001, MNRAS, 328, 726, doi: 10.1046/j.1365-8711.2001.04912.x

Starobinsky, A. A. 1982, Physics Letters B, 117, 175, doi: 10.1016/0370-2693(82)90541-X

Takahashi, T. 2014, PTEP, 2014, 06B105, doi: 10.1093/ptep/ptu060

Tinker, J., Kravtsov, A. V., Klypin, A., et al. 2008, ApJ, 688, 709, doi: 10.1086/591439

Tinker, J. L., Robertson, B. E., Kravtsov, A. V., et al. 2010, ApJ, 724, 878, doi: 10.1088/0004-637X/724/2/878

Troxel, M. A., MacCrann, N., Zuntz, J., et al. 2018, Phys. Rev. D, 98, 043528, doi: 10.1103/PhysRevD.98.043528

Valcin, D., Bernal, J. L., Jimenez, R., Verde, L., & Wandelt, B. D. 2020, J. Cosmology Astropart. Phys., 2020, 002, doi: 10.1088/1475-7516/2020/12/002

Villaescusa-Navarro, F. 2018, Pylians: Python libraries for the analysis of numerical simulations, Astrophysics Source Code Library, record ascl:1811.008. http://ascl.net/1811.008

Villaescusa-Navarro, F., Hahn, C., Massara, E., et al. 2020, ApJS, 250, 2, doi: 10.3847/1538-4365/ab9d82

Wagner, C., & Verde, L. 2012, J. Cosmology Astropart. Phys., 2012, 002, doi: 10.1088/1475-7516/2012/03/002

Wagner, C., Verde, L., & Boubekeur, L. 2010, J. Cosmology Astropart. Phys., 2010, 022, doi: 10.1088/1475-7516/2010/10/022

Warren, M. S., Abazajian, K., Holz, D. E., & Teodoro, L. 2006, ApJ, 646, 881, doi: 10.1086/504962

Yokoyama, S., Sugiyama, N., Zaroubi, S., & Silk, J. 2011, MNRAS, 417, 1074, doi: 10.1111/j.1365-2966.2011.19323.x

Yoo, J., & Watanabe, Y. 2012, International Journal of Modern Physics D, 21, 1230002, doi: 10.1142/S0218271812300029

Zentner, A. R. 2007, International Journal of Modern Physics D, 16, 763, doi: 10.1142/S0218271807010511