Scuola Internazionale Superiore di Studi Avanzati - Trieste

Topics on Cosmological Non-Gaussianity

Jorge Iván Noreña Sánchez

Astroparticle Physics

Scuola Internazionale Superiore di Studi Avanzati (SISSA)

A thesis submitted for the degree of

PhilosophiæDoctor (PhD)

September 2010

SISSA - Via Bonomea 265 - 34136 TRIESTE - ITALY

Topics on Cosmological Non-Gaussianity

Jorge Iván Noreña Sánchez

Astroparticle Physics

Scuola Internazionale Superiore di Studi Avanzati (SISSA)

A thesis submitted for the degree of

PhilosophiæDoctor (PhD)

September 2010

Abstract

We study nonlinear cosmological perturbations during the post-inflationary evolution, using the equivalence between a perfect barotropic fluid and a derivatively coupled scalar field with Lagrangian $[-(\partial\phi)^2]^{(1+w)/2w}$. This approach is analogous to the one employed in the study of non-gaussianities from inflation. We use this method to derive the second order metric during matter dominance in the comoving gauge directly as function of the primordial inflationary perturbation ζ. This calculation is a first step in the estimation of the gravitational contribution to the non-gaussianity observed in the CMB. We then calculate the Cosmic Microwave Background anisotropy bispectrum on large angular scales in the absence of primordial non-gaussianities, assuming exact matter dominance and extending at second order the classic Sachs-Wolfe result $\delta T/T = \Phi/3$. The bispectrum is explicitly computed in the flat-sky approximation. It scales as l^{-4} in the scale invariant limit and the shape dependence of its various contributions is represented in 3d plots. In the squeezed limit this corresponds to $f_{\rm NL}^{\rm local} = -1/6 - \cos(2\theta)$, where θ is the angle between the short and the long modes; the angle dependent contribution comes from lensing.

Turning to the study of non-gaussianity in the large scale structure, we improve upon the existing results for the non-gaussian mass function by combining path integral methods and saddle point techniques (which have been separately applied in previous approaches). Additionally, we carefully account for the various scale dependent combinations of small parameters which appear. Some of these combinations in fact become of order unity for large mass scales and at high redshifts, and must therefore be treated non-perturbatively. We thus derive an accurate expression for the mass function which is based on approximations that are valid over a larger range of mass scales and redshifts than those of other authors. By tracking the terms ignored in the analysis, we estimate theoretical errors for our result and also for the results of others.

Acknowledgements

It has been both a pleasure and an honor to work under the supervision of Paolo Creminelli. I want to thank him not only for his insightful academic guidance, but also for showing me, with his great enthusiasm, the pleasure of doing physics.

During this work I had the luck to cross paths with several talented scientists and great friends with whom I greatly enjoyed working, and from whom I learnt much. Their names are Lotfi Boubekeur, Marcello Muso, Aseem Paranjape, and Filippo Vernizzi. Among them I especially wish to thank Guido D'Amico who shared with me the small difficulties and joys of our daily work.

My life would have been sad and dull where it not for Antonella, with whom I shared many simple joys, and many more strong emotions.

Few people, especially when living in a foreign country, have the opportunity that I had of meeting so many good friends. The list of names is happily very long, but I will take the risk of offending some by naming my classmates and companions of random adventures (of course in alphabetical order), they are Tom Barley, Luca Lepori, Emiliano Molinaro, Marco Nardecchia, Andrea Prudenziati, Raffaele Savelli, Yue Ling Tsai, and Robert Ziegler.

Finally, I wish to thank my family for their unending support.

Contents

List of Figures

List of Tables

1

Introduction

1.1 Preliminaries

In the last few decades, cosmology has seen great advances that changed the way we understand not only the history of our universe, but also the way we view nature. Both theory and observations have opened new paths of research that proved to be revolutionary. This thesis is but a small contribution to this successful field, based on part of the work I developed during the last few years, references [1; 2; 3]. The unifying topic of this thesis is the subject of non-gaussianity which is one of the windows to probe the physics of the very early universe, as I will shortly explain.

The big bang model of the universe predicts the energy density at very early times to be so high that the laws of physics governing its evolution are beyond what we can test in our laboratories. The leading hypothesis is that this very early universe underwent a period of accelerated expansion (termed inflation) driven by some field, during which quantum fluctuations of this field became much larger than the causal horizon thus becoming classical. After this period of inflation ended the perturbations reentered the horizon providing the inhomogeneities in the gravitational potential that later seeded the formation of structure like galaxies and clusters of galaxies. The evolution of the universe during inflation is thus expected to leave an imprint on the statistics of the temperature perturbations we observe in the cosmic microwave background (CMB) and the large-scale distribution of matter present in the universe (LSS).

Inflation predicts the distribution of these perturbations to be nearly Gaussian, with small deviations from "gaussianity" that contain information about the nature of the physics driving it. It is the study of the generation of these deviations from gaussianity and their observation that is called "non-gaussianity" in the context of cosmology. The purpose of this thesis is not to provide a thorough introduction to this important topic (though some of that can be found in this introduction) but rather to present the original developments regarding this subject in which I participated during these last few years.

This thesis is organized as follows: In this first chapter we will shortly describe

how inflation generates *nearly* gaussian perturbations, and how deviations from gaussianity can be observed. We will see that it is non-linear physics that provides the source for the non-gaussianity; one therefore expects the non-linear nature of gravity to change the primordial non-gaussian signal as the universe evolves. In chapter 2 we will start studying this by computing the second-order metric during the matter dominated epoch, for this task we develop a novel method in which the cosmological fluid is treated as a scalar field. In chapter 3 we take another step in the study of the contribution of non-linear gravity to the non-gaussianity observed in the temperature fluctuations of the CMB, by actually computing this contribution in a simplified but physical case. In chapter 4 we turn to a slightly different topic, there we will study the modification that non-gaussianity induces in the number density of massive dark matter haloes, specifically the mass function. We will improve upon previous computations of this effect resulting in what we believe is one of the most rigorous analytical estimates for the non-gaussian mass function. We draw conclusions in chapter 5.

In this introduction we will start by briefly describing the standard cosmological model and the framework of inflation in section 1.2; this will be useful in order to present some terminology and notation. We will then turn to the quantum generation of fluctuations in section 1.3. The deviations from non-gaussianity induced on these fluctuations in several models of inflation are studied in section 1.4. Current constraints on the magnitude of this non-gaussianity come mainly from the CMB, which we will study in section 1.5, and the large scale structure, which we will study in section 1.6. None of this material is original, and I took most of it from references [4; 5; 6; 7]. This should provide a good background for the rest of this thesis. For a more detailed study of each of these topics we refer the reader to the references cited above.

1.2 The standard cosmological model.

In this section I will very shortly describe what has become the standard description of our universe and briefly mention some of the evidence supporting it. We will use this to motivate the introduction of the inflationary hypothesis, which motivates the study of non-gaussianity. The main purpose of this section is actually to fix some notation and terminology that will be used throughout the following chapters. A more detailed introduction to these topics can be found in standard textbooks on cosmology.

Observations of the distribution of matter on scales larger than about 100 Mpc show that it is nearly homogeneous, and the observation of a nearly isotropic microwave background (called cosmic microwave radiation background or CMB) with a temperature of 2.74 K strongly suggests that the universe itself is isotropic. Of course, homogeneity and isotropy are not invariant under Lorentz transformations, so to be precise we assume the existence of a set of observers for which the universe looks homogeneous and isotropic. Under these assumptions, the only metric compatible with the symmetries is the Friedmann-Lemaître-Robertson-Walker (FLRW) metric

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2(t)\left(\frac{\mathrm{d}r^2}{1-kr^2} + r^2\mathrm{d}\Omega^2\right), \tag{1.1}$$

where $a(t)$ is a time-dependent function, often called the scale factor, to be solved for by using Einstein's equation; the curvature constant k can take the values $-1, 0, 1$ depending on the geometry of the space-time (-1 for a closed universe, 0 for a flat universe, and 1 for an open universe); and we define $\mathrm{d}\Omega^2 \equiv \mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}\varphi^2$. Note that the full time dependence of the spatial part of the metric is encoded in the scale factor, for this reason the spatial coordinates in which the FLRW metric takes this form are often called comoving coordinates. The time coordinate t appearing in relation (1.1) is often called cosmic time.

The only form that the stress-energy tensor can take which is compatible with the symmetries of the FLRW metric can be written in the form of a perfect fluid

$$T^{\mu}{}_{\nu} = (\rho + p)u^{\mu}u_{\nu} + p\delta^{\mu}{}_{\nu}\,, \tag{1.2}$$

where p and ρ are the pressure and energy densities of the fluid respectively, and u^{μ} is a timelike vector representing the fluid 4-velocity. One can also formulate an equation of state that relates p to ρ. A parametrization that describes a wide range of cases is $p = w\rho$ where w is often called the equation of state parameter. A perfect fluid composed of non-relativistic particles (for example dark matter) has $w = 0$, and a perfect fluid composed of composed of highly relativistic particles (for example photons) has $w = 1/3$.

The observation of shifts towards the red in the frequencies of spectral lines from distant galaxies, along with increasingly precise measurements of their distances, provide overwhelming evidence for the expansion of the universe, $\dot{a}(t) > 0$. The precise time dependence of the scale factor can be found by solving Einstein's equation. By using the metric given by equation (1.1) and the stress-energy tensor given by relation (1.2) in Einstein's equation, one obtains Friedmann's equations:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2}\,, \tag{1.3}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p)\,. \tag{1.4}$$

From equation (1.4) one can see that for a universe filled with matter that satisfies the weak energy condition (which for a stress-energy tensor of the form (1.2) is written as $w \geq -1/3$) the expansion slows down, $\ddot{a} < 0$. However, observations of distant supernovae [8; 9] have shown the universe to be accelerating $\ddot{a} > 0$. This indicates the presence of some component that violates the weak energy condition, often termed dark energy. The simplest explanation of this phenomenon, and one that is fully favored by current data, is the introduction of a cosmological constant, that is the addition of a term of the form $\Lambda g^{\mu\nu}/8\pi G$ to the stress energy-tensor that can be interpreted as having an equation of state parameter $w = -1$. While the value that Λ must have in order to account for the present acceleration of the universe remains unexplained by current physics, we will not dwell here on this interesting topic since it would take us too far off-track, and simply refer the reader to reference [10].

Let us now define the Hubble parameter $H \equiv \dot{a}/a$, the density ratios $\Omega_i \equiv 3\rho_i/(8\pi G H^2)$ where i stands for each component of the cosmic fluid (radiation r, matter m, dark matter dm, baryons b, etc)and $\Omega_k \equiv \frac{k}{a^2 H^2}$. Using these definitions one can rewrite equation (1.3) as

$$H(a) = H_0 \left[\frac{\Omega_{m\,0}}{a^3} + \frac{\Omega_{r\,0}}{a^4} + \frac{\Omega_{k\,0}}{a^2} + \Omega_\Lambda \right]^{1/2}, \tag{1.5}$$

where a subindex 0 indicates the value a quantity takes at the present time. Thus, a measurement of the energy density of the universe is a direct measurement of its geometry. A combination of measurements of the temperature fluctuations of the CMB, distances to type Ia supernovae and baryon acoustic oscillations yield Ω_k to be zero to one part in 10^{-2} at the present time; we will therefore often assume implicitly that the universe is flat ($k = 0$). These measurements also give (see reference [11])

$$\Omega_\Lambda = 0.722 \pm 0.015\,, \qquad \Omega_b = 0.0461 \pm 0.0015\,, \qquad \Omega_{dm} = 0.232 \pm 0.013\,. \tag{1.6}$$

To summarize, the universe is believed to be well described by the metric (1.1) with $k = 0$, a scale factor that is growing in time as given by equations (1.3) and (1.4), and composed today mainly of dark energy in the form of a cosmological constant, cold dark matter, and "baryons", with the percentages as given by (1.6). This model is often called ΛCDM.

Note also that even though today dark energy is the dominant component of the universe, going back in time dark matter and then radiation become dominant. The CMB was emitted during the matter dominated era, shortly after the moment when atomic nuclei and electrons combined to form neutral atoms, and the density of free electrons became so low that the mean free path of the photons became of the order of the size of the horizon. Going even further back in time, the energy density of the universe was expected to be so high that even atomic nuclei where dissociated. One can compute the abundance of light elements produced from the cooling of this plasma composed initially of free electrons, protons, neutrons, photons and neutrinos, to find that it coincides extremely well with observations. This primordial nucleosynthesis is considered one of the key observational proofs of the standard cosmological model, for more information on this we refer the reader to reference [12].

1.2.1 Inflation

We have described a model of the universe that is supported by firm observational evidence. If this description of the universe is naïvely extrapolated to the extreme, the scale factor is expected to be zero at some finite time in the past, $a = 0$ at $t = 0$. Thus $t = 0$ is sometimes called "the origin of the universe". We will call this picture "standard big bang cosmology". However, this naïve extrapolation leaves some open questions that are solved by the introduction of an additional simple hypothesis, namely inflation. In order to formulate these questions let us first introduce some additional

notation. The metric given by relation (1.1) can be rewritten as

$$\mathrm{d}s^2 = a^2(\tau)\Big(-\mathrm{d}\tau^2 + \frac{\mathrm{d}r^2}{1-kr^2} + r^2\mathrm{d}\Omega^2\Big)\,, \tag{1.7}$$

where τ is called conformal time and it is related to cosmic time by

$$\tau \equiv \int_0^t \frac{\mathrm{d}t'}{a(t')} = \int_0^a \mathrm{d}\log a' \Big(\frac{1}{a'H(a')}\Big)\,. \tag{1.8}$$

Note that this integral is also the maximum distance that signals can travel since the origin of the universe $t = 0$ at a given time t, *i.e.* the particle horizon of the FLRW metric.

During the matter dominated era the particle horizon goes like $\tau \propto a$, and during radiation domination $\tau \propto a^{1/2}$, so the particle horizon decreases when going backwards in time. This means that in standard big bang cosmology the integral of equation (1.8) converges and the universe at a given time is made of several patches which were never in causal contact. The size of this patches at the CMB is of order ~ 1 square degree in the sky, so the CMB is made of hundreds of patches causally disconnected from each other. However, the CMB has nearly the same temperature in every direction, a fact that is hard to explain in the light of what we've just discussed. This is often called the horizon problem.

Another related problem of the standard big bang cosmology can be easily formulated after rewriting the Friedmann equation (1.3) as

$$\Omega - 1 = \frac{k}{H^2a^2}\,, \tag{1.9}$$

where $\Omega = \Omega_\Lambda + \Omega_m + \Omega_r$. During radiation and matter domination the quantity aH decreases with time. From equation (1.9) one finds that in order for the universe to be flat today ($\Omega = 1$) to one part in 10^{-2}, it must be flat to one part in 10^{-16} at the time of the big bang nucleosynthesis, and to one part in 10^{-55} when the energy density of the universe was at the GUT scale. This large fine-tuning is often called the flatness problem.

An elegant solution to both the flatness problem and the horizon problem is to include the additional hypothesis that before radiation domination there was an epoch when the "comoving horizon" $(aH)^{-1}$ decreased with time

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big(\frac{1}{aH}\Big) < 0\,. \tag{1.10}$$

This hypothesis is equivalent to $\ddot{a} > 0$, and for this reason it is called "inflation". By making use of equations (1.3) and (1.4), another way to rewrite this condition is

$$\varepsilon \equiv -\frac{\dot{H}}{H} < 1\,. \tag{1.11}$$

If the epoch of inflation lasted long enough, an initial curvature term in equation (1.9) can be rendered small enough to achieve the required fine-tuning needed to solve the flatness problem. One can also require that inflation last long enough to give a contribution to the integral in equation (1.8) which is arbitrarily large and thus solve the horizon problem. Notice also that due to this early phase of accelerated expansion, it is not necessarily true that $a = 0$ at some time in the past. Inflation was first postulated as a solution to the flatness and horizon problems in reference [13].

Both the flatness and horizon problems are merely related to initial conditions; if the universe started exactly flat and homogeneous, but with the necessary small inhomogeneites to seed structure formation, there would be no need for inflation. However, the inflationary hypothesis is a dynamical solution to these problems that, as we will see in section 1.3, also provides a physical explanation for the generation of small perturbations seen in the CMB and which later grow gravitationally to form the large scale structure.

A simple way to realize inflation is to postulate the existence of a scalar field ϕ minimally coupled to gravity with an action

$$S = \int \mathrm{d}^4x \, \sqrt{-g}\Big(-\frac{1}{2}\partial_\mu\phi\partial^\mu\phi - V(\phi)\Big) \,. \tag{1.12}$$

In an FLRW metric, the scalar field must depend only on time, and its energy-momentum tensor can be computed from the action, giving

$$\rho = \frac{1}{2}\dot{\phi}^2 + V(\phi) \,, \tag{1.13}$$

$$p = \frac{1}{2}\dot{\phi}^2 - V(\phi) \,. \tag{1.14}$$

Inflation can then be realized for a universe dominated by such a scalar field if its potential dominates over its kinetic energy in such a way that the equation of state parameter for the scalar field satisfies $w < -1/3$ thus making $\ddot{a} > 0$ according to equation (1.4). The dynamics of this scalar field are determined by its equation of motion and the Friedmann equation

$$\ddot{\phi} + 3H\dot{\phi} + V_{,\phi}(\phi) = 0 \,, \qquad H^2 = \frac{8\pi G}{3}\Big(\frac{1}{2}\dot{\phi}^2 + V(\phi)\Big) \,. \tag{1.15}$$

Using equation (1.4) with equations (1.13) and (1.14), one can write the condition for inflation, equation (1.11), as

$$\varepsilon = 4\pi G\frac{\dot{\phi}^2}{H^2} < 1 \,. \tag{1.16}$$

In order for the first derivative of the field to remain small for a large period of time such that inflation lasts long enough to solve the horizon and flatnees problems, one

also requires its second derivative to be small when compared to the "Hubble friction" term $3H\dot{\phi}$ in equation (1.15), that is

$$|\eta| < 1, \qquad \text{with} \quad \eta \equiv -\frac{\ddot{\phi}}{H\dot{\phi}}. \tag{1.17}$$

Models of inflation for which $\epsilon, \eta \ll 1$ are called "slow-roll inflation". For $\epsilon \approx 0$, the scale factor is $a(t) \approx e^{Ht}$, so in slow roll inflation the space-time will be close to de Sitter, and equations (1.15) give

$$\dot{\phi} \approx -\frac{V_{,\phi}}{3H}, \qquad H^2 \approx \frac{8\pi G}{3} V(\phi) \approx \text{const}. \tag{1.18}$$

In slow-roll models, the period of inflation will come to an end when the potential no longer satisfies the slow-roll conditions. At this point the scalar field oscillates coherently and is expected to decay into other fields that populate the universe. This process is called reheating and we choose not to dwell on it here, the interested reader can take a look at [14].

One often measures how long infl ation lasts according to a given model by its "number of e-folds"

$$N \equiv \ln \frac{a_{end}}{a_{begin}} = \int_{t_{begin}}^{t_{end}} H \mathrm{d}t \approx 8\pi G \int_{\phi_{end}}^{\phi_{begin}} \frac{V}{V_{,\phi}} \mathrm{d}\phi, \tag{1.19}$$

where we have used equations (1.18), and the labels *begin* and *end* denote the beginning and end of inflation respectively. A necessary condition to solve the horizon problem is that the comoving horizon at the start of inflation be larger than the present comoving horizon. This can be stated as follows

$$\frac{a_{end}}{a_0 H_0} \frac{a_{begin}}{a_{end}} \lesssim \frac{1}{H_I}, \tag{1.20}$$

where H_I is the value of the Hubble parameter during inflation, which we assume to be nearly constant. Using the definition of N, this implies a bound on the number of e-folds that inflation must last

$$N \gtrsim \ln\left(\frac{T_0}{H_0}\right) - \ln\left(\frac{T_{end}}{H_I}\right) \approx 67 - \ln\left(\frac{T_{end}}{H_I}\right), \tag{1.21}$$

where we used the fact that the temperature of radiation goes like $T \propto 1/a$. This means that inflation must last for roughly 60 e-folds in order to solve the horizon problem.

1.3 Quantum generation of fluctuations during inflation.

In the previous section we briefly described how inflation solves the flatness and horizon problems of standard big bang cosmology. However, the most compelling feature

of inflation is actually that it provides a physical mechanism for the generation of perturbations in the gravitational potential that can be observed in the CMB, and which grow gravitationally to form the large scale structure of the universe. In this section we describe this process following the treatment of [4; 7].

The calculation will be carried out in several steps, first we perturb the FLRW metric and the stress-energy tensor, then we compute the action for the classical perturbed field, we then turn to the quantization of the field, and finally we present the primordial power spectrum generated by inflation.

1.3.1 The perturbed FLRW metric.

The generic perturbed flat FLRW metric can be written as

$$\mathrm{d}s^2 = -(1+2\Phi)\mathrm{d}t^2 + a(t)\omega_i \mathrm{d}t\,\mathrm{d}x^i + a(t)^2\left[(1-2\Psi)\delta_{ij} + \chi_{ij}\right]\mathrm{d}x^i\mathrm{d}x^j\,. \tag{1.22}$$

One can furthermore decompose ω_i into a scalar and a pure vector parts[1]

$$\omega_i = \partial_i\omega + \omega_i^{\perp}\,, \qquad \text{with} \quad \partial^i\omega_i^{\perp} = 0\,, \tag{1.23}$$

and also decompose the traceless tensor χ_{ij} as[2]

$$\chi_{ij} = \partial_i\partial_j\chi + 2\partial_{(i}\chi_{j)} + \gamma_{ij}\,, \qquad \text{with} \quad \partial^i\chi_i = \gamma^i{}_i = \partial^i\gamma_{ij} = 0\,. \tag{1.24}$$

Similarly, we perturb the stress-energy tensor (bars denote unperturbed quantities)

$$T^{\mu}{}_{\nu} = (\bar{\rho}+\delta\rho)u^{\mu}u_{\nu} + (\bar{p}+\delta p)(u^{\mu}u_{\nu} + \delta^{\mu}{}_{\nu}) + \Sigma^{\mu}{}_{\nu}\,, \tag{1.25}$$

where now u^{μ} are perturbed velocities, δp and $\delta\rho$ are the pressure and energy density perturbations, and $\Sigma^{\mu}{}_{\nu}$ is the anisotropic stress which satisfies $\Sigma^{\mu}{}_{\mu} = \Sigma^{\mu}{}_{\nu}u^{\nu} = 0$. It will prove useful to define $T_{0i} \equiv (\bar{\rho}+\bar{p})q_i$, where q_i is sometimes called the momentum transfer.

We also split the scalar field into background and perturbations

$$\phi(x) \equiv \bar{\phi}(t) + \delta\phi(x)\,. \tag{1.26}$$

Computing the stress energy tensor for this scalar field with an action given by equation (1.12) gives

$$T^{\mu}{}_{\nu} = \partial^{\mu}\phi\partial_{\nu}\phi - \delta^{\mu}{}_{\nu}\left(\frac{1}{2}\partial^{\alpha}\phi\partial_{\alpha}\phi + V(\phi)\right). \tag{1.27}$$

[1] This decomposition has to do with the transformation properties of each of these quantities under three-dimensional rotations. Indeed, under rotations around the Fourier mode vector $\mathbf{k}$ scalars have helicity 0, vectors have helicity ±1, and tensors have helicity ±2. Furthermore, in linear perturbation theory one can show that degrees of freedom with different helicity do not mix, see reference [7] for a clear proof. This of course will no longer hold at the non-linear level, see equation (2.66).

[2] Throughout this thesis we use the notation $T_{(\mu\nu)} \equiv \frac{1}{2}(T_{\mu\nu} + T_{\nu\mu})$, and $T_{[\mu\nu]} \equiv \frac{1}{2}(T_{\mu\nu} - T_{\nu\mu})$ for an arbitrary tensor T.

By doing the following identifications one can write it in the form of equation (1.25)

$$u_\mu = \frac{\partial_\mu \phi}{\sqrt{-\partial^\alpha \phi \partial_\alpha \phi}}\,, \tag{1.28}$$

$$\bar{\rho} + \delta\rho = -\frac{1}{2}\partial_\mu\phi\partial^\mu\phi + V(\phi)\,, \tag{1.29}$$

$$\bar{p} + \delta p = -\frac{1}{2}\partial_\mu\phi\partial^\mu\phi - V(\phi)\,. \tag{1.30}$$

Thus, for a scalar field the momentum transfer is simply $q_i = \partial_i \delta\phi$ as seen from equation (1.27).

In the perturbed metric, given by equation (1.22), there are ten degrees of freedom: the four scalars Ψ, Φ, ω and χ, two transverse vectors $\omega_i^\perp$ and χ_i (having two degrees of freedom each), and one transverse and traceless tensor γ_{ij} (with two degrees of freedom). If we add a scalar field (the inflation) this is another scalar degree of freedom of the theory. Four of these degrees of freedom, two scalars and one transverse vector, are related to the freedom in choosing coordinates $x^\mu \to x^\mu + \xi^\mu(x)$. A specific choice of coordinates is called a gauge. Furthermore, in general relativity four of Einstein's equations can be interpreted as being constraint equations that fix one scalar and one transvere vector degrees of freedom (see equations (1.37) and (1.36) below). This leaves us only with three dynamical degrees of freedom: the two dynamical degrees of freedom contained in the transverse traceless tensor γ_{ij} and one dynamical scalar degree of freedom which comes from the introduction of the inflaton.

In this thesis we will make use of two gauges: the Poisson gauge and the comoving gauge. These gauges are defined by setting to zero four degrees of freedom (two scalar and one transverse vector) through a coordinate choice. In the Poisson gauge one fixes to zero the two scalar degrees of freedom χ and ω, and the transverse vector χ_i. Thus, the metric in Poisson gauge takes the form

$$\mathrm{d}s^2 = -(1+2\Phi)\mathrm{d}t^2 + a(t)\omega_i^\perp \mathrm{d}t\,\mathrm{d}x^i + a(t)^2\big[(1-2\Psi)\delta_{ij} + \gamma_{ij}\big]\mathrm{d}x^i\mathrm{d}x^j\,. \tag{1.31}$$

The potentials Φ and Ψ in Poisson gauge at linear order are often called Bardeen potentials (see reference [15]).

The comoving gauge is defined by setting to zero the scalar part of the momentum transfer q, the scalar χ, and the transverse vector χ_i. For a universe dominated by a scalar field, this is equivalent to setting the equal time hypersurfaces to coincide with the constant field hypersurfaces, *i.e.* set the perturbations of the field to zero. The remaining scalar degree of freedom will then be a scalar component of the metric. The relations connecting the Poisson gauge to the comoving gauge can be found in section 2.5. For a full discussion of several useful gauges and gauge transformations at linear order, we refer the reader to reference [16].

Another gauge which will be of some use in the following discussions is the uniform density gauge, in which we fix the scalar χ and the transverse vector χ_i to zero, and also require the constant time hypersurfaces to coincide with the constant density

hypersurfaces. In this gauge we will use the non-standard notation $\Psi \equiv \zeta_{cd}$, where cd stands for constant density[1].

1.3.2 Classical dynamics of the perturbed field.

In order to quantize the field, we must first write the perturbed action which encodes its dynamics, for this we will follow the treatment of reference [4]. We will achieve this by using the ADM splitting of the metric (see reference [17])

$$ds^2 = -N^2 dt^2 + h_{ij}(dx^i + N^i dt)(dx^j + N^j dt) \,. \tag{1.32}$$

Here h_{ij} is the metric of constant time hypersurfaces, while N and N^i are called the lapse and the shift respectively. The action (1.12) becomes

$$S = \frac{1}{2}\int \mathrm{d}^4x\sqrt{h}\Big[\frac{1}{M_{pl}^2}NR^{(3)} + \frac{1}{M_{pl}^2}N^{-1}\left(E_{ij}E^{ij} - E^2\right) - 2NV + N^{-1}\left(\dot{\phi} - N^i\partial_i\phi\right)^2 - Nh^{ij}\partial_i\phi\partial_j\phi\Big] \,, \tag{1.33}$$

where the Planck mass is $M_{pl}^2 \equiv 1/8\pi G$, $R^{(3)}$ is the curvature scalar computed from h_{ij}, $E \equiv E^i_{\ i}$ and

$$E_{ij} \equiv \frac{1}{2}\left(\dot{h}_{ij} - \nabla_i N_j - \nabla_j N_i\right) \,. \tag{1.34}$$

The covariant derivatives ∇_i are with respect to the 3-metric h_{ij} and all roman indeces $i, j, \ldots$ are raised and lowered with this metric. The first line in equation (1.33) is simply the Einstein-Hilbert action for the metric (1.32) and the second line is the scalar field action. In this section we will work in comoving gauge following reference [4]. The gauge conditions for the comoving gauge are written in this language as

$$\begin{gathered}\delta\phi = 0, \qquad h_{ij} = a^2 e^{2\zeta}\hat{h}_{ij} \,, \qquad \hat{h}_{ij} = \delta_{ij} + \gamma_{ij} + \tfrac{1}{2}\gamma_{il}\gamma_{lj} + \ldots \,, \\ \det\hat{h} = 1 \,, \qquad \gamma_{ii} = 0 \,, \qquad \partial_i\gamma_{ij} = 0 \,,\end{gathered} \tag{1.35}$$

where ζ describes the dynamical scalar degree of freedom and γ the tensor ones[2].

As the action does not contain time derivatives of N and N_i, these variables act as Lagrange multipliers, i.e. their equations of motion are algebraic constraints. These equations are the momentum and Hamiltonian constraints of Einstein equations:

$$\nabla_i\left[N^{-1}\left(E^i_j - \delta^i_j E\right)\right] = 0 \,, \tag{1.36}$$

[1] This strange notation is due to the fact that we reserve ζ for the curvature perturbation in the comoving gauge.

[2] We follow the notation of [4] using ζ to denote the curvature perturbation in the comoving gauge. Often this quantity is instead called $\mathcal{R}$ reserving ζ to denote the curvature perturbation in the uniform density gauge.

$$R^{(3)} - 2M_{pl}^2 V - M_{pl}^2 N^{-2}\dot{\phi}^2 - \frac{1}{N^2}\left(E_{ij}E^{ij} - E^2\right) = 0\,. \tag{1.37}$$

For now we are interested only in the leading order perturbation theory results, so we will work here at linear order (which corresponds to quadratic order in ζ and γ in the action). The first step is to solve the linear constrain equations, it turns out that the quadratic-order solutions will not be necessary since they will appear in the second-order action multiplying the background equations. Separating the shifts into a scalar and a transverse vector part $N^i \equiv \partial_i\psi + N_T^i$, with $\partial_i N_T^i = 0$, and writing $N = 1 + N_1$, one obtains

$$N_1 = \frac{\dot{\zeta}}{H}\,, \qquad N_T^i = 0\,, \qquad \psi = -\frac{\zeta}{Ha^2} + \frac{\dot{\phi}^2}{2M_{pl}^2 H^2}\partial^{-2}\dot{\zeta}\,, \tag{1.38}$$

where we defined $\partial^{-2}\partial_i\partial^i \equiv 1$. Using these solutions in the action, equation (1.33), and after several integrations by parts, one obtains the leading order action for the scalar degree of freedom ζ

$$S = \frac{1}{2}\int \mathrm{d}^4x\; a^3\frac{\dot{\phi}^2}{H^2}\left(\dot{\zeta}^2 - a^{-2}\partial_i\zeta\partial^i\zeta\right). \tag{1.39}$$

The equation of motion for ζ can be derived from the action to give

$$\ddot{\zeta} + \frac{\partial}{\partial t}\ln\left(a^3\frac{\dot{\phi}^2}{H^2}\right)\dot{\zeta} - \frac{\nabla^2}{a^2}\zeta = 0\,. \tag{1.40}$$

For the sake of clarity we will also limit our calculations to leading order in the slow-roll parameters defined in equations (1.16) and (1.17). By going to conformal time, and changing variables from ζ to the Sasaki-Mukhanov variable $v \equiv a(\dot{\phi}/H)\zeta$, one can rewrite the equation of motion at linear order in perturbations and zeroth order in slow roll as

$$v'' - \frac{a''}{a}v - \nabla^2 v = 0\,, \tag{1.41}$$

where primes denote derivatives with respect to conformal time. The solution to this equation in Fourier space is simply given by

$$v_{\mathbf{k}}(\tau) = C_1\frac{e^{-ik\tau}}{\sqrt{2k}}\left(1 - \frac{i}{k\tau}\right) + C_2\frac{e^{ik\tau}}{\sqrt{2k}}\left(1 + \frac{i}{k\tau}\right). \tag{1.42}$$

Here, C_1 and C_2 are integration variables to be fixed by initial conditions. We will see in the next subsection that at the quantum level these are fixed by the choice of vacuum.

Similarly, one can derive the action for the tensor degrees of freedom (see section 2.4.2 for more details)

$$S = \frac{M_{pl}^2}{8}\int \mathrm{d}^4x\; a^3\left(\dot{\gamma}_{ij}\dot{\gamma}^{ij} - a^{-2}\partial_k\gamma_{ij}\partial^k\gamma^{ij}\right), \tag{1.43}$$

and the equation of motion for each component is simply the equation of motion for a scalar field in an FLRW background

$$\ddot{\gamma}_{ij} + 3H\dot{\gamma}_{ij} - \frac{\nabla^2}{a^2}\gamma_{ij} = 0\,. \tag{1.44}$$

Before turning to the quantization of these fields, let us motivate the use of the comoving gauge. The advantage of this gauge is that ζ can be shown to be constant outside of the horizon for a universe filled with an adiabatic fluid. This means that, given a Fourier mode for ζ with wave vector k, once $k \ll aH$ its amplitude will remain constant independent of the evolution of the universe. Thus, once we compute the amplitude of the perturbations when they exit the horizon, they will remain unaffected by the potentially unknown evolution of the universe between the end of inflation and the moment when they reenter the horizon.

In order to show that ζ is conserved outside of the horizon we will take the approach of reference [18], and while we will be somewhat imprecise here due to lack of space, we point the reader to this reference for further details. The starting point is the ADM form of the metric, equation (1.32).

The first step is to perform a gradient expansion. That is, we expand perturbatively in the quantity k/aH, meaning that we will focus on the case in which all the Fourier modes are outside of the horizon. Another key assumption is that in the limit of very large scales $k/aH \to 0$ the universe becomes an FLRW spacetime locally. This implies that we can assume the shifts N^i to be at least of order k/aH if we are to recover the FLRW metric in the form of equation (1.1). If we split the metric of constant time hypersurfaces h_{ij} as

$$h_{ij} = a^2(t)e^{2\alpha}(\delta_{ij} + \gamma_{ij})\,, \tag{1.45}$$

we can further assume that $\gamma_{ij} = \mathcal{O}(k/aH)$. No assumptions can be made for the lapse N or the quantity α since they are not locally measurable.

Next consider a universe filled with a perfect fluid (*i.e.* $\Sigma^i{}_j = 0$ in equation (1.25) at least to order $\mathcal{O}(k^2/a^2H^2)$), and choose coordinates that are comoving with the fluid $u^i = 0$ (this fixes three out of the four degrees of freedom that we can fix with a gauge choice). In these coordinates the 4-velocity is given by

$$u^\mu = (1/N, 0) + \mathcal{O}(k^2/a^2H^2)\,, \tag{1.46}$$

from which we can compute

$$\theta \equiv \nabla_\mu u^\mu = \frac{3(\dot{\alpha} + H)}{N} + \mathcal{O}(k^2/a^2H^2)\,. \tag{1.47}$$

We can now write the energy conservation equation using the perfect fluid form for the stress-energy tensor

$$-u_\mu \nabla_\nu T^{\mu\nu} = \frac{\mathrm{d}}{\mathrm{d}\tau}\rho + (\rho + p)\theta = 0\,, \tag{1.48}$$

which, using equation (1.47), becomes

$$\dot{\rho} + 3(\rho + p)(\dot{\alpha} + H) = \mathcal{O}(k^2/a^2H^2)\,. \tag{1.49}$$

Now choose coordinates such that constant time hypersurfaces coincide with constant ρ hypersurfaces (fixing the one remaining degree of freedom that we can fix with a gauge choice). We argue that if p is a unique function of ρ, at least to first order in the gradient expansion, then both ρ and p will be unperturbed in equation (1.49) and using the unperturbed energy conservation equation gives

$$\dot{\alpha} = \mathcal{O}(k^2/a^2H^2)\,. \tag{1.50}$$

To complete the proof we must connect this coordinate choice in which $u^i = 0$ and $\rho(x) = \bar{\rho}(t)$ (which we will call the pseudo uniform density gauge) with the comoving gauge. It is easy to convince oneself that to leading order in k/aH both gauges coincide under the assumption that $N^i = \mathcal{O}(k/aH)$ and $\gamma_{ij} = \mathcal{O}(k/aH)$ since the metric in the pseudo uniform density gauge has the large scale form

$$\mathrm{d}s^2 = -N^2\mathrm{d}t^2 + a^2(t)e^{2\alpha}\mathrm{d}x^2\,, \tag{1.51}$$

which has the same form as the comoving gauge metric at large scales with $\zeta = \alpha$. One can also check that $T_{0i} = \mathcal{O}(k/aH)$ in the pseudo uniform density gauge. This means that the gauge in which $u^i = 0$ and $\rho(x) = \bar{\rho}(t)$ satisfies trivially the conditions that define the comoving gauge, equation (1.35), at large scales [1]. Equation (1.50) then implies

$$\dot{\zeta} = \mathcal{O}(k/aH)\,. \tag{1.52}$$

Notice that we never expanded in perturbations, so this result is expected to hold nonlinearly when all the modes are outside of the horizon. However, modes inside of the horizon can change the evolution of ζ, as will be discussed in section 2.6.

1.3.3 Quantization of a scalar field in an expanding background.

In this subsection we follow the treatment of reference [19] and perform the computation to leading order in slow-roll. The starting point is the equation of motion for the Sasaki-Mukhanov variable, equation (1.41), which can be derived from an action of the form

$$S = \frac{1}{2}\int \mathrm{d}^4x \left(v'^2 - \partial_i v \partial^i v + \frac{a''}{a}v^2\right)\,. \tag{1.53}$$

The conjugate momentum for v is then $\pi = v'$ and the Hamiltonian is

$$H = \frac{1}{2}\int \mathrm{d}^4x \left(v'^2 + \partial_i v \partial^i v - \frac{a''}{a}v^2\right)\,. \tag{1.54}$$

In Fourier space the equation of motion for each mode v_k has the form of an equation of motion for a harmonic oscillator with a time dependent frequency

$$v_k'' + \omega_k^2(\tau)v_k = 0\,, \qquad \omega_k(\tau) \equiv \sqrt{k^2 - a''/a}\,. \tag{1.55}$$

[1] A similar argument shows that at superhorizon scales the curvature perturbation in comoving gauge and the curvature perturbation in constant density gauge both coincide, $\zeta = \zeta_{cd} + \mathcal{O}(k/aH)$.

By analogy with the harmonic oscillator, we interpret the field v as an operator, and decompose it in creation and annihilation operators

$$v(\mathbf{x},\tau) = \int \frac{\mathrm{d}^3k}{(2\pi)^3}\frac{1}{\sqrt{2}}\left(e^{i\mathbf{k}\cdot\mathbf{x}}f_k^*(\tau)a_\mathbf{k} + e^{-i\mathbf{k}\cdot\mathbf{x}}f_k(\tau)a_\mathbf{k}^\dagger\right), \tag{1.56}$$

where $f(\tau)$ is a complex solution of the equation of motion (1.55). In order for the field and its conjugate momentum to satisfy the canonical commutation relation

$$[v(\mathbf{x},\tau),\pi(\mathbf{y},\tau)] = i\delta(\mathbf{x}-\mathbf{y}), \tag{1.57}$$

one imposes on the creation and annihilation operators the usual relations

$$[a_\mathbf{k},a_{\mathbf{k}'}^\dagger] = \delta(\mathbf{k}-\mathbf{k}'), \qquad [a_\mathbf{k},a_{\mathbf{k}'}] = [a_\mathbf{k}^\dagger,a_{\mathbf{k}'}^\dagger] = 0, \tag{1.58}$$

along with a precise condition for the normalization of the function $f(\tau)$[1]

$$\mathrm{Im}(f_k'(\tau)f_k^*(\tau)) = 1. \tag{1.59}$$

One defines the vacuum as the state that is annihilated by any annihilation operator, $a_\mathbf{k}|0\rangle \equiv 0$. As usual, one can then define particle states through the action of the creation operators on the vacuum,

$$|m_{\mathbf{k}_1},n_{\mathbf{k}_2},\dots\rangle \equiv \frac{1}{\sqrt{m!n!\dots}}\left[(a_{\mathbf{k}_1}^\dagger)^m(a_{\mathbf{k}_2}^\dagger)^n\dots\right]|0\rangle. \tag{1.60}$$

However, the definition of the vacuum, the creation and annihilation operators and the mode functions $f_k(\tau)$ are all related in the sense that a choice of one fixes all the others. Furthermore, the definition of the vacuum is not straightforward (or even well defined). The problem resides in the fact that the minimum of the energy will change with time for a harmonic oscillator with a time-dependent frequency. Suppose that at some fixed time τ_1 the energy is minimized by the vacuum $|_{\tau_1}0\rangle$ related to some creation operators $a_\mathbf{k}^\dagger$, and let us look for the associated mode functions by finding the minimum of the energy. The energy of the vacuum is given by the expectation value of the Hamiltonian which can be shown to be

$$\langle_{\tau_1}0|H(\tau_1)|_{\tau_1}0\rangle = \frac{1}{4}\delta^{(3)}(0)\int \mathrm{d}^3k\ \left(|f_k'|^2+\omega_k^2(\tau_1)|f_k|^2\right). \tag{1.61}$$

The minimum of this quantity as a function of f_k for each k can be shown to be located at

$$f_k(\tau_1) = \frac{1}{\sqrt{\omega_k(\tau_1)}}, \qquad f_k'(\tau_1) = i\sqrt{\omega_k(\tau_1)}, \tag{1.62}$$

and the energy is

$$\langle_{\tau_1}0|H(\tau_1)|_{\tau_1}0\rangle = \frac{1}{2}\delta^{(3)}(0)\int \mathrm{d}^3k\,\omega_k(\tau_1). \tag{1.63}$$

[1] One can impose a different normalization on $f(\tau)$, but this will also include additional factors in the commutation relations.

But the problem is that the Hamiltonian will change with time. The function that minimizes the energy at some other time τ_2 will of course be given by a linear combination of $f_k(\tau)$ and $f_k^*(\tau)$ since it must also be a solution of the equations of motion

$$g^*(\tau) \equiv \alpha_k f_k^*(\tau) + \beta_k f_k(\tau)\,, \tag{1.64}$$

where α_k and β_k are called Bogolyubov coefficients for which the normalization condition (1.59) implies $|\alpha_k|^2 - |\beta_k|^2 = 1$. The mode function in the new vacuum $g(\tau)$ will have different creation operators $b_{\mathbf{k}}^\dagger$ associated with it, which are fixed by the requirement that $v(\mathbf{x},\tau)$ be equal when expressed in any of the two basis of solutions

$$b_{\mathbf{k}} = \alpha_{\mathbf{k}} a_k + \beta_k^* a_{-\mathbf{k}}^\dagger\,, \qquad b_{\mathbf{k}}^\dagger = \alpha_{\mathbf{k}}^* a_k^\dagger + \beta_k a_{-\mathbf{k}}\,. \tag{1.65}$$

Using these relations one can show that the energy at a time τ_2 of the vacuum $|_{\tau_1}0\rangle$ will be

$$\langle_{\tau_1}0|H(\tau_2)|_{\tau_1}0\rangle = \delta^{(3)}(0)\int \mathrm{d}^3k\;\omega_k(\tau_2)\left(\frac{1}{2} + |\beta_k|^2\right)\,. \tag{1.66}$$

which is different from the minimum energy at $\tau = \tau_2$. A further problem arises since, as seen from equation (1.55), there are modes $k^2 < a''/a$ for which $\omega_k^2 < 0$. For these modes one can show that the energy cannot even be minimized (see equation (1.61)).

The way out of these problems (which is not uncontested) is to define the vacuum by the physical requirement that it be equal to the vacuum of a Minkowski spacetime when the modes are much smaller than the horizon

$$f_k(\tau) \to \frac{1}{\sqrt{2k}} e^{-ik\tau}, \qquad \text{for} \quad k \gg aH\,. \tag{1.67}$$

The associated vacuum is called the Bunch-Davies vacuum. If inflation lasted long enough, all the modes that are now inside the horizon where much inside the horizon at the beginning of inflation, such that the condition $k \gg aH$ is satisfied and the Bunch-Davies vacuum for such modes is well defined. The choice of the Bunch-Davies vacuum as the physical vacuum during inflation fixes the coefficients of the solution to the equation of motion, given by equation (1.42), to be $C_1 = 1$ and $C_2 = 0$, that is

$$f_k(\tau) = \frac{e^{-ik\tau}}{\sqrt{2k}}\left(1 - \frac{i}{k\tau}\right)\,. \tag{1.68}$$

We will see in the following subsection that the interesting thing to compute is the two-point correlation function for $\zeta = a^{-1}(H/\dot{\phi})v$, which gives

$$\langle\zeta_{\mathbf{k}}(\tau)\zeta_{\mathbf{k}'}(\tau)\rangle = (2\pi)^3\delta(\mathbf{k}+\mathbf{k}')\frac{H^2}{a^2\dot{\phi}^2}|f_k(\tau)|^2\,. \tag{1.69}$$

During inflation the comoving horizon $(aH)^{-1}$ shrinks and eventually becomes much smaller than the characteristic length of a given mode k. In this limit, using the form of the mode functions in the Bunch-Davies vacuum one obtains

$$\langle\zeta_{\mathbf{k}}(\tau)\zeta_{\mathbf{k}'}(\tau)\rangle = (2\pi)^3\delta(\mathbf{k}+\mathbf{k}')\frac{H_*^4}{\dot{\phi}_*^2}\frac{1}{2k^3}\,, \tag{1.70}$$

where the subscript $*$ indicates the value that a quantity takes when a the mode k crosses the horizon $k \approx a_* H_*$. Each mode will cross the horizon at slightly different values of H and $\dot{\phi}$, thus slightly modifying the dependence on k of the two-point function. We will compute this explicitly in the following subsection. We saw that the value of ζ remains constant outside of the horizon until it reenters after the end of inflation. It is then possible to connect the two-point function of the CMB with this primordial two-point function.

One can perform the same steps for the tensor degrees of freedom, starting from their equation of motion (1.44). The resulting two point function outside of the horizon is given by

$$\langle \gamma^s_{\mathbf{k}}(\tau) \gamma^s_{\mathbf{k}'}(\tau) \rangle = (2\pi)^3 \delta(\mathbf{k}+\mathbf{k}') \frac{4H_*^2}{M_{pl}^2} \frac{1}{2k^3}\,, \tag{1.71}$$

where the index s indicates each of the two degrees of freedom (polarizations) of γ_{ij}. The absence of the factor $H^2/\dot{\phi}^2$ in this relation is due to the fact that the equation of motion for each component of γ is simply that of a free scalar field in an FLRW background.

1.3.4 The inflationary power spectrum.

In this subsection we will compute the power spectrum of fluctuations generated by inflation, and then derive some interesting consequences. We will follow the treatment of reference [7]. Let us start by defining the power spectrum $P_\zeta(k)$ as

$$\langle \zeta_{\mathbf{k}} \zeta_{\mathbf{k}'} \rangle \equiv (2\pi)^3 \delta(\mathbf{k}+\mathbf{k}') P_\zeta(k)\,. \tag{1.72}$$

Making use of equation (1.70) we can write the power spectrum approximately as[1]

$$P_\zeta(k) = \frac{2\pi^2 A_s}{k^3} \left(\frac{k}{k_*}\right)^{n_s - 1}\,, \tag{1.73}$$

where A_s is often called the amplitude of the fluctuations, which can be measured from the amplitude of the temperature fluctuations in the CMB and k_* is some reference scale. A power spectrum with the form $P_\zeta(k) \propto 1/k^3$ is termed scale-invariant, and the scalar spectral index n_s is a number close to one that encodes the deviation from this scale invariance. In order to compute $(n_s - 1)$ from (1.70) let us recall that $\mathrm{d}N = H \mathrm{d}t$, so the slow-roll parameters can be written as

$$\varepsilon \equiv -\frac{\dot{H}}{H^2} = -\frac{\mathrm{d}\ln H}{\mathrm{d}N}\,, \qquad \eta \equiv -\frac{\ddot{\phi}}{H\dot{\phi}} = \varepsilon - \frac{1}{2\varepsilon}\frac{\mathrm{d}\varepsilon}{\mathrm{d}N}\,, \tag{1.74}$$

[1] One can also go one step further and take into account the fact that the spectral index is expected to depend slightly on scale and write

$$P_\zeta(k) \equiv \frac{2\pi^2 A_s}{k^3} \left(\frac{k}{k_*}\right)^{n_s - 1 + \alpha_s \ln(k/k_*)}\,,$$

where α_s is called the running of the spectral index, defined as $\alpha_s \equiv \mathrm{d}n_s / \mathrm{d}\ln k$.

and we can use this to compute

$$n_s - 4 = \frac{\mathrm{d}\ln P_\zeta(k)}{\mathrm{d}\ln k} = \frac{\mathrm{d}\ln P_\zeta(k)}{\mathrm{d}N}\,\frac{\mathrm{d}N}{\mathrm{d}\ln k}\,. \tag{1.75}$$

Now we compute each piece of this equation. Making use of the form of the power spectrum given by equation (1.70), $P_\zeta(k) \propto H^2/k^3\varepsilon$, one obtains

$$\frac{\mathrm{d}\ln P_\zeta(k)}{\mathrm{d}N} = -3 + 2\frac{\mathrm{d}\ln H}{\mathrm{d}N} - \frac{\mathrm{d}\ln\varepsilon}{\mathrm{d}N} = -3 - 2\varepsilon - 2(\varepsilon - \eta)\,, \tag{1.76}$$

where we used the expressions (1.74). Similarly, recalling that the power spectrum is frozen at the moment of horizon crossing $k = aH$, one can compute

$$\frac{\mathrm{d}N}{\mathrm{d}\ln k} = \left(\frac{\mathrm{d}\ln k}{\mathrm{d}N}\right)^{-1} = (1-\varepsilon)^{-1}\,. \tag{1.77}$$

Putting together equations (1.75), (1.76) and (1.77), finally gives the leading order slow-roll result

$$n_s - 1 = 2\eta - 4\varepsilon\,. \tag{1.78}$$

Current limits are $n_s = 0.982^{+0.020}_{-0.019}$, see reference [11], which is close to one and thus compatible with what we have just computed for slow-roll inflation[1].

One can perform an analogous analysis for the tensor modes. Let us then define the power spectrum of tensor perturbations analogously to equation (1.73) as

$$\sum_s \langle \gamma^s_{\mathbf{k}}\gamma^s_{\mathbf{k}'}\rangle \equiv (2\pi)^3\delta(\mathbf{k}+\mathbf{k}')P_t(k)\,, \tag{1.79}$$

where we sum over the two polarizations. Taking into account the factor 2 coming from this sum and using equations (1.70) and (1.71) one obtains the following expression for the ratio of the tensor to scalar power spectra at leading order in slow-roll

$$r \equiv \frac{P_t(k)}{P_\zeta(k)} = 16\varepsilon\,. \tag{1.80}$$

This is a consistency relation that must be satisfied by slow-roll models of inflation. The ratio r also contains information on the energy scale of inflation since for a given power spectrum of scalar fluctuations (which has been measured at the CMB) it is proportional to $H^2 \sim V$, so

$$V^{1/4} \sim \left(\frac{r}{0.01}\right)^{1/4} 10^{16}\ \mathrm{GeV}\,, \tag{1.81}$$

current limits are $r < 0.36$ at the 95% confidence level (see reference [11]). This means that the energy scale of inflation in this type of models must be $V \lesssim 0.4 \times 10^{18}$ GeV. In analogy to equation (1.73) we write

$$P_t(k) = \frac{2\pi^2 A_t}{k^3}\left(\frac{k}{k_*}\right)^{n_t}\,, \tag{1.82}$$

[1] Here we cite the limits which include the tensor to scalar ratio r in the analysis.

where again A_t is the amplitude of the tensor perturbations, k_* is some reference scale, and n_t encodes the deviation from scale-invariance. One can compute n_t to leading order in slow roll obtaining

$$n_t = -2\varepsilon\,. \tag{1.83}$$

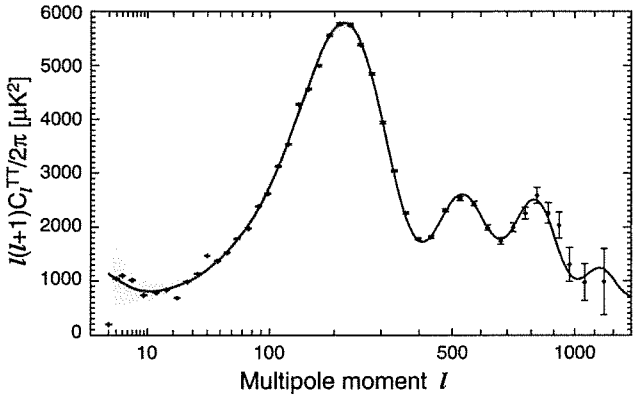


Figure 1.1: The CMB TT angular power spectrum as observed by the WMAP satellite, taken from reference [20].

Now we will very shortly describe how to connect the primordial power spectra to the angular power spectrum of temperature fluctuations in the CMB (figure 1.1). Since the sky is spherical, it is convenient to decompose the observed temperature fluctuations into spherical harmonics

$$\frac{\delta T(\hat{n})}{\bar{T}} = \sum_{\ell m} a_{\ell m} Y_{\ell m}(\hat{n})\,, \tag{1.84}$$

where $\bar{T}$ is the average temperature of the CMB, and $\delta T(\hat{n}) \equiv T(\hat{n}) - \bar{T}$ is the deviation with respect to this average value in a given direction of observation $\hat{n}$. Inverting this relation gives

$$a_{\ell m} = \int \mathrm{d}\Omega\; Y^*_{\ell m}(\hat{n}) \frac{\delta T(\hat{n})}{\bar{T}}\,. \tag{1.85}$$

One can then define the angular power spectrum as

$$C_\ell^{TT} \equiv \frac{1}{2\ell+1} \sum_m \langle a^*_{\ell m} a_{\ell m} \rangle\,, \tag{1.86}$$

where the superindex TT simply indicates that this is a temperature power spectrum, and angular brackets denote an ensemble average. Upon inversion, this relation gives

$$\langle a^*_{\ell m} a_{\ell' m'} \rangle = C_\ell^{TT} \delta_{\ell' \ell} \delta_{m' m}\,. \tag{1.87}$$

At the linear level, and under the assumption that tensor perturbations are subdominant with respect to scalar perturbations, one connects the observed fluctuations in

the temperature of the CMB $a_{\ell m}$ with the primordial perturbations in the gravitational potential ζ using the following relation

$$a_{\ell m} = 4\pi(-i)^{\ell} \int \frac{\mathrm{d}^3 k}{(2\pi)^3} \Delta_{T\ell}(k) \zeta_{\mathbf{k}} Y_{\ell m}(\hat{\mathbf{k}}) \,, \tag{1.88}$$

where the function $\Delta_{T\ell}(k)$ is called the transfer function. The transfer function takes into account the evolution of the perturbations after horizon reentry and the connection between the gravitational potential ζ to the temperature fluctuations. This function can be computed for example by using the publicly available codes CAMB [21] or CMBfast [22]. For an explanation of the physics involved, see references [23], [24] and [25]. Using this we can finally write down the relation between the observed angular power spectrum C_ℓ^{TT} and the primordial scalar power spectrum $P_\zeta(k)$:

$$C_\ell^{TT} = \frac{2}{\pi} \int \mathrm{d}k \, k^2 \Delta_{T\ell}(k) \Delta_{T\ell}(k) P_\zeta(k) \,. \tag{1.89}$$

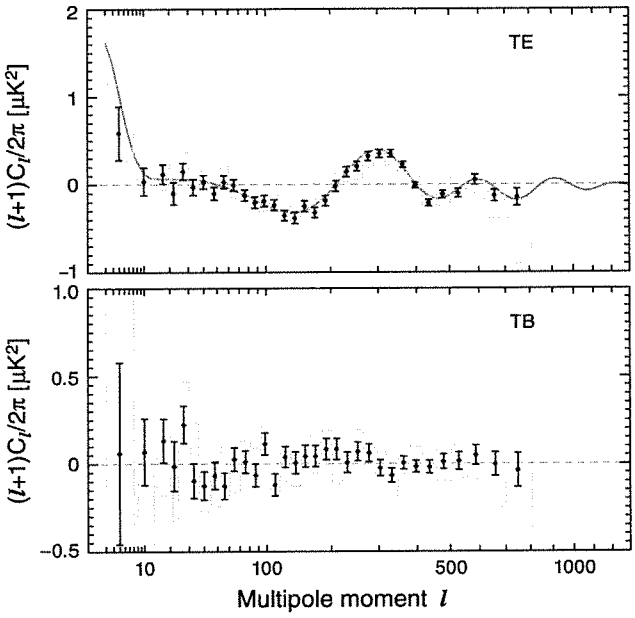


Figure 1.2: The CMB TE and TB angular power spectra as observed by the WMAP satellite, taken from reference [20].

Though the inflationary hypothesis has yet to be proven, evidence in its favor is beginning to accumulate. The most compelling evidence for inflation is the oscillations observed in the angular power spectra. Since the perturbations generated at inflation are created with coherent phases, *i.e.* all modes with the same wavelength will have the same phase, once a mode reenters the horizon it will oscillate coherently and produce the observed peaks in the CMB (figures 1.1 and 1.2). If the oscillations were not

coherent, they would interfere and wash out this peak structure. It is possible in principle that some other physics produces the observed peak structure in the TT angular power spectrum that does not involve inflation. Indeed, processes that occur after the last scattering, like the Integrated Sachs Wolfe (ISW) effect can leave an imprint on the TT angular power spectrum at scales larger than the horizon at recombination. However, there is a negative peak in the TE angular power spectrum (that is, the power spectrum for the correlation of the temperature with the E-mode polarization, see reference [23]) at around $\ell \sim 150$ that corresponds to a scale larger than the horizon at recombination, see figure 1.2. While inflation naturally predicts this (see reference [26]), a model involving causal physics at recombination would not be able generate coherent oscillations outside of the horizon at the surface of last scattering. Furthermore, polarization fluctuations are produced by Thompson scatterings at the epoch of last scattering and are not affected by the ISW effect.

We will see in section 1.4 that inflation generically predicts the fluctuations in ζ to obey nearly gaussian statistics. This hypothesis has ben tested to order $\mathcal{O}(10^{-4})$ in the CMB (see section 1.5), making the small constraints on a deviation from non-gaussianity one of the most precise quantitative evidences for inflation. However, the observation of deviations from non-gaussianity of that order is not only compatible with the inflationary hypothesis, but would also contain information about the physics behind it.

1.4 Deviations from gaussianity.

By gaussianity we mean that given the random field $\phi(x)$ (*eg.* the field ζ will take a random value at each point x as dictated by its generation as quantum fluctuations during inflation) its only connected correlation functions are its two point correlation functions

$$\langle \phi(x_1)\phi(x_2)\rangle_c \neq 0\,, \qquad \text{and} \quad \langle \phi(x_1)\dots\phi(x_{n>2})\rangle_c = 0\,, \tag{1.90}$$

where angular brackets denote ensemble averages. This implies for example that the probability that it take a value ϕ_o at some point x is given by a Gaussian probability distribution function (PDF), as seen by writing it down in terms of the moment generating function

$$P(\phi_o, x) = \int \frac{\mathrm{d}\lambda}{2\pi} \exp\left[\sum_{p=2}^{\infty} \frac{i^p}{p!}\lambda^p \langle \phi^p(x)\rangle_c\right] e^{-i\lambda\phi_o} \tag{1.91}$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\phi_o^2}{2\sigma^2}} \tag{1.92}$$

where $\sigma^2 \equiv \langle \phi^2(x)\rangle_c$ is the variance and we assumed for simplicity that the mean value of the field is zero.

However, we will see that since the dynamics governing ζ is non-linear, it will in general have a three-point function which is different from zero, thus deviating from

gaussianity. We will now study the three-point function generated in several models of inflation and its shapes. We will only present the main results and point the reader to some useful references for more complete derivations and references to the original literature. Before diving into the calculations let us define the bispectrum

$$\langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \rangle \equiv (2\pi)^3 \delta(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) B(k_1, k_2, k_3) \,. \tag{1.93}$$

Here we used the fact that the three point function must be proportional to the Dirac delta due to translational invariance of the underlying theory and the bispectrum $B(k_1, k_2, k_3)$ depends only on the moduli of the momenta due to the rotational invariance[1].

1.4.1 Single field slow-roll models.

We will start by studying the three point function in the models that we have so far considered in this chapter. Here we only sketch the computation and point the interested reader to the original literature, references [4] and [27]. To begin, we compute the third order action for ζ. The procedure is similar to what was done in the previous section: start by writing down the action in the ADM form (1.33) and solve the constraint equations (1.36) and (1.37). It turns out that it is only necessary to solve these constraints at the linear level, since the second order pieces of the lapse N and the shift N^i will always multiply the first order constraint equations, while their third order pieces will have factors that vanish due to the background equations of motion. Replacing the linear solution for the constraint equations, one finds the third order piece of the action

$$\begin{aligned} S = \int \mathrm{d}^4x \Big\{ & M_{pl}^2 a e^{\zeta} \Big(1 + \frac{\dot{\zeta}}{H}\Big) \big[-2\nabla^2\zeta - (\partial_i\zeta)^2 \big] + a^3 e^{3\zeta} \frac{\dot{\phi}^2}{H^2} \dot{\zeta}^2 \Big(1 - \frac{\dot{\zeta}}{H}\Big) \\ & + M_{pl}^2 a^3 e^{3\zeta} \Big[\frac{1}{2} \big(\partial_i\partial_j\psi\partial^i\partial^j\psi - (\nabla^2\psi)^2\big) \Big(1 - \frac{\dot{\zeta}}{H}\Big) - 2\partial_i\psi\partial^i\zeta\nabla^2\psi \Big] \Big\} \,. \end{aligned} \tag{1.94}$$

After performing the field redefinition

$$\zeta = \zeta_c + \frac{1}{2}\frac{\ddot{\phi}}{\dot{\phi}H}\zeta_c^2 + \frac{1}{8M_{pl}^2}\frac{\dot{\phi}^2}{H^2}\zeta_c^2 + \frac{1}{4M_{pl}^2}\frac{\dot{\phi}^2}{H}\partial^{-2}(\zeta_c\nabla^2\zeta_c) + \dots \,, \tag{1.95}$$

[1] To prove this, simply take the inverse Fourier transform of $\langle\zeta^3\rangle$ with respect to all the three modes,

$$\int \mathrm{d}^3x_1 \mathrm{d}^3x_2 \mathrm{d}^3x_3 \; e^{-i(\mathbf{x}_1\cdot\mathbf{k}_1 + \mathbf{x}_2\cdot\mathbf{k}_2 + \mathbf{x}_3\cdot\mathbf{k}_3)} \langle\zeta^3\rangle \,,$$

under a translation of the three spatial coordinates by the same vector $\delta\mathbf{x}$, this will involve a factor $e^{\delta\mathbf{x}\cdot(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)}$, and thus the three $\mathbf{k}$ must form a triangle to guarantee translational invariance. Furthermore, if B depended on the directions of the momenta $\mathbf{k}$, a rotation of the spatial coordinates $\mathbf{x}$ would rotate the momenta thus changing B.

where terms that vanish out of the horizon or that are of a higher order in slow-roll were omitted, Maldacena finds

$$S = \int \mathrm{d}^4x \frac{\dot{\phi}^4}{M_{pl}^2 H^4} a^5 H \dot{\zeta}_c^2 \partial^{-2} \dot{\zeta}_c + \dots . \tag{1.96}$$

Notice that this action is multiplied by an additional factor of ε with respect to the second-order action, this expresses the fact that a large self-coupling of the field would break the slow-roll approximation. From this equation he then computes the three point function for the fluctuations

$$\langle \zeta_c \zeta_c \zeta_c \rangle = (2\pi)^3 \delta(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) \frac{1}{2k_1^3 k_2^3 k_3^3} \frac{H_*^6}{M_{pl}^2 \dot{\phi}_*^2} \frac{\sum_{i>j} k_i^2 k_j^2}{k_1 + k_2 + k_3} . \tag{1.97}$$

After going back to ζ he finally gets

$$\begin{aligned} \langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \rangle = (2\pi)^3 \delta(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) \frac{H_*^8}{\dot{\phi}_*^4} \frac{1}{8k_1^3 k_2^3 k_3^3} \\ \times \left[(-2\eta + \varepsilon) \sum_i k_i^3 + 2\varepsilon \left(\frac{1}{2} \sum_{i \neq j} k_i k_j^2 + 4 \frac{\sum_{i>j} k_i^2 k_j^2}{k_1 + k_2 + k_3} \right) \right] . \end{aligned} \tag{1.98}$$

This means that the level of non-gaussianity generated by slow-roll models of inflation is suppressed by the slow-roll parameters

$$\frac{B(k_1, k_2, k_3)}{P_\zeta(k_1) P_\zeta(k_2)} \propto \mathcal{O}(\epsilon, \eta) , \tag{1.99}$$

and thus at least of order 10^{-1}. We will see in the following subsection that present experiments have a precision of order $\mathcal{O}(30)$ when measuring this quantity, and planned experiments will achieve a precision of $\mathcal{O}(3)$. Such a small value of the non-gaussianity is thus unobservable.

1.4.2 The consistency relation.

Now we wish to prove that in all single-field models of inflation the three-point function satisfies (see references [4] and [28], here we will follow the derivation of reference [29])

$$\lim_{k_1 \to 0} \langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_1} \rangle = -(2\pi)^3 \delta(\mathbf{k_1} + \mathbf{k_2} + \mathbf{k_3}) P_\zeta(k_1) P_\zeta(k_2) \frac{\mathrm{d} \ln k_2^3 P_\zeta(k_2)}{\mathrm{d} \ln k_2} . \tag{1.100}$$

In order to do so we follow a separate universe approach similar to the one we used in section 1.3 to prove the conservation of ζ outside of the horizon. There, we performed a perturbative expansion in gradients and showed that it is consistent to assume that the universe at scales larger than the horizon can be well described locally by an FLRW metric

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2(t) e^{2\zeta_B(\mathbf{x})} \mathrm{d}\mathbf{x}^2 , \tag{1.101}$$

where the subindex B indicates that this can be thought of as a background on top of which one perturbs, and ζ_B depends only on space since it is constant at superhorizon scales. Let us then compute the correlation function of ζ in the presence of a background wave ζ_B at two points $\mathbf{x}_1$ and $\mathbf{x}_2$ separated by a distance much smaller than the scale of variation of this background

$$\langle\zeta\zeta\rangle_B(\mathbf{x}_1,\mathbf{x}_2) \approx \langle\zeta\zeta\rangle_0(|\mathbf{x}_1-\mathbf{x}_2|) + \zeta_B\left(\frac{\mathbf{x}_1+\mathbf{x}_2}{2}\right)\frac{\mathrm{d}}{\mathrm{d}\zeta_B}\langle\zeta\zeta\rangle(|\mathbf{x}_1-\mathbf{x}_2|)\Big|_0 , \tag{1.102}$$

where the subindex 0 means an evaluation of a quantity in the absence of the background wave. The presence of the background mode ζ_B can be seen from equation (1.101) to be equivalent to a rescaling of the spatial coordinates, and we can thus exchange the derivative with respect to ζ_B for a derivative with respect to $\ln(|\mathbf{x}_1-\mathbf{x}_2|)$. After performing a Fourier transform of this equation one obtains

$$\langle\zeta\zeta\rangle_B(\mathbf{k}_1,\mathbf{k}_2) \approx \langle\zeta\zeta\rangle_0(k_S) + \zeta_B(k_L)\frac{1}{k_S^3}\frac{\mathrm{d}}{\mathrm{d}\ln k_S}[k_S^3\langle\zeta\zeta\rangle(k_S)] , \tag{1.103}$$

where $\mathbf{k}_L \equiv \mathbf{k}_1+\mathbf{k}_2$ and $\mathbf{k}_S \equiv (\mathbf{k}_1-\mathbf{k}_2)/2$. The correlation of one mode larger than the horizon with two short modes then gives

$$\langle\zeta_{\mathbf{k}_1}\zeta_{\mathbf{k}_2}\zeta_{\mathbf{k}_3}\rangle = -\langle\zeta_{\mathbf{k}_1}\zeta_{\mathbf{k}_L}\rangle\frac{1}{k_S^3}\frac{\mathrm{d}}{\mathrm{d}\ln k_S}[k_S^3\langle\zeta\zeta\rangle(k_S)] \tag{1.104}$$

$$= -(2\pi)^3\delta(\mathbf{k_1}+\mathbf{k_2}+\mathbf{k_3})P_\zeta(k_1)P_\zeta(k_2)\frac{\mathrm{d}\ln {k_2}^3P_\zeta(k_2)}{\mathrm{d}\ln k_2} \tag{1.105}$$

$$= -(2\pi)^3\delta(\mathbf{k_1}+\mathbf{k_2}+\mathbf{k_3})(n_s-1)P_\zeta(k_1)P_\zeta(k_2) , \tag{1.106}$$

where the mode $\mathbf{k}_1$ has been taken to be long and it correlates only with the background mode. The two short modes will eventually also exit the horizon during inflation. This means that for all single field models of inflation the bispectrum in the limit in which one of the modes is much longer than the other two (called the squeezed limit $k_1 \ll k_2, k_3$) will be

$$\lim_{k_1\to 0}\frac{B(k_1,k_2,k_3)}{P_\zeta(k_2)P_\zeta(k_3)} = (1-n_s) , \tag{1.107}$$

and remember that (n_s-1) has been observed to be of the order 10^{-1}. Notice that this is true independently of whether the slow-roll conditions hold as they where never used in the proof. This also means that if a large non-gaussianity in the squeezed limit is observed, single-field models of inflation are ruled out and one must resort to scenarios with multiple fields like the curvaton, which will be explained below.

1.4.3 Simple models that generate large non-gaussianities

Here we will take a quick glance at two simple models that generate large non-gaussianities. This is just the tip of the iceberg and much more information about these and other models can be found in references [30; 31].

Even though all models with a single field must satisfy the consistency relation presented in the previous section, the non-gaussianity in other configurations of the bispectrum can be rendered large. Let us now present the simplest example of this case, in which we consider a single scalar field minimally coupled to gravity with an action of the form

$$S = \int \mathrm{d}^4x \, \sqrt{-g} P(X, \phi) \,, \tag{1.108}$$

where $X \equiv \partial_\mu \phi \partial^\mu \phi$ and P is a generic function of X and ϕ. Upon computing the stress-energy tensor one obtains the following value for the speed of sound[1]

$$c_s^2 = \frac{P_{,X}}{P_{,X} + 2X P_{,XX}} \,. \tag{1.109}$$

Computing the second order action for ζ we obtain

$$S = \int \mathrm{d}^4x \, \varepsilon \left[a^3 \frac{\dot{\zeta}^2}{c_s^2} - a(\partial_i \zeta)^2 \right] , \tag{1.110}$$

where the time kinetic term for ζ is now multiplied by a factor of $1/c_s^2$. The third order action will be multiplied by an additional slow roll parameter, as in the simple slow-roll case, but the time derivatives will again be multiplied by factors of $1/c_s$. Thus, for a very small speed of sound $c_s^2 \ll 1$ it is possible to enhance some pieces of the third-order action. Of course this model must still satisfy the consistency relation of the previous subsection, thus producing a non-gaussianity in the squeezed limit which is proportional to $(n_s - 1)$, but it can induce a large bispectrum in the equilateral limit $k_1 = k_2 = k_3$ of the order $\mathcal{O}(1/c_s^2)$. For more details on these computations we point the reader to reference [30]. The shape of the generated bispectrum is such that it is similar to the equilateral shape that we will introduce in section 1.5.

Let us now discuss one of the simplest models that can generate a large non-gaussianity in the squeezed limit, we will follow reference [32]. Suppose that apart from the inflaton there is another scalar field σ, termed curvaton, with a mass much smaller than the Hubble parameter during inflation $\partial^2/\partial\sigma^2 V(\sigma) \ll H$. Inflation will generate perturbations of this field on superhorizon scales with a power spectrum of the form (see section 1.3)

$$P_{\delta\sigma}(k) = \frac{H_*^2}{2k_*^3} \,. \tag{1.111}$$

Furthermore, assume that the inflaton perturbations are negligible, and the curvaton field has negligible couplings with other fields (such that it decays into radiation much after the end of inflation but much before nucleosynthesis). Its perturbations will then have an equation of motion at superhorizon scales of the form

$$\ddot{\delta\sigma} + 3H\dot{\delta\sigma} + \frac{\partial^2}{\partial\sigma^2} V(\sigma) = 0 \,. \tag{1.112}$$

[1]The speed of sound is a gauge dependent quantity, here we use the one defined in the comoving gauge.

Inflation will eventually come to an end and the Hubble parameter will start to decrease. When H becomes smaller than the mass of the curvaton, the fluctuations will start to oscillate. After a few Hubble times one can take the approximation $V \approx m_\sigma^2 \sigma^2$, and the energy density of the field can be written as

$$\rho_\sigma = m_\sigma^2 \tilde{\sigma}^2 \,, \tag{1.113}$$

where $\tilde{\sigma}^2$ is the amplitude of the oscillations. We will see that this quadratic relation will be the source for large non-gaussianity in the squeezed limit. Writing the fractional energy density perturbation up to quadratic order in the fractional curvaton perturbation gives

$$\frac{\delta\rho_\sigma}{\rho_\sigma} \approx 2\frac{\delta\sigma}{\sigma} + \left(\frac{\delta\sigma}{\sigma}\right)^2 . \tag{1.114}$$

In order to compute the bispectrum in the squeezed limit, write down the curvature perturbation in the constant energy density gauge, which in an arbitrary gauge can be computed to be

$$\zeta_{cd} = -\Psi - H\left(\frac{\delta\rho}{\dot{\rho}}\right) \tag{1.115}$$

$$= -\Psi - \frac{\delta\rho_\sigma}{4\rho_r + 3\rho_\sigma} \,, \tag{1.116}$$

where in the second line we computed its value before the decay of the curvaton. In the limit of instantaneous decay, the radiation and the curvaton form a single adiabatic fluid and the curvature perturbation is conserved outside of the horizon, so that this expression is valid always. If we now chose a gauge in which $\Psi = 0$ we obtain

$$\begin{aligned} \zeta_{cd} &= -\frac{\rho_\sigma}{4\rho_r + 3\rho_\sigma}\left(\frac{\delta\rho_\sigma}{\rho_\sigma}\right) \\ &= -\frac{\rho_\sigma}{4\rho_r + 3\rho_\sigma}\left[2\frac{\delta\sigma}{\sigma} + \left(\frac{\delta\sigma}{\sigma}\right)^2\right] . \end{aligned} \tag{1.117}$$

We have been working in the superhorizon limit, where $\zeta = \zeta_{cd}$, so let us drop the subindex cd from now on. Rewriting this expression in terms of its linear approximation ζ^l gives

$$\zeta = \zeta^l - \frac{4\rho_r + 3\rho_\sigma}{2\rho_\sigma}\zeta^{l\,2} . \tag{1.118}$$

This is precisely a non-gaussianity of the local type (see section 1.5). It is now easy to compute the three point function for ζ using this expression, the result is

$$\lim_{k_1 \to 0} \frac{B(k_1, k_2, k_3)}{P_\zeta(k_2) P_\zeta(k_3)} \approx -\frac{24}{5}\frac{4\rho_r + 3\rho_\sigma}{\rho_\sigma} . \tag{1.119}$$

Which can be much larger than one.

In studying the curvaton we made many rough approximations like neglecting the gradient terms in the transformation between ζ_{cd} and ζ, and non-linear corrections to

the gravitational potentials. A more rigorous approach can be found in reference [31]. The generation of a large non-gaussianity in the local limit is actually a feature of several models involving more than one field. This is due to the fact that the relation of the fluctuations of these additional fields with ζ outside of the horizon is in general non-linear.

1.5 Constraints on non-gaussianity from the CMB.

As explained in the previous sections, non-gaussianity provides a unique window to the physics of the very early universe, providing a way to test the inflationary hypothesis, discriminate between inflationary models, and maybe even measure the coefficients in the Lagrangian governing its dynamics. Now we wish to shortly describe how non-gaussianity is observed and constrained. In this section we will describe constraints coming from the CMB which are the most accurate to date. Most of this material is taken from reference [6], for a more comprehensive introduction see reference [33].

If the CMB were gaussian, the probability of observing a certain distribution of temperature fluctuations in the CMB given an experiment with a number of ℓ, m modes N_{harms} is given by the normal probability distribution

$$\begin{aligned} P(a) &= \frac{1}{(2\pi)^{N_{harm}/2}|C|^{1/2}} \exp\left[-\frac{1}{2} \sum_{\ell m \ell' m'} a_{\ell m} (C^{-1})_{\ell m, \ell' m'} a_{\ell' m'} \right] \\ &= \prod_{\ell m} \frac{e^{-|a_{\ell m}|^2/(2C_\ell)}}{\sqrt{2\pi C_\ell}}, \end{aligned} \tag{1.120}$$

where $C_{\ell m, \ell' m'} \equiv \langle a_{\ell m} a_{\ell' m'} \rangle$ and we used the fact that $C_{\ell m, \ell' m'} = C_\ell \delta_{\ell \ell'} \delta_{m m'}$. As was said before, the normal distribution is completely determined by the two point function. The generalization of this relation can be obtained by using a formula similar to equation (1.92) in order to obtain, upon linearizing in the three point function,

$$\begin{aligned} P(a) = &\left[1 - \frac{1}{6} \sum_{\ell_1 m_1} \sum_{\ell_2 m_2} \sum_{\ell_3 m_3} \langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \rangle \frac{\partial}{\partial a_{\ell_1 m_1}} \frac{\partial}{\partial a_{\ell_2 m_2}} \frac{\partial}{\partial a_{\ell_2 m_2}} \right] \\ &\times \frac{1}{(2\pi)^{N_{harm}/2}|C|^{1/2}} \exp\left[-\frac{1}{2} \sum_{\ell m \ell' m'} a_{\ell m} (C^{-1})_{\ell m, \ell' m'} a_{\ell' m'} \right]. \end{aligned} \tag{1.121}$$

What is done by the people who analyze the data is then to parametrize the angular bispectrum $\langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \rangle$ in terms of a few quantities and then estimate the value of those parameters for which the PDF is maximized.

The ideal way to estimate the contribution of some primordial model to the observed non-gaussianity would be then to connect ζ to $\delta T/\bar{T}$ as done when we studied the power spectrum earlier[1] and decompose the resulting temperature bispectrum in spherical

[1] Of course non-linearities in this connection will introduce a "contamination" of the primordial signal, one of the purposes of this thesis is to estimate these effects.

harmonics. The angular bispectrum must have the following shape

$$\langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \rangle = \mathcal{G}^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3} f_{NL} b_{\ell_1 \ell_2 \ell_3} \,, \tag{1.122}$$

where $\mathcal{G}$ is called the Gaunt integral to be defined below, $b_{\ell_1 \ell_2 \ell_3}$ is called the reduced bispectrum containing the information about the "shape" of the angular bispectrum, and f_{NL} contains information about its magnitude. The fact that the angular bispectrum is proportional to the Gaunt integral comes from the fact that three-point function in k-space must be proportional to a Dirac delta, *i.e.* it comes from the fact that the underlying theory is invariant under translations. Similarly the fact that the reduced bispectrum does not contain the indexes m_1, m_2, m_3 is a consequence of rotational invariance. The Gaunt integral is defined as

$$\mathcal{G}^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3} \equiv \int \mathrm{d}\Omega \, Y_{\ell_1 m_1}(\hat{\mathbf{n}}) Y_{\ell_2 m_2}(\hat{\mathbf{n}}) Y_{\ell_3 m_3}(\hat{\mathbf{n}}) \,. \tag{1.123}$$

One can indeed check that in the flat sky approximation (see appendix 3.A) the Gaunt integral goes to $(2\pi)^2 \delta(\mathbf{l}_1 + \mathbf{l}_2 + \mathbf{l}_3)$.

In reality the complicated k-dependence of the primordial bispectra generated by different inflationary models make this approach computationally unfeasible and what is done instead is to propose some templates for the bispectrum $b^{(i)}$ such that

$$\langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \rangle \approx \sum_i \mathcal{G}^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3} f^{(i)}_{NL} b^{(i)}_{\ell_1 \ell_2 \ell_3} \,. \tag{1.124}$$

One can then measure the parameters $f^{(i)}_{NL}$ for each template and compare that with the theoretical estimates. We will be more explicit on this point below. Using this can maximize the PDF by solving $\mathrm{d} \ln P / \mathrm{d} f^{(i)}_{NL} = 0$, thus obtaining

$$f^{(i)}_{NL} = \sum_j \left(F^{-1}\right)_{ij} S_i \,, \tag{1.125}$$

where S_i is defined as

$$\begin{aligned} S_i = \frac{1}{6} \sum_{\ell_1 \ell_2 \ell_3} \sum_{m_1 m_2 m_3} & \mathcal{G}^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3} b^{(i)}_{\ell_1 \ell_2 \ell_3} \\ & \times \left[\left(C^{-1} a\right)_{\ell_1 m_1} \left(C^{-1} a\right)_{\ell_2 m_2} \left(C^{-1} a\right)_{\ell_3 m_3} - 3 \left(C^{-1}\right)_{\ell_1 m_1, \ell_2 m_2} \left(C^{-1} a\right)_{\ell_3 m_3} \right] , \end{aligned} \tag{1.126}$$

and F_{ij} is the Fisher matrix

$$F_{ij} = \frac{f_{sky}}{6} \sum_{\ell_1 \ell_2 \ell_3} \sum_{m_1 m_2 m_3} \left(\mathcal{G}^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3} \right)^2 \frac{b^{(i)}_{\ell_1 \ell_2 \ell_3} b^{(j)}_{\ell_1 \ell_2 \ell_3}}{C_{\ell_1} C_{\ell_2} C_{\ell_3}} \,, \tag{1.127}$$

where f_{sky} is the fraction of the sky observed by the experiment. The Fisher matrix, when normalized to its amplitude, is also sometimes called the "cosine" between two

bispectrum shapes, if it is close to one the two bispectrum shapes i and j are said to be "similar". Given a theoretical model one computes the cosine of its predicted primordial bispectrum (taking into account secondary and second-order effects) with one of the standard shapes constrained by observations; if the result is outside the constraints the model is ruled out. It thus makes sense to study shapes that are similar to the ones predicted by the most interesting (or popular) models but which are also simple computationally-wise.

One of the shapes that is often studied is the local shape. A way to define it is to suppose that ζ can be expressed in terms of some gaussian field ζ_g as[1]

$$\zeta(x) = \zeta_g(x) + \frac{3}{5} f_{NL}^{loc} (\zeta_g^2(x) - \langle \zeta_g^2(x) \rangle) \,. \tag{1.128}$$

Using this expression one obtains a shape for the three point function in k space

$$\begin{aligned} \langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \rangle &= \int \frac{\mathrm{d}^3 q}{(2\pi)^3} \langle \zeta_{g\,\mathbf{k}_1} \zeta_{g\,\mathbf{k}_2} \zeta_{g\,\mathbf{k}_3 - \mathbf{q}} \zeta_{g\,\mathbf{q}} \rangle + 2 \text{ cyclic permutations} \\ &= (2\pi)^3 \delta(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) \\ &\quad \times \frac{6}{5} f_{NL}^{loc} \Big[P_\zeta(k_1) P_\zeta(k_2) + P_\zeta(k_2) P_\zeta(k_3) + P_\zeta(k_3) P_\zeta(k_1) \Big] \,. \end{aligned} \tag{1.129}$$

From this expression it is easy to compute the reduced bispectrum. Instead of doing that let us make a few comments. The last line of equation (1.129) is the k space bispectrum which in the squeezed limit $k_1 \ll k_2, k_3$ goes like $B^{loc}(k_1, k_2, k_3) = (12/5) f_{NL}^{loc} P_\zeta(k_1) P_\zeta(k_2)$, so that the quantity that in the previous section we have been calling informally the non-gaussianity goes to a constant

$$\lim_{k_1 \to 0} \frac{B^{loc}(k_1, k_2, k_3)}{P_\zeta(k_1) P_\zeta(k_3)} = \frac{12}{5} f_{NL}^{loc} \,. \tag{1.130}$$

Equation (1.130) also means that, according to the consistency relation that we proved in the previous section, all single-field models must have a small local non-gaussianity. If a large local non-gaussianity is observed, we would then have to resort to models like the curvaton which involve several fields.

Another shape which is often studied is the equilateral shape [34], which is defined as

$$\begin{aligned} B^{equil}(k_1, k_2, k_3) = \frac{36}{5} f_{NL}^{equil} (2\pi^2)^2 A_\zeta^2 k_*^{2(1-n_s)} \Bigg[&- \frac{1}{2 k_1^{4-n_s} k_2^{4-n_s}} - \frac{1}{3 (k_1 k_2 k_3)^{2(4-n_s)/3}} \\ &+ \frac{1}{k_1^{(4-n_s)/3} k_2^{(4-n_s)/2} k_1^{4-n_s}} + 5 \text{ permutations} \Bigg] \,. \end{aligned} \tag{1.131}$$

[1] This odd definition comes from the fact that it is conventionally defined in terms of the Bardeen potential Φ rather than the curvature perturbation ζ.

This shape is normalized such that it equals the local shape when $f_{NL}^{loc} = f_{NL}^{equil} = 1$ and all the three modes are equal. It was defined such that it has a cosine with the shape generated in models with a small speed of sound which is close to one while also having a simple form. It is called equilateral shape since, when normalized with a factor $(k_2/k_1)^2(k_3/k_1)^2$, it peaks for the equilateral triangle configuration $k_1 = k_2 = k_3$. Notice also that, by looking at the form of the three point function for the simple single field slow roll model, equation (1.98),one concludes that the non-gaussianity it generates in the equilateral shape is $f_{NL}^{equil} \approx (5/3)\varepsilon$. It is also interesting to look at references [35] and [36] where it is argued that the measurement of the non-gaussianity for different shapes (namely the local, equilateral and orthogonal shapes), is a direct measurement of the coefficients of the effective theory for single field inflation.

Up to now we have been ignoring secondary effects, that is non-primordial sources of non-gaussianity. The local form couples a long-wavelength fluctuation with two short-wavelength modes, therefore a mechanism like gravitational lensing can contribute to this shape. Indeed, when the gravitational potential Φ evolves due to the presence of dark energy and the non-linear growth of structure, it induces a change in the temperature perturbations usually called integraged Sachs Wolfe effect (ISW). The correlation of a long ISW mode with the lensing potential generates a contribution to the non-gaussianity which is of order $f_{NL} \sim 10$, and is thought to be the largest secondary source (see reference [37]). However, its shape is known to good accuracy and can be subtracted from the data, leaving a "contamination" in the primordial of $\Delta f_{NL} \sim 2$ for the WMAP analysis (see reference [6]).

We saw in several points of the previous calculation that non-linear effects in the dynamics of ζ are the source of the non-gaussianity (primordial or otherwise). Therefore another non-primordial effect that must be taken into account is the non-linearity of the transfer function. The precise computation of this effect is rather involved, and needs the use of numerics (see references [38; 39; 40; 41; 42]), but an analytical estimate can be done for the physical limit in which one observes very large scales and supposes the universe to be always matter dominated. This limit is a way to isolate the non-linear gravitational effects while preserving the gauge-invariance of the final result (see reference [3]). The following two chapters of this thesis will be dedicated to this topic.

Current 2σ limits on f_{NL}^{loc} and f_{NL}^{equil} from the WMAP satellite (reference [11]) give

$$f_{NL}^{local} = 32 \pm 21\,, \qquad f_{NL}^{equil} = 26 \pm 140\,. \tag{1.132}$$

The 2σ deviation of the local value from zero could be an indication that we are close to measuring it. The PLANCK satellite, which at the time of writing has finished taking its first year of data, promises a much higher sensitivity of $\Delta f_{NL} \sim 5$. If this sensitivity is actually achieved, the local non-gaussianity will be either measured (and thus single field models will be ruled out due to the consistency relation) or the constraint will be so small that second order and secondary effects are expected to become relevant and maybe dominant. Since the fluctuations are already of the order $\mathcal{O}(10^{-5})$, these constraints on f_{NL}^{loc} actually imply that the gaussianity of the fluctuations has been tested to order $\mathcal{O}(10^{-4})$ in the CMB, making it one of the most precise quantitative

evidences in favor of inflation.

1.6 Constraints on non-gaussianity from the LSS.

The study of the possibility of observing non-gaussianity in the LSS has recently become a very active topic of research. This is due in part to the advent of many observations of large numbers of galaxies with the potential to accumulate sufficient statistics. The observation of the non-gaussianity can be done in three ways: the direct observation of the bispectrum for the number density of dark matter haloes, very precise measurements of the mass function (that is the number density of dark matter haloes per mass interval), or a scale-dependent effect on the halo bias. For comprehensive recent reviews see references [5] and [43]. We will study here the halo bias (for which we follow reference [44]) and the bispectrum (for which we follow reference [45], for a more recent study see reference [46]), and refer the reader to chapter 4 for the study of the mass function.

In order to study the bispectrum of the dark matter density contrast $\delta \equiv \delta\rho/\bar{\rho}$, first we connect ζ with the dark matter density perturbation which is a non-linear relation but at the linear level is simply

$$\delta(\mathbf{k}, z) \equiv M(k, z)\zeta(k) \equiv -\frac{2k^2}{5\Omega_m H_0^2} T(k) D(z) \zeta(k) \,. \tag{1.133}$$

Here $T(k)$ is the transfer function of perturbations normalized to unity at $k \to 0$ and encodes the scale dependence coming from the fact that larger scales reenter the horizon later; $D(z)$ is the linear growth factor that encodes the growth of density perturbations with time. Thus one can write the contribution of the primordial bispectrum to the total observed three point function as

$$\langle \delta_{\mathbf{k}_1} \delta_{\mathbf{k}_2} \delta_{\mathbf{k}_3} \rangle = M(k_1, z) M(k_2, z) M(k_3, z) \langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \rangle \,. \tag{1.134}$$

The non-linear evolution of density perturbations at small scales induces a bispectrum on the number density of haloes even for gaussian initial conditions. Its precise computation is rather complex (see reference [46]) but at leading order one can write

$$\delta_{\mathbf{k}} \approx \delta^{(1)}_{\mathbf{k}} + \int \frac{\mathrm{d}^3 q}{(2\pi)^3} F(\mathbf{q}, \mathbf{q} - \mathbf{k}) \delta^{(1)}_{\mathbf{q}} \delta^{(1)}_{\mathbf{k}-\mathbf{q}} \,, \tag{1.135}$$

where $\delta^{(1)}$ is the linear solution for the density constrast, and we define

$$F_2(\mathbf{k}_1, \mathbf{k}_2) \equiv \frac{5}{7} + \frac{\hat{\mathbf{k}}_1 \cdot \hat{\mathbf{k}}_2}{2} \left(\frac{k_1}{k_2} + \frac{k_2}{k_1} \right) + \frac{2}{7} (\hat{\mathbf{k}}_1 \cdot \hat{\mathbf{k}}_2)^2 \,, \tag{1.136}$$

which then gives a contribution to the bispectrum of the form

$$B^G(k_1, k_2, k_3) = 2 F_2(\mathbf{k}_1, \mathbf{k}_2) P_\delta(k_1) P_\delta(k_2) + \text{cyclic permutations} \,. \tag{1.137}$$

where P_δ is the dark matter density contrast power spectrum $P_\delta(k,z) \equiv M^2(k,z)P_\zeta(k)$. In most of the planned surveys what will be observed are galaxies and not dark matter haloes. Thus there is one more contribution which must be taken into account, and it comes from the connection of the dark matter density contrast with the galaxy number density contrast $\delta_g \equiv (n_g - \bar{n}_g)/\bar{n}_g$ where n_g is the number density of galaxies. This connection is sometimes written down in the following phenomenological way

$$\delta_g = b_1\delta + \frac{1}{2}b_2\delta^2 + \dots , \tag{1.138}$$

where b_1 and b_2 are called the linear and non-linear galaxy bias respectively. Thus the total galaxy bispectrum will be

$$\begin{aligned} B_g(k_1,k_2,k_3) \approx\ & b_1^3 M(k_1,z)M(k_2,z)M(k_3,z)B_\zeta(k_1,k_2,k_3) \\ & + b_1^3 B^G(k_1,k_2k_3) + b_1^2 b_2\big[P_\delta(k_1)P_\delta(k_2) + \text{cyc. perms.}\big] , \end{aligned} \tag{1.139}$$

where B_ζ is the primordial bispectrum for ζ. A precise study of the gravitational non-linear effects and the non-linear bias are thus necessary if one wishes to go to very high precisions. The claim is that the precision to 1σ of future planned surveys for the measurement of this bispectrum, if non-linear effects are correctly accounted for, is of the order $\Delta f_{NL}^{loc} \sim 0(3)$, which is competitive with PLANCK (see reference [45]). Though this sensitivity is below the one of the scale dependance in the halo bias that we will study below, the big advantage of this method is its sensitivity to the shape of the bispectrum.

In order to study the scale dependent effect on the halo bias, let us follow reference [44] and work in the peak-background split formalism by splitting the density as

$$\rho = \bar{\rho}(1 + \delta_l + \delta_s) , \tag{1.140}$$

where δ_l are the long wavelength modes and δ_s the short wavelength ones. The long wavelength modes are expected to be much smaller on average than the short ones which are the ones that collapse to form haloes. The local Lagrangian density of haloes n will depend on the long wavelength piece since haloes will form more easily when inside a long wavelength overdensity. Thus we write

$$n = \bar{n}(1 + b_L\delta_l) , \tag{1.141}$$

where the Lagrangian halo bias b_L is defined as

$$b_L = \bar{n}^{-1}\frac{\partial n}{\partial \delta_l} . \tag{1.142}$$

The Eulerian halo bias b is related to the Lagrangian bias by $b = b_L + 1$. Now suppose we introduce a non-gaussianity with a local shape (1.128), and we split ζ_g in short and long modes

$$\zeta = \zeta_{g\,l} + \frac{3}{5}f_{NL}^{loc}\zeta_{g\,l}^2 + (1 + \frac{6}{5}f_{NL}^{loc}\zeta_{g\,l})\zeta_{g\,s} + \frac{3}{5}f_{NL}^{loc}\zeta_{g\,s}^2 . \tag{1.143}$$

We can connect the density contrast with ζ by using expression (1.133), thus the long wavelength mode will be

$$\delta_l(k) = M(k,z)\zeta_{g\,l}(k)\,, \tag{1.144}$$

where we ignored the much smaller contribution $\propto f_{NL}^{loc}\zeta_{g\,l}^2$. The short part of δ can similarly be written as

$$\delta_s(k) = M(k,z)\left[\left(1+\frac{6}{5}f_{NL}^{loc}\zeta_{g\,l}\right)\zeta_{g\,s}+\frac{3}{5}f_{NL}^{loc}\zeta_{g\,s}\right]. \tag{1.145}$$

Thus, the presence of the non-gaussianity and of the long mode will modify the amplitude of the perturbations at small scales even at the linear level in these small scale quantities by a factor of $(1+(6/5)f_{NL}^{loc}\zeta_{g\,l})$. From this expression we can again compute the Lagrangian bias using equation (1.142) to obtain

$$b_L = \bar{n}^{-1}\left[\frac{\partial n}{\partial\delta_l}+\frac{6}{5}f_{NL}^{loc}\frac{\partial\zeta_l}{\partial\delta_l}\frac{\partial n}{\partial(1+(6/5)f_{NL}^{loc}\zeta_l)}\right]. \tag{1.146}$$

Notice that the presence of the non-gaussianity modifies the bias. Now we can use the fact that the long wavelength perturbation modifies the amplitude of the short wavelength ones to write

$$\sigma_8^{local} = \left(1+\frac{6}{5}f_{NL}^{loc}\zeta_{g\,l}\right)\sigma_8\,, \tag{1.147}$$

where σ_8 is the variance of the dark matter density contrast when averaged on a scale of 8 $\mathrm{Mpc\,h^{-1}}$. We can now write, using $\partial\zeta_l/\partial\delta_l = M^{-1}(k,z)$, the expression for the total modification of the bias due to the presence of the local non-gaussianity

$$\Delta b(k,z) = \frac{6}{5}f_{NL}^{loc}M^{-1}(k,z)\frac{\partial\ln n}{\partial\ln\sigma_8}\,, \tag{1.148}$$

where we made explicit the fact that this modification depends on the scale k. The characteristic dependence on scale of this expression makes it possible to distinguish it from other effects that change the magnitude of the bias. Using this, reference [44] puts bounds on the local non-gaussianity of

$$-29 < f_{NL}^{loc} < +70\,, \tag{1.149}$$

which is already competitive with current WMAP constraints. Future surveys promise a 1σ precision of $\Delta f_{NL}^{loc}\sim 1$ (see reference [47]), making it one of the most promising probes for the local non-gaussianity.

2

Action approach to cosmological perturbations: the 2nd order metric in matter dominance

2.1 Introduction

As we saw in chapter 1, there has recently been an extraordinary improvement in the accuracy of cosmological observations, especially regarding the statistical properties and the evolution of perturbations around a FLRW Universe. In order to fully exploit the data, it has become necessary to go beyond linear perturbation theory. This is mandatory, for instance, if one wants to study primordial non-Gaussianities which, by definition, are sensitive to the non-linear interactions of the theory.

One expects secon-order effects at recombination to "contaminate" the signal of non-Gaussianity in the CMB. In this chapter and the following we will study the simplified (but physical) case of a universe with no dark energy, such that after the matter-radiation equality it is matter dominated at all times and the gravitational potentials appearing in the metric are constant at first order. In such a Universe one may ask which is the contribution of second-order gravitational effects to the non-Gaussianity (*i.e.* observing only scales which are much larger than the scale of the scale of the first peak of the CMB power spectrum), such calculation will be the second-order analogue of the classic Sachs-Wolfe result [48]. This chapter will draw heavily from [1], where we computed the second-order metric during radiation domination as a first step towards this goal. We will take the further steps in this direction in chapter 3.

The calculation of primordial non-Gaussianities in various models is based on a perturbative expansion of the action – or of the equations of motion – of one or more scalar fields. In general, the mixing of these scalar fields with gravity cannot be neglected. This can be done by using an ADM approach which allows to write an action for the relevant degrees of freedom taking into account gravitational perturbations [4]. In this

chapter we want to extend this formalism to the study of cosmological perturbations *after* inflation, during the standard FLRW evolution.

During its evolution, the Universe is filled with cosmological fluids such as radiation or cold dark matter. At first sight a fluid is very different from a scalar field, so that it is not clear how to extend the aforementioned approach to our case. However, in most cases the cosmological fluids can be considered as perfect and barotropic. In this case, as we will review in section 2.2, the dynamics of the fluid is exactly equivalent to the one of a derivatively coupled scalar field, playing the role of the fluid potential velocity. Using this equivalence, the usual treatment of perturbations in the primordial Universe, used when they are generated and stretched out of the Hubble radius, can be extended to their subsequent evolution in the FLRW Universe, when the modes re-enter the Hubble radius.

In this chapter this approach is applied to find the exact metric at second order during the matter dominated era. Our results are compatible with those obtained previously in [49; 50], by directly working at the level of the Einstein equations. The knowledge of the second order metric is the first step for the calculation of any observable beyond the linear approximation. This calculations are quite relevant: in the case of the CMB anisotropy the expected magnitude of a generic second order effect could in principle be comparable to the sensitivity of the forthcoming Planck satellite.

As we will see, the calculation of the second order metric will closely parallel the one of the three-point function of scalar and tensor perturbations generated during inflation. In particular, we will make use of the action calculated for this purpose for k-inflation [51; 52], expanded up to third order. In this context, it is natural to use the comoving curvature perturbation ζ as the variable describing the scalar perturbations, the same which is commonly employed for the quantization of scalar fluctuations during inflation (it is defined as the scalar part of the metric perturbation in a gauge in which the scalar part of the 3-momentum density is set to zero $T^0_i = 0$, see equation (2.25)). This variable is also well-known to be nonlinearly conserved on super-Hubble scales independently of the details of the cosmological evolution [53]. For this reason it is the natural variable to set the initial conditions for the post-inflationary dynamics. Therefore, an advantage of our approach is that the dynamics of perturbations is described by the same variable and the same approach, from its generation at early times to the late times observations.

As a warm-up, we start with linear perturbation theory in section 2.3. Then, in section 2.4, we extend our calculation to second order for scalar and tensor modes, assuming that primordial gravitational waves are negligible. In this way we obtain the full second order metric in the comoving gauge including tensor modes generated by scalars. We will see that ζ, thought to be conserved outside the horizon, will actually recieve a piece at second order which evolves (see equation (2.51)), but one must connect ζ to observable quantities in order to correctly interpret this result, we will try to give this interpretation in section 2.6. In section 2.5 we check our results with the ones obtained in Poisson gauge [49; 50] after a suitable second order gauge transformation. Conclusions are drawn in section 2.7.

2.2 A fluid as a scalar field

In cosmology, when dissipative phenomena are negligible, the energy content of the Universe can be approximated as a sum of perfect fluids. A perfect fluid is defined to have a stress-energy tensor of the form

$$T_{\mu\nu} = (\rho + p)u_\mu u_\nu + p g_{\mu\nu} \,, \tag{2.1}$$

where ρ and p are the energy density and the pressure, while u^μ is the fluid 4-velocity. Cosmological fluids are usually taken to be irrotational. This assumption is justified by the absence of vorticity in the initial conditions set by inflation and by the fact that vorticity is diluted by the expansion of the Universe. The fluids are also taken to be barotropic, i.e., their pressure is a function of the energy density only, $p = p(\rho)$. Under these conditions, each fluid is characterized by a single scalar function, so that it is not surprising that its dynamics can be described in terms of a scalar field.

Indeed, let us consider a derivatively coupled scalar ϕ in Minkowski spacetime with Lagrangian density[1]

$$\mathcal{L} = P(X) \,, \qquad X \equiv -\partial_\mu\phi\partial^\mu\phi \,. \tag{2.2}$$

Varying the action yields the equation of motion

$$\partial_\mu[P'(X)\partial^\mu\phi] = 0 \,. \tag{2.3}$$

The stress-energy tensor of this field is given by

$$T_{\mu\nu} = 2P'(X)\partial_\mu\phi\partial_\nu\phi + P(X)g_{\mu\nu} \,, \tag{2.4}$$

which is of the perfect fluid form (2.1) if we identify

$$\rho = 2P'X - P \,, \qquad p = P \,, \qquad u_\mu = \frac{\partial_\mu\phi}{\sqrt{X}} \,. \tag{2.5}$$

The perfect fluid interpretation makes sense only if $\partial_\mu\phi$ is everywhere timelike and future directed. Projecting the conservation equation of the stress-energy tensor of a fluid, $\partial_\mu T^\mu_\nu = 0$, along and orthogonal to the fluid flux yields the equation of conservation of energy and the Euler equation. In the case of the stress-energy tensor (2.4), the Euler equation is a trivial identity, while the conservation of the energy is equivalent to the equation of motion (2.3).

The equation of motion (2.3) can be interpreted as the conservation of the current

$$J^\mu = 2\sqrt{X}P'(X)\cdot u^\mu \,. \tag{2.6}$$

This conservation is a consequence (by Nöther theorem) of the invariance of the action under shift of $\phi : \phi \to \phi + \text{const}$. From the fluid point of view this current describes

[1] For Lagrangian approaches which describe also vorticous motion and non-barotropic perfect fluids see [54; 55; 56].

the conserved particle density flux $J^\mu \equiv nu^\mu$, where n is the number particle density. Therefore one can identify $n = 2\sqrt{X}P'(X)$. This yields a physical interpretation of the norm of $\partial_\mu\phi$: $\sqrt{X} = (\rho + P)/n$ is the so called specific inertial mass [54; 55]. It is also straightforward to verify that

$$\sqrt{X} = \frac{d\rho}{dn} , \tag{2.7}$$

so that $\sqrt{X}$ acts as a sort of conjugate variable with respect to n [56].

It is well known that for a perfect fluid the entropy per particle is conserved along the fluid flow. This can be checked using the continuity equation $\partial_\mu J^\mu = 0$, i.e. eq. (2.3), and the energy conservation $u^\mu \partial_\nu T^\nu_\mu = 0$ [57]. Furthermore, as we have discussed, our approach describes a barotropic fluid which implies that the entropy per particle is everywhere constant. In other words, the Lagrangian (2.2) can only describe mechanical excitations of the fluid. It cannot take into account dissipative irreversible processes like heat conduction or viscosity.

We will be interested in studying perturbations around a homogeneous configuration. In Minkowski spacetime $\phi = ct$ is a solution of the equation of motion for any c. Different values of c describe different unperturbed values of the energy density $\rho(c^2)$. The dynamics of fluctuations around this background can be studied by expanding the Lagrangian using $\phi = ct + \delta\phi(t, \vec{x})$, where $\delta\phi$ describes the compressional mode of the fluid. At second order we obtain

$$\mathcal{L} = P'(c^2)[\dot{\delta\phi}^2 - (\nabla\delta\phi)^2] + 2P''(c^2)c^2\dot{\delta\phi}^2 . \tag{2.8}$$

From this expression one sees that the speed of sound of the excitations is given by

$$c_s^2 = \left.\frac{P'(X)}{P'(X) + 2XP''(X)}\right|_{X=c^2} , \tag{2.9}$$

which, as expected, is the usual adiabatic speed of sound in a barotropic fluid,

$$c_s^2 = \left.\frac{p'(X)}{\rho'(X)}\right|_{X=c^2} = \frac{dp}{d\rho} . \tag{2.10}$$

Note that although the Lagrangian (2.2) is Lorentz invariant, this symmetry is spontaneously broken by the vacuum $\phi = ct$. For this reason the speed of perturbations will in general differ from the speed of light $c_s^2 = 1$.

A standard case of barotropic fluid is given by a linear equation of state $p = w\rho$ with $w = \text{const}$. From the first two equalities in eq. (2.5) one deduces (see for instance [58; 59])

$$P = X^{\frac{1+w}{2w}} , \quad w \neq 0 , \tag{2.11}$$

up to a proportionality constant which is irrelevant for the classical theory. In this case, the speed of sound in eq. (2.10) reduces to $c_s^2 = w$. As an example, one can consider a radiation fluid with equation of state $w = 1/3$. In this case the Lagrangian (2.11) reduces to

$$\mathcal{L} = X^2 = (-\partial_\mu\phi\partial^\mu\phi)^2 . \tag{2.12}$$

The inclusion of gravity is completely straightforward. The equation of motion (2.3) becomes

$$\partial_\mu[\sqrt{-g}P'(X)\partial^\mu\phi] = 0 \,. \tag{2.13}$$

In an expanding FLRW background, the homogeneous solution satisfies now

$$n = P'\dot{\phi} \propto a^{-3} \,. \tag{2.14}$$

Indeed, for a constant w this is the standard redshift of the energy density,

$$\rho \propto \dot{\phi}^{\frac{1+w}{w}} \propto a^{-3(1+w)} \,. \tag{2.15}$$

To summarize, we are able to describe the dynamics of one or more fluids coupled with gravity, within a Lagrangian formalism. This can be employed to study cosmological perturbations in the presence of barotropic fluids. Note that the Lagrangian (2.2) is a particular case of the so called $k-$essence/$k-$inflation scenarios [60; 61]. For this class of Lagrangians, cosmological perturbations have been extensively studied starting from [59] and more recently extended to second [51; 52] and third order [62] to study the non-Gaussianities produced during inflation. In this chapter these calculations can be reinterpreted as describing the nonlinearities of a fluid.

In the following we will be interested in the study of dark matter perturbations, which would correspond to the dust case $w = 0$. Obviously the zero pressure limit must be taken with care, as from eq. (2.5) the Lagrangian strictly vanishes for a pressureless fluid.

2.3 Background and linear dynamics

Following the discussion of the last section, the dynamics of a perfect fluid coupled with gravity can be described by the action

$$S = \frac{1}{2}\int \mathrm{d}^4x\sqrt{-g}[R + 2P(X)] \,, \tag{2.16}$$

where we chose units such that $M_P^{-2} \equiv 8\pi G_N = 1$. In particular we are interested in the case $p = w\rho$ with constant w, when P is given by eq. (2.11).

We assume a flat FLRW metric $ds^2 = -dt^2 + a^2(t)d\vec{x}^2$. Friedmann equations read

$$H^2 = \frac{1}{3}\left(2XP' - P\right) \,, \tag{2.17}$$

$$2\dot{H} + 3H^2 = -P \,, \tag{2.18}$$

where $H \equiv \dot{a}/a$ is the Hubble expansion parameter. It is useful to define

$$\epsilon \equiv -\frac{\dot{H}}{H^2} = \frac{3}{2}(1+w) \,. \tag{2.19}$$

Note that, although the notation is inspired by the inflationary case, we are not assuming that ϵ is small.[1]

Following Maldacena [4], the study of perturbations can be done using the ADM splitting of the metric [17],

$$ds^2 = -N^2 dt^2 + h_{ij}(dx^i + N^i dt)(dx^j + N^j dt) \,. \tag{2.22}$$

The action (2.16) becomes

$$S = \frac{1}{2}\int \mathrm{d}t\, \mathrm{d}^3x \sqrt{h}\left[N(R^{(3)} + 2P) + N^{-1}\left(E_{ij}E^{ij} - E^2\right)\right] \,, \tag{2.23}$$

where $R^{(3)}$ is the curvature scalar computed from h_{ij}, $E \equiv E^i{}_i$ and

$$E_{ij} \equiv \frac{1}{2}\left(\dot{h}_{ij} - \nabla_i N_j - \nabla_j N_i\right) \,. \tag{2.24}$$

The covariant derivatives ∇_i are with respect to the 3-metric h_{ij} and all roman indeces $i, j, \ldots$ are raised and lowered with this metric.

The action (2.16) describes 3 dynamical degrees of freedom: one scalar mode for the fluid excitations and the 2 tensor helicities of the gravity waves. In the ADM formalism, these degrees of freedom are contained in the scalar field ϕ and in the 3-metric h_{ij}, while the lapse N and the shift N_i are not dynamical. Following [4] we choose to work in the comoving gauge

$$\begin{gathered} \delta\phi = 0, \qquad h_{ij} = a^2 e^{2\zeta}\hat{h}_{ij} \,, \qquad \hat{h}_{ij} = \delta_{ij} + \gamma_{ij} + \tfrac{1}{2}\gamma_{il}\gamma_{lj} + \ldots \,, \\ \det \hat{h} = 1 \,, \qquad \gamma_{ii} = 0 \,, \qquad \partial_i \gamma_{ij} = 0 \,, \end{gathered} \tag{2.25}$$

where ζ describes the scalar mode and γ the tensor ones. The momentum and Hamiltonian constraints are

$$\nabla_i\left[N^{-1}\left(E^i_j - \delta^i_j E\right)\right] = 0 \,, \tag{2.26}$$

$$R^{(3)} + 2P - 4XP' - \frac{1}{N^2}\left(E_{ij}E^{ij} - E^2\right) = 0 \,. \tag{2.27}$$

To warm up let us calculate the action for the scalar mode ζ at second order, in order to find the linear equation of motion and the first order metric. As there is no mixing between ζ and γ at second order in the action, in this section we can set $\gamma = 0$.

[1] To simplify the comparison with the literature we give here the parameters used in refs. [51; 52] for the $w = $ const case:

$$\eta \equiv \frac{\dot{\epsilon}}{\epsilon H} = 0 \,, \qquad u \equiv 1 - \frac{1}{c_s^2} = 1 - \frac{1}{w} \,, \qquad s \equiv \frac{1}{H}\frac{\dot{c}_s}{c_s} = 0 \,, \tag{2.20}$$

and

$$\Sigma = XP' + 2X^2P'' = \frac{H^2\epsilon}{w} \,, \qquad \lambda = X^2P'' + \frac{2}{3}X^3P''' = -\frac{H^2\epsilon}{6w}\left(1 - \frac{1}{w}\right) \,. \tag{2.21}$$

To compute the second order action, we need to solve the constraints (2.26,2.27) and plug their solution for N and N_i back into the action (2.23). This needs to be done at first order only, since the second-order solutions for N and N^i will multiply $\delta\mathcal{L}/\delta N$ or $\delta\mathcal{L}/\delta N^i$ at zeroth order, which vanish on the background [4]. To solve the momentum constraint (2.26) we decompose N_i as $N_i \equiv \partial_i\psi + N_{Ti}$ where $\partial_i N_{Ti} = 0$. By defining $N = 1 + \delta N$, one finds, at first order,

$$\delta N = \frac{\dot{\zeta}}{H}\,, \qquad N_{T\,i} = 0\,. \tag{2.28}$$

Furthermore, one can find ψ at first order by solving the energy constraint[1] (2.27),

$$\psi = -\frac{\zeta}{H} + \frac{a^2\epsilon}{w}\partial^{-2}\dot{\zeta}\,. \tag{2.29}$$

Note that the last term of the previous equation contains w at the denominator so that one is not allowed to take the limit $w \to 0$ at this stage.

Substituting the solutions (2.28,2.29) into the action (2.23) one obtains, after some integration by parts, the second order action for ζ,

$$S_2 = \int \mathrm{d}t\,\mathrm{d}^3x\, a^3 \frac{\epsilon}{w}\left[\dot{\zeta}^2 - \frac{w}{a^2}(\partial\zeta)^2\right]\,. \tag{2.30}$$

The equation of motion derived from this action is thus

$$\ddot{\zeta} + 3H\dot{\zeta} - \frac{w}{a^2}\partial^2\zeta = 0\,. \tag{2.31}$$

We are interested in the limit $w \to 0$. Setting $w = 0$, the growing solution of eq. (2.31) is simply a constant, $\zeta = \zeta_0$, where ζ_0 is the perturbation generated during inflation, which remains constant on super-Hubble scales independently of the equation of state[2]. However, to find the metric we need to plug the solution for ζ into eqs. (2.28,2.29) and for $\dot{\zeta}$ in eq. (2.29) we need to keep the leading order correction in w in eq. (2.31).

The general solution of this equation in Fourier space can be written as a linear combination of Bessel functions,

$$\zeta_{\vec{k}}(\tau) = C_1\tau^{-\nu}J_\nu(k\tau\sqrt{w}) + C_2\tau^{-\nu}Y_\nu(k\tau\sqrt{w}), \qquad \nu = \frac{3(1-w)}{2(1+3w)}\,, \tag{2.34}$$

[1] Here and in the following $\partial^2 \equiv \partial_i\partial_i$ and ∂^{-2} is its inverse.

[2] For a generic equation of state $p(\rho)$ the action for ζ is

$$S = \int \mathrm{d}t\,\mathrm{d}^3x\, a^3 \frac{\epsilon}{c_s^2}\left[\dot{\zeta}^2 - \frac{c_s^2}{a^2}(\partial\zeta)^2\right]\,, \tag{2.32}$$

where c_s^2 is given by eq. (2.9). This gives the following equation of motion

$$\ddot{\zeta} + \dot{\zeta}\frac{d}{dt}\ln\left(a^3\frac{\epsilon}{c_s^2}\right) - \frac{c_s^2}{a^2}\partial^2\zeta = 0\,. \tag{2.33}$$

From this equation we see that ζ is constant on large scales, $k/aH \ll 1$, independently of the energy content. This is not true, for instance, for the Newtonian potential.

where we have introduced the conformal time $\tau = \int dt/a$. The growing mode of this solution goes as

$$\zeta = \zeta_0 + \frac{2w}{5a^2H^2}\partial^2\zeta_0 + \mathcal{O}(w^2) \,, \tag{2.35}$$

where ζ_0 is the asymptotic value of ζ on super-Hubble scales. The time derivative of (2.35) is

$$\dot{\zeta} = \frac{2w}{5a^2H}\partial^2\zeta_0 + \mathcal{O}(w^2) \,. \tag{2.36}$$

This can be plugged into eqs. (2.28,2.29) which yield, at lowest order in $w \to 0$,

$$\delta N = 0 \,, \qquad \psi = -\frac{2}{5}\frac{\zeta_0}{H} \,. \tag{2.37}$$

The metric at first order in perturbations is thus given by

$$ds^2 = -dt^2 - \frac{4}{5H}\partial_i\zeta_0 \, dt dx^i + a^2(1 + 2\zeta_0)d\vec{x}^2 \,. \tag{2.38}$$

Note that under a coordinate transformation

$$t \to t + \frac{2\zeta_0}{5H} \,, \qquad x^i \to x^i \,, \tag{2.39}$$

this metric takes the Newtonian gauge form

$$ds^2 = -(1 + 2\Phi)dt^2 + a^2(1 - 2\Psi)d\vec{x}^2 \,, \tag{2.40}$$

with

$$\Phi = \Psi = -\frac{3}{5}\zeta_0 \,, \tag{2.41}$$

which is the well known linear relation between the Newtonian potential and ζ_0 in matter dominance.

2.4 Second order perturbations

2.4.1 Scalar perturbations

We are interested in deriving the third order action for ζ. In order to do that we follow the same procedure we did at linear order and initially set $\gamma = 0$ everywhere. Tensor modes generated at second order by scalar perturbations will be considered in the next section. To derive the third order action we only need the solutions to the constraint equations at linear order, i.e. eqs. (2.28,2.29). This is because the second and third order solutions would multiply the first and zeroth order constraint equations, respectively [4]. We will be interested in the second order solutions of the constraint equations only later, when we will derive the explicit expression of the metric at second order.

It is straightforward to derive

$$E_{ij}E^{ij} - E^2 = -6(H+\dot{\zeta})^2 + 4\frac{e^{-2\zeta}}{a^2}(H+\dot{\zeta})\left(\partial_i N_i + \partial_i\zeta N_i\right) + \frac{e^{-4\zeta}}{a^4}\left[\frac{1}{4}(\partial_i N_j + \partial_j N_i)^2 - (\partial_i N_j + \partial_j N_i)(N_i\partial_j\zeta + N_j\partial_i\zeta) - (\partial_i N_i)^2\right], \tag{2.42}$$

and

$$R^{(3)} = -2\frac{e^{-2\zeta}}{a^2}[2\partial^2\zeta + (\partial\zeta)^2]\,. \tag{2.43}$$

Expanding $P(X)$ with $X = \dot{\phi}^2/N^2$ and using the first equality in (2.28) one obtains [51]

$$P(X) = P + P'\left(-2\frac{\dot{\zeta}}{H} + 3\frac{\dot{\zeta}^2}{H^2} - 2\,\delta N_2 - 4\frac{\dot{\zeta}^3}{H^3} + 6\,\delta N_2\frac{\dot{\zeta}}{H}\right)X + \frac{1}{2}P''\left(4\frac{\dot{\zeta}^2}{H^2} + 8\,\delta N_2\frac{\dot{\zeta}}{H} - 12\frac{\dot{\zeta}^3}{H^3}\right)X^2 - \frac{4}{3}P'''\frac{\dot{\zeta}^3}{H^3}X^3\,, \tag{2.44}$$

where, on the right hand side, X, P and its derivatives are evaluated at zeroth order. The lapse has been split as $N = 1 + \delta N_1 + \delta N_2$, where $\delta N_1 = \dot{\zeta}/H$ is the first order contribution, eq. (2.28), while δN_2 is the second order one. As we discussed, terms containing δN_2 will not appear in the action; however they will be important to derive the metric.

Plugging the previous expressions into the action and using the solutions to the constraint equation (2.28) one obtains, after several integrations by parts, the third order action for ζ,

$$S_3 = \int \mathrm{d}t\,\mathrm{d}^3x\,a^3\left[\frac{3\epsilon}{w}\dot{\zeta}^2\zeta - \frac{\epsilon}{3Hw}(2+\frac{1}{w})\dot{\zeta}^3 - \frac{\epsilon}{a^2}(\partial\zeta)^2\zeta - \frac{2}{a^4}\partial_i\zeta\partial_i\psi\partial^2\psi + \frac{1}{2a^4}\left(\partial_i\partial_j\psi\partial_i\partial_j\psi - (\partial^2\psi)^2\right)(3\zeta - \frac{\dot{\zeta}}{H})\right]\,, \tag{2.45}$$

where ψ is given by eq. (2.29),

$$\psi = -\frac{\zeta}{H} + \frac{a^2\epsilon}{w}\partial^{-2}\dot{\zeta}\,. \tag{2.46}$$

This action describes the cosmological scalar perturbations of a perfect fluid with constant equation of state w. Notice that it is just a particular case of the general action derived in refs. [51; 52] in the context of generalized inflationary models, although it describes here a quite different physical system.

The action is simplified by a field redefinition, $\zeta_n = \zeta - f(\zeta)$, where [52]

$$f(\zeta) = \frac{1}{wH}\zeta\dot{\zeta} + \frac{1}{4a^2H^2}\left[-(\partial_i\zeta)^2 + \partial^{-2}\partial_i\partial_j(\partial_i\zeta\partial_j\zeta)\right] + \frac{\epsilon}{2H^2w}\left[\partial_i\zeta\partial_i\partial^{-2}\dot{\zeta} - \partial^{-2}\partial_i\partial_j(\partial_i\zeta\partial_j\partial^{-2}\dot{\zeta})\right] . \quad (2.47)$$

As f is quadratic in ζ, the field redefinition does not change the second order action. Note that the function $f(\zeta)$ contains either spatial gradients or time derivatives of ζ. Thus, for any constant w, the field redefinition (2.47) vanishes on super-Hubble scales, so that in this regime ζ and ζ_n coincide. In terms of ζ_n the action reads

$$S_3 = \int \mathrm{d}t\,\mathrm{d}^3x\, a^3\frac{\epsilon}{w}\left[\frac{2}{3}\left(\frac{1}{w}-1\right)\frac{\dot{\zeta}_n^3}{H} + \frac{3}{2}\left(3-\frac{1}{w}\right)\zeta_n\dot{\zeta}_n^2 + \frac{1}{2a^2}(5+w)\zeta_n(\partial_i\zeta_n)^2 - \left(2-\frac{\epsilon}{2}\right)\frac{\epsilon}{w}\dot{\zeta}_n\partial_i\zeta_n\partial_i\partial^{-2}\dot{\zeta}_n + \frac{\epsilon^2}{4w}\partial^2\zeta_n(\partial_i\partial^{-2}\dot{\zeta}_n)^2\right] . \quad (2.48)$$

We are interested in a dust fluid $w = 0$. The equation of motion at second order derived from this action can be solved perturbatively by plugging the first order solutions $\zeta = \zeta_n = \zeta_0$ into the second order terms. At lowest order in $w \to 0$ one finds

$$\ddot{\zeta}_n + 3H\dot{\zeta}_n = \frac{1}{a^2}F_0 , \quad (2.49)$$

where

$$F_0 = -\zeta_0\partial^2\zeta_0 - \frac{5}{16}(\partial_i\zeta_0)^2 - \frac{3}{8}\partial_i\partial^{-2}(\partial^2\zeta_0\partial_i\zeta_0) . \quad (2.50)$$

The right hand side of eq. (2.49) is negligible at very early times, when all the modes are well out of the Hubble radius, $k/aH \ll 1$, so that $\zeta_n = \text{const}$ is a solution. As in this regime ζ and ζ_n coincide as discussed, this corresponds to the well-known fact that ζ is conserved on super-Hubble scales, even nonlinearly. In this regime the Fourier modes of ζ are decoupled and equal to the initial condition set by inflation, $\zeta = \zeta_0$ for $k/aH \ll 1$. Solving the equation above and going back to ζ using (2.47) yields

$$\zeta = \zeta_0 - \frac{1}{5a^2H^2}\partial^{-2}\partial_i\partial_j(\partial_i\zeta_0\partial_j\zeta_0) . \quad (2.51)$$

This expression gives ζ at second order during matter dominance, as a function of its initial condition ζ_0. We postpone its discussion to the end of this section and we proceed to conclude the calculation of the second order metric: we need to solve for N and N_i at second order.

As at first order, to solve for the whole metric we need to keep terms of order w in the second order evolution equation which now reads

$$\ddot{\zeta}_n + 3H\dot{\zeta}_n - \frac{w}{a^2}\partial^2\zeta_n = \frac{1}{a^2}(F_0 + wF_1) + \frac{w}{a^4H^2}G , \quad (2.52)$$

where

$$F_1 = -5\zeta_0\partial^2\zeta_0 + \frac{1}{8}(\partial_i\zeta_0)^2 + \frac{3}{4}\partial_i\partial^{-2}(\partial^2\zeta_0\partial_i\zeta_0)\ , \tag{2.53}$$

$$G = -\frac{21}{25}(\partial^2\zeta_0)^2 - \frac{2}{5}\zeta_0\partial^4\zeta_0 + \frac{1}{20}\partial_i\zeta_0\partial_i\partial^2\zeta_0 + \frac{9}{200}\partial^2(\partial_i\zeta_0)^2 - \frac{3}{50}\partial_i\partial^{-2}(\partial_i\zeta_0\partial^4\zeta_0)\ . \tag{2.54}$$

To obtain the w corrections to the right-hand side we used the linear evolution of ζ at order w^2:

$$\dot{\zeta} = \frac{2w}{5a^2H}\partial^2\zeta_0 - \frac{6w^2}{25a^2H}\partial^2\zeta_0 + \frac{4w^2}{35a^4H^3}\partial^4\zeta_0 + \mathcal{O}(w^3). \tag{2.55}$$

Solving eq. (2.52) and writing the solution in terms of ζ using the field redefinition (2.47) one finds

$$\begin{aligned}\zeta = \zeta_0 &- \frac{1}{5a^2H^2}\partial^{-2}\partial_i\partial_j(\partial_i\zeta_0\partial_j\zeta_0) \\ &+\frac{w}{5a^2H^2}\Big[2\partial^2\zeta_0 - 4\zeta_0\partial^2\zeta_0 + \frac{14}{5}(\partial_i\zeta_0)^2 + \frac{18}{5}\partial_i\partial^{-2}(\partial^2\zeta_0\partial_i\zeta_0) \\ &\quad - \frac{1}{7a^2H^2}\Big(\frac{27}{20}(\partial^2\zeta_0)^2 + \frac{51}{10}\partial_i\zeta_0\partial_i\partial^2\zeta_0 + \frac{4}{5}(\partial_i\partial_j\zeta_0)^2 \\ &\quad + \frac{1}{2}\partial_i\partial^{-2}(\partial_i\zeta_0\partial^4\zeta_0) + \frac{1}{5}\partial^{-2}\partial_i(\partial_i\partial_j\zeta_0\partial_j\partial^2\zeta_0)\Big)\Big] + \mathcal{O}(w^2)\ .\end{aligned} \tag{2.56}$$

Now that we have a solution for ζ including $\mathcal{O}(w)$ corrections, we can proceed to solve the constraints to get the final second order metric. The momentum constraint is a vector equation; taking its transverse part with the projector $\delta_{ij} - \partial_i\partial_j/\partial^2$ we obtain an equation for the transverse part of the shift vector N_{Ti} ($\partial_i N_{Ti} = 0$),

$$N_{Ti} = -\frac{4}{5}\frac{1}{H}\partial^{-2}\left[\partial_i\zeta_0\partial^2\zeta_0 - \partial^{-2}\partial_i\partial_k(\partial_k\zeta_0\partial^2\zeta_0)\right]\ . \tag{2.57}$$

Notice that the transverse part of the shift vanished at first order, see eq. (2.28). The longitudinal part of the momentum constraint gives an equation for the lapse function at second order,

$$\begin{aligned}\delta N_2 = &\frac{w}{5a^2H^2}\left[(\partial\zeta_0)^2 - 4\zeta_0\partial^2\zeta_0\right] \\ &- \frac{2w}{175a^4H^4}\left[3(\partial^2\zeta_0)^2 + 14\partial_i\zeta_0\partial_i\partial^2\zeta_0 + 4(\partial_i\partial_j\zeta_0)^2\right] + \mathcal{O}(w^2)\ .\end{aligned} \tag{2.58}$$

Although the lapse perturbation vanishes in the limit $w \to 0$, its expression at order w is necessary to solve the energy constraint, similarly to what happened at first order. This gives the second order correction to the shift function

$$\begin{aligned}\psi_2 = \frac{1}{5H}\partial^{-2}\Big[&(\partial_i\zeta_0)^2 - 3\partial^{-2}\partial_i\partial_j(\partial_i\zeta_0\partial_j\zeta_0) \\ &+ \frac{4}{5a^2H^2}\Big(\frac{3}{7}(\partial^2\zeta_0)^2 + \partial_i\zeta_0\partial_i\partial^2\zeta_0 + \frac{4}{7}(\partial_i\partial_j\zeta_0)^2\Big)\Big]\ .\end{aligned} \tag{2.59}$$

Plugging in the ADM metric (2.22) the first order results (2.28), (2.29) and the second order ones (2.56), (2.57), (2.58) and (2.59), we finally obtain the second order metric for $w = 0$,

$$g_{00} = -1 + \frac{4}{25a^2H^2}(\partial_i\zeta_0)^2 \,, \tag{2.60}$$

$$\begin{aligned} g_{0i} = &-\frac{1}{5H}\partial_i\Big[2\zeta_0 - \partial^{-2}(\partial_j\zeta_0)^2 + 3\partial^{-4}\partial_j\partial_k(\partial_j\zeta_0\partial_k\zeta_0) \\ &- \frac{4}{5a^2H^2}\partial^{-2}\Big(\frac{3}{7}(\partial^2\zeta_0)^2 + \partial_i\zeta_0\partial_i\partial^2\zeta_0 + \frac{4}{7}(\partial_i\partial_j\zeta_0)^2\Big)\Big] \\ &-\frac{4}{5}\frac{1}{H}\partial^{-2}\left[\partial_i\zeta_0\partial^2\zeta_0 - \partial^{-2}\partial_i\partial_k(\partial_k\zeta_0\partial^2\zeta_0)\right] \,, \end{aligned} \tag{2.61}$$

$$g_{ij} = a^2\left[1 + 2\zeta_0 + 2\zeta_0^2 - \frac{2}{5a^2H^2}\partial^{-2}\partial_k\partial_l(\partial_k\zeta_0\partial_l\zeta_0)\right]\delta_{ij} + a^2\gamma_{ij} \,. \tag{2.62}$$

Before moving to the calculation of the tensor contribution γ, let us discuss eq. (2.51) which describes the second order evolution of ζ. In Fourier space this equation reads

$$\zeta_{\vec{k}} = \zeta_{0\vec{k}} + \frac{1}{5a^2H^2}\int\frac{d^3q}{(2\pi)^3}\,\frac{(\vec{k}\cdot\vec{q})(k^2 - \vec{k}\cdot\vec{q})}{k^2}\,\zeta_{0\vec{q}}\,\zeta_{0\vec{k}-\vec{q}} \,. \tag{2.63}$$

It is important to stress that the kernel inside the integral does not vanish in the limit $k \to 0$. Surprisingly, this implies that a very long wavelength mode well outside the Hubble radius evolves in the presence of short wavelength perturbations: ζ is *not* conserved on super-Hubble scales as its second order contribution in the matter dominated era grows like the scale factor $a \propto 1/a^2H^2$. However, this is not in contrast with the literature on the conservation of ζ [53], as ζ is conserved when all the modes are out of the Hubble radius, so that the second term on the right hand side of eqs. (2.51) and (2.63) can be neglected. Recently, an interpretation of this result has been given in [63], we will comment on it in section 2.6.

2.4.2 Gravitational waves

So far we did not discuss gravitational waves. At linear order tensor and scalar modes are decoupled, so that gravitational waves can be completely neglected in the limit where the primordial contribution generated by inflation is very small. Beyond the linear approximation scalar and tensor modes mix, so that it is not consistent to set $\gamma = 0$. In the following we are going to assume that the amplitude of primordial tensor modes is very small: gravitational waves will only be generated by the scalar modes through couplings of the form $\gamma\zeta\zeta$.

To study the generation of gravitational waves we need the quadratic action for γ and the cubic terms of the form $\gamma\zeta\zeta$. The constraint equations will not enter neither in the derivation of the action nor in obtaining the expression of the second order metric. Indeed γ does not appear at first order in the constraint equations and quadratic terms are negligible as $\gamma = \mathcal{O}(\zeta^2)$.

Let us start with the quadratic action for γ. Making use of the expressions

$$E_{ij}E^{ij} - E^2 \supset \frac{1}{4}\dot{\gamma}_{ij}\dot{\gamma}_{ij} \tag{2.64}$$

and

$$R^{(3)} \supset \frac{1}{a^2}\left(\gamma_{ij}\partial^2\gamma_{ij} + \frac{3}{4}(\partial_i\gamma_{jk})^2 - \frac{1}{2}\partial_i\gamma_{kj}\partial_k\gamma_{ij}\right) , \tag{2.65}$$

after integration by parts this yields

$$S_{\gamma\gamma} = \frac{1}{8}\int \mathrm{d}t\,\mathrm{d}^3x\,a^3\left(\dot{\gamma}_{ij}\dot{\gamma}_{ij} - \frac{1}{a^2}\partial_k\gamma_{ij}\partial_k\gamma_{ij}\right) . \tag{2.66}$$

For the $\zeta\zeta\gamma$ terms we use the expressions

$$R^{(3)} \supset \frac{4}{a^2}\gamma_{ij}\partial_i\partial_j\zeta - \frac{10}{a^2}\gamma_{ij}\zeta\partial_i\partial_j\zeta \tag{2.67}$$

and

$$e^{3\zeta}N^{-1}a^3\left(E_{ij}E^{ij} - E^2\right) \supset -a\dot{\gamma}_{ij}\left(3\zeta - \frac{\dot{\zeta}}{H}\right)\partial_i\partial_j\psi + \frac{1}{a}\partial_k\gamma_{ij}\partial_i\partial_j\psi\partial_k\psi , \tag{2.68}$$

and we obtain

$$\begin{aligned} S_{\gamma\zeta\zeta} = \int \mathrm{d}t\,\mathrm{d}^3x\,a^3\Big[&-\frac{2}{Ha^2}\gamma_{ij}\partial_i\dot{\zeta}\partial_j\zeta - \frac{1}{a^2}\gamma_{ij}\partial_i\zeta\partial_j\zeta \\ &- \frac{1}{2a^2}\left(3\zeta - \frac{\dot{\zeta}}{H}\right)\dot{\gamma}_{ij}\partial_i\partial_j\psi + \frac{1}{2a^4}\partial_l\gamma_{ij}\partial_i\partial_j\psi\partial_l\psi\Big] . \end{aligned} \tag{2.69}$$

This part of the action is the same as the one derived by Maldacena [4] for a scalar field with standard kinetic term. This is not surprising as these couplings do not depend on $P(X)$ in (2.23).

The equation of motion for γ can be obtained by varying the actions (2.66) and (2.69) with respect to γ_{ij}. Here there are no subtleties in the limit $w \to 0$ so that one can drop terms containing $\dot{\zeta}$. Notice that γ_{ij} is constrained to msatisfy the transverse and traceless conditions $\partial_i\gamma_{ij} = 0$ and $\gamma_{ii} = 0$. As such, the γ's in the action above must be understood as projected by the transverse and traceless projector[1], i.e.,

$$P^{\mathrm{TT}}_{ij\,kl} = \frac{1}{2}\left(P_{ik}P_{jl} + P_{jk}P_{il} - P_{ij}P_{kl}\right) , \tag{2.70}$$

[1] One can also see this in the following way: add to the action for the tensors (2.69) the following two terms $\lambda\gamma^i_i$ and $\lambda_i\partial_j\gamma^{ij}$ where λ and λ_j are Lagrange multipliers that impose the transverse and traceless gauge conditions on γ_{ij}. By carefully working through the equations of motion, and explicitly solving for the Lagrange multipliers, one obtains equation (2.72).

where P_{ij} is a symmetric transverse projector given by

$$P_{ij} \equiv \delta_{ij} - \frac{\partial_i \partial_j}{\partial^2} \,. \tag{2.71}$$

Thus, the evolution equation reads

$$\ddot{\gamma}_{ij} + 3H\dot{\gamma}_{ij} - \frac{\partial^2}{a^2}\gamma_{ij} = P^{\rm TT}_{ij\,kl}\left[\frac{2}{a^2}\partial_k\zeta_0\partial_l\zeta_0 + \frac{4}{25a^4H^2}\partial^2\left(\partial_k\zeta_0\partial_l\zeta_0\right)\right] \,, \tag{2.72}$$

where we have simplified the term in the square braket on the right hand side, taking into account that we are interested only in its transverse component. The solution of this equation, with the appropriate initial condition $\gamma_{ij} = 0$, reads

$$\gamma_{ij} = -\frac{4}{5}\left[9\left(\frac{1}{3} - \frac{j_1(k\tau)}{k\tau}\right)\partial^{-2} + \frac{1}{5a^2H^2}\right] P^{\rm TT}_{ij\,kl}\left(\partial_k\zeta_0\partial_l\zeta_0\right) \,, \tag{2.73}$$

where the spherical Bessel function $j_1(x)$ can be written in terms of trigonometric functions as $j_1(x) = \sin(x)/x^2 - \cos(x)/x$.

2.5 Transforming to Poisson gauge

In this section we want to write our second order metric in conformal Poisson gauge, i.e. in the form

$$\begin{aligned} {\rm d}s^2 = a^2(\tau)\Big\{&-(1+2\Phi_P){\rm d}\tau^2 + 2\omega^\perp_{P\,i}d\tau dx^i \\ &+ \left[(1-2\Psi_P)\delta_{ij} + \gamma_{P\,ij}\right]{\rm d}x^i{\rm d}x^j\Big\} \,, \end{aligned} \tag{2.74}$$

where $\omega^\perp_{P\,i}$ is transverse, $\partial_i\omega^\perp_{P\,i} = 0$, and γ_P is transverse and traceless. Going to this gauge will enable us to compare our results with those obtained in [49; 50].

For the second order gauge transformation we use the notation of [49] (apart from exchanging the name of the scalar potentials Φ and Ψ) and we write our metric as

$$\begin{aligned} {\rm d}s^2 = a^2(\tau)\Big\{&-(1+2\Phi_\zeta){\rm d}\tau^2 + 2(\omega^\perp_{\zeta\,i} + \partial_i\omega_\zeta){\rm d}\tau{\rm d}x^i \\ &+ \left[(1-2\Psi_\zeta)\delta_{ij} + \gamma_{ij}\right]{\rm d}x^i{\rm d}x^j\Big\} \,, \end{aligned} \tag{2.75}$$

where

$$\Phi_\zeta = -\frac{2}{25a^2H^2}(\partial_i\zeta_0)^2 \,, \tag{2.76}$$

$$\begin{aligned} \omega_\zeta = &-\frac{2}{5aH}\zeta_0 + \frac{1}{5aH}\partial^{-2}\left[(\partial_i\zeta_0)^2 - 3\partial^{-2}\partial_i\partial_j(\partial_i\zeta_0\partial_j\zeta_0)\right] \\ &+\frac{4}{25a^3H^3}\partial^{-2}\left[\frac{3}{7}(\partial^2\zeta_0)^2 + \partial_i\zeta_0\partial_i\partial^2\zeta_0 + \frac{4}{7}(\partial_i\partial_j\zeta_0)^2\right] \,, \end{aligned} \tag{2.77}$$

$$\omega^\perp_{\zeta i} = -\frac{4}{5aH}\partial^{-2}\left[\partial_i\zeta_0\partial^2\zeta_0 - \partial^{-2}\partial_i\partial_j(\partial_j\zeta_0\partial^2\zeta_0)\right] \,, \tag{2.78}$$

$$\Psi_\zeta = -\zeta_0 - \zeta_0^2 + \frac{1}{5a^2H^2}\partial^{-2}\partial_i\partial_j(\partial_i\zeta_0\partial_j\zeta_0) \,, \tag{2.79}$$

and γ_{ij} is given by eq. (2.73). The gauge transformation up to second order is, in conformal coordinates $x^\mu = (\tau, x^i)$,

$$x^\mu \to x^\mu - \xi_1^\mu - \xi_2^\mu + \frac{1}{2}\xi_1^\nu \xi_{1,\nu}^\mu \tag{2.80}$$

where at each order $\xi^0 = \alpha$ and $\xi^i = \partial^i \beta + d^i$ with the vector d^i transverse, $\partial_i d_i = 0$. The first order piece is fixed by eq. (2.39), i.e. $\alpha_1 = -2\zeta_0/(5aH)$ and $\beta_1 = 0 = d_{1\,i}$. The parameters of the gauge transformation at second order can be obtained from the second order components of the metric

$$\begin{aligned}\omega_{P\,i}^{\perp} + \partial_i \omega_P = \omega_{\zeta\,i}^{\perp} + \partial_i \omega_\zeta + \frac{\alpha_1}{2}\partial_i \left[2(\alpha_1' + 2aH\alpha_1) - \alpha_1' - 4aH\alpha_1\right] \\ - \frac{1}{2}\alpha_1' \partial_i \alpha_1 - \partial_i \alpha_2 + \partial_i \beta_2' + d_{2\,i}' \,,\end{aligned} \tag{2.81}$$

and

$$\begin{aligned}\gamma_{P\,ij} = \gamma_{ij} + \partial_i \alpha_1 \partial_j \alpha_1 - \frac{1}{3}\delta_{ij}(\partial_k \alpha_1)^2 \\ + 2\left(\partial_i \partial_j - \frac{1}{3}\delta_{ij}\partial^2\right)\beta_2 + \partial_i d_{2\,j} + \partial_j d_{2\,i} \,.\end{aligned} \tag{2.82}$$

In the previous equations we used that $\omega_\zeta = -2\zeta_0/(5aH) = \alpha_1$ at first order. Imposing that γ_P is transverse ($\chi_P^{\parallel} = \chi_P^{\perp} = 0$ in the notation of [49]) allows to solve for β_2 and $d_{2\,i}$:

$$\beta_2 = -\frac{3}{25a^2H^2}\partial^{-2}\left[\partial^{-2}\partial_i\partial_j(\partial_i\zeta_0\partial_j\zeta_0) - \frac{1}{3}(\partial_i\zeta_0)^2\right] \,, \tag{2.83}$$

$$d_{2\,i} = -\frac{4}{25a^2H^2}\partial^{-2}\left[\partial^2\zeta_0\partial_i\zeta_0 - \partial^{-2}\partial_i\partial_j(\partial^2\zeta_0\partial_j\zeta_0)\right] \,. \tag{2.84}$$

Imposing $\omega_P = 0$ in eq. (2.81) gives

$$\alpha_1 + \alpha_2 = \omega_\zeta + \beta_2' \,. \tag{2.85}$$

At second order, the other components of the metric transform as [49]

$$\Phi_P = \Phi_\zeta + \frac{\alpha_1}{2}\left(\alpha_1'' + 5aH\alpha_1' + \frac{3}{2}a^2H^2\alpha_1\right) + \alpha_1'^2 + \alpha_2' + aH\alpha_2 \,, \tag{2.86}$$

$$\begin{aligned}\Psi_P = \Psi_\zeta + \frac{\alpha_1}{2}\left[-\frac{3}{2}a^2H^2\alpha_1 - aH\alpha_1' + 2\left(\Psi_\zeta' + 2aH\Psi_\zeta\right)\right] \\ - \frac{1}{6}(\partial_i\alpha_1)^2 - aH\alpha_2 - \frac{1}{3}\partial^2\beta_2 \,,\end{aligned} \tag{2.87}$$

which, with the conditions (2.83), (2.84) and (2.85) above, give

$$\Phi_P = -\frac{3}{5}\zeta_0 + \frac{9}{25}\left[\zeta_0^2 + \partial^{-2}(\partial_j\zeta_0)^2 - 3\partial^{-4}\partial_i\partial_j(\partial_i\zeta_0\partial_j\zeta_0)\right] + \frac{6}{175a^2H^2}\partial^{-2}\left[2(\partial_i\partial_j\zeta_0)^2 + 5(\partial^2\zeta_0)^2 + 7\partial_i\zeta_0\partial_i\partial^2\zeta_0\right] , \quad (2.88)$$

$$\Psi_P = -\frac{3}{5}\zeta_0 - \frac{9}{25}\left[\zeta_0^2 + \frac{2}{3}\partial^{-2}(\partial_i\zeta_0)^2 - 2\partial^{-4}\partial_i\partial_j(\partial_i\zeta_0\partial_j\zeta_0)\right] + \frac{6}{175a^2H^2}\partial^{-2}\left[2(\partial_i\partial_j\zeta_0)^2 + 5(\partial^2\zeta_0)^2 + 7\partial_i\zeta_0\partial_i\partial^2\zeta_0\right] , \quad (2.89)$$

$$\omega^{\perp}_{P\,i} = -\frac{24}{25aH}\partial^{-2}\left[\partial^2\zeta_0\partial_i\zeta_0 - \partial^{-2}\partial_i\partial_j(\partial^2\zeta_0\partial_j\zeta_0)\right] , \quad (2.90)$$

$$\gamma_{P\,ij} = -\frac{36}{5}\left(\frac{1}{3} - \frac{j_1(k\tau)}{k\tau}\right)\partial^{-2}P^{\rm TT}_{ij\,kl}\left(\partial_k\zeta_0\partial_l\zeta_0\right) . \quad (2.91)$$

The transverse traceless projector above can be expanded to give

$$P^{\rm TT}_{ij\,kl}\left(\partial_k\zeta_0\partial_l\zeta_0\right) = -\partial^{-2}\left[\partial^2\Theta_0\delta_{ij} + \partial_i\partial_j\Theta_0 + 2(\partial^2\zeta_0\partial_i\partial_j\zeta_0 - \partial_i\partial_k\zeta_0\partial_j\partial_k\zeta_0)\right] , \quad (2.92)$$

with

$$\Theta_0 = -\frac{1}{2}\partial^{-2}\left[(\partial^2\zeta_0)^2 - (\partial_i\partial_j\zeta_0)^2\right] . \quad (2.93)$$

This result can be compared with ref. [50] by taking into account that $\zeta_0 = -(5/3)\Phi$ at linear order and on super-Hubble scales. The resulting metrics coincide up to a typo in eq. (A.29) of [50].[1] Notice that in order to perform the calculation in Poisson gauge, one has to supplement the second order equations obtained in [49] by the proper second order matching with the initial condition provided by ζ [50]. In our calculation this matching is already taken into account by the fact that we are always using the conserved quantity ζ.

2.6 Comments on the apparent non-conservation of ζ

We saw in section 2.4 that the second-order solution for the scalar metric fluctuation in comoving gauge ζ can be written down as

$$\zeta'_{\vec{k}} = -\frac{4}{5aH}\int\frac{d^3q}{(2\pi)^3}\,\frac{(\vec{k}\cdot\vec{q})(k^2 - \vec{k}\cdot\vec{q})}{k^2}\,\zeta_{0\vec{q}}\,\zeta_{0\vec{k}-\vec{q}} , \quad (2.94)$$

[1] To obtain the correct result the right hand side of eq. (A.29) of [50] should be multiplied by a factor -2. Note that in their notation a factor of $1/2$ multiplies all the second order contributions, i.e., $\omega_i \equiv \omega_{1i} + \frac{1}{2}\omega_{2i}$ while in our notation there is no factor of $1/2$.

where the second-order piece evolves in time and does not go to zero as $k \ll aH$. Indeed, it goes like

$$\zeta'_{k\ll aH} \approx -\frac{4}{5aH}\int \frac{d^3q}{(2\pi)^3}\,\frac{(\vec{k}\cdot\vec{q})(\vec{k}\cdot\vec{q})}{k^2}\,\zeta_{0\vec{q}}\,\zeta_{0\vec{q}} \tag{2.95}$$

$$= -\frac{4}{15aH}\int \frac{dq}{(2\pi^2)}\,q^4\zeta_{0\vec{q}}\,\zeta_{0\vec{q}}\,. \tag{2.96}$$

This means that modes of ζ inside of the horizon source a variation of modes outside of the horizon. As we pointed out, this does not contradict the literature, which states that ζ is conserved if one considders all modes to be outside of the horizon. This can be worrisome since something similar can in principle happend during radiation domination, and one would have to take into account the unknown thermal history of the very early universe in order to compute the ζ beyond the linear order.

A recent reinterpretation of this phenomenon was proposed in [63]. There, it is argued that "not all gauges are created equal" as we will now explain. When writing down the perturbed FLRW metric one is implicitly averaging over small-scale modes. One may write for the metric

$$g_{\mu\nu} = \bar{g}_{\mu\nu} + H_{\mu\nu} + h_{\mu\nu}\,, \tag{2.97}$$

where $\bar{g}$ is the FLRW metric, and we split the metric perturbations in a long mode ($k \ll \Lambda$) piece H and a short mode ($k \gg \Lambda$) part h, where Λ is the scale at which averaging takes place. In order for this framework to be consistent, *i.e.* in order for the large-scale perturbed FLRW metric to be correctly defined, one must be able to define for example the inverse metric at large scales:

$$g^{\mu\nu}|_{k\ll\Lambda} = \bar{g}^{\mu\nu} - H^{\mu\nu} + h^{\mu\alpha}h^{\nu}{}_{\alpha}|_{k\ll\Lambda}\,, \tag{2.98}$$

where the leading term in h vanishes by definition in the limit of large scales. The quadratic term in h will in general be a convolution of modes that may be non-zero. So in order for (2.98) to be the inverse of (2.97) in the large-scale limit the following condition must be satisfied

$$H^{(2)}_{\mu\nu} \gg (h_{\mu\alpha}h^{\alpha}_{\nu})|_{k\ll\Lambda}\,. \tag{2.99}$$

The point is that not all gauges satisfy this condition, as we will soon see.

Equation (2.79) shows that the metric potential Ψ defined in equation (2.74) in the comoving gauge has the form $\Psi^{(1)}_{\zeta} \sim \zeta$ and, in order to check whether the condition (2.99) is satisfied, we write the schematic form for the large-scale second order piece [1] $\Psi^{(2)}_{\zeta} \sim \zeta^2 + \frac{\Lambda^2}{a^2H^2}\zeta^2$ in the limit $k \ll \Lambda$. This means that the metric is not well defined if one averages over scales comparable with the horizon since in this case the condition (2.99) is not satisfied: $\Psi^{(2)}_{\zeta}|_{k\ll\Lambda} \sim \Psi^{(1)\,2}_{\zeta}|_{k\ll\Lambda}$.

[1] To do this we did all integrations in the moduli of momenta from 0 to Λ since, if the metric is consistently defined, one should be able to write it down only in terms of the long-scale modes. After that, we imposed the condition $\Lambda \gg k$ in order to obtain the final schematic form.

Let us now study what happens for the Poisson gauge. As can be seen from equation (2.89), the form of the first-order piece for the metric potentials is $\Psi_P^{(1)} \sim \zeta$ and the second order piece can be written schematically as $\Psi_P^{(2)} \sim \zeta^2 + \frac{\Lambda^2}{k^2}\zeta^2 + \frac{\Lambda^2}{a^2H^2}\zeta^2$. Thus, the condition (2.99) expressed as $\Psi_P^{(2)}|_{k\ll\Lambda} \gg \Psi_P^{(1)\,2}|_{k\ll\Lambda}$ is always satisfied.

Using the fact that the Poisson gauge is always well defined, the authors of [63] argue that a correct way to define the non-linear equivalent of ζ is to use the linear relation between the spatial curvature perturbation in comoving gauge ζ and the metric components in Poisson gauge Φ_P and Ψ_P during matter domination:

$$\zeta_\ell \equiv \Phi_P + \frac{2}{3}\frac{\dot{\Phi}_P + aH\Psi_P}{a^H}\,. \tag{2.100}$$

With this definition one can show that this variable also evolves outside of the horizon as [63]

$$\zeta'_\ell|_{\Lambda\ll(aH)^{-1}} \approx \frac{aH}{3\bar{\rho}}[\tau^k_k] \approx \frac{5}{3aH}[\partial_i\zeta_0\partial_i\zeta_0]_\Lambda\,, \tag{2.101}$$

where the square brackets denote averaging with a characteristic scale $\Lambda \gg (aH)^{-1}$, and τ^k_k is an effective stress-energy tensor generated by small wavelength modes and matter, see [63]. For comparison we used the explicit form of τ for averaging scales much larger than the horizon. You can notice that ζ_ℓ is again not conserved outside of the horizon. Moreover, it is easy to check that in superhorizon scales it is different from the comoving gauge ζ we used in this article, equation (2.96), by taking its Fourier transform

$$\zeta'_\ell|_{k\ll aH} \approx \frac{5}{3aH}\int\frac{dq}{(2\pi^2)}\,q^4\zeta_{0\vec{q}}\,\zeta_{0\vec{q}}\,. \tag{2.102}$$

However, the right hand side of equation (2.101) is simply proportional the effective pressure generated by the short wavelength modes, and it appears in our equations because we have not yet renormalized the background to take it into account. Once the background is renormalized, the resulting ζ_ℓ will be conserved.

We believe that the correct question to ask should be formulated in terms of observable quantities for which it should be clear that there is no paradox. In chapter 3, we will compute the connection between temperature fluctuations in the CMB and initial conditions using the Poisson gauge metric. We will see there, in equation (3.32), that the large scale expression for the temperature fluctuation does not diverge due to integrals of second order in ζ_0, and indeed it goes to a constant at large scales, being thus well defined. The fact that the calculation of the temperature fluctuations performed in Poisson gauge does not introduce further paradoxes might lead one to believe that ζ_ℓ, which is defined in that gauge, is the correct variable to consider since it is furthermore consistently defined and conserved outside of the horizon. However, a more precise analysis of the connection between ζ_ℓ and observable quantities is beyond the scope of this thesis.

2.7 Discussion

In this chapter we have discussed a new approach to nonlinear cosmological perturbations, based on the equivalence between cosmological perfect fluids and derivatively coupled scalar fields. In this approach, perturbations can be studied using an ADM perturbative expansion of the scalar field action, similarly to what is done for the study of non-Gaussianities during inflation.

Using this method, we have calculated the second order metric during matter domination in the comoving gauge (2.25), where the scalar fluctuations are described by the comoving curvature perturbation ζ. The metric, which is the main result of the chapter, is given by

$$
\begin{aligned}
g_{00} &= -1 + \frac{4}{25a^2H^2}(\partial_i\zeta_0)^2 \,, && (2.103)\\
g_{0i} &= -\frac{1}{5H}\partial_i\Big[2\zeta_0 - \partial^{-2}(\partial_j\zeta_0)^2 + 3\partial^{-4}\partial_j\partial_k(\partial_j\zeta_0\partial_k\zeta_0) \\
&\qquad - \frac{4}{5a^2H^2}\partial^{-2}\Big(\frac{3}{7}(\partial^2\zeta_0)^2 + \partial_i\zeta_0\partial_i\partial^2\zeta_0 + \frac{4}{7}(\partial_i\partial_j\zeta_0)^2\Big)\Big] \\
&\quad -\frac{4}{5}\frac{1}{H}\partial^{-2}\left[\partial_i\zeta_0\partial^2\zeta_0 - \partial^{-2}\partial_i\partial_j(\partial_j\zeta_0\partial^2\zeta_0)\right] \,, && (2.104)\\
g_{ij} &= a^2\exp[2\zeta(t)]\delta_{ij} + a^2\gamma_{ij} \,, && (2.105)
\end{aligned}
$$

where

$$
\zeta(t) = \zeta_0 - \frac{1}{5a^2H^2}\partial^{-2}\partial_k\partial_l(\partial_k\zeta_0\partial_l\zeta_0) \,, \qquad (2.106)
$$

ζ_0 is the perturbation generated during inflation, conserved on super-Hubble scales, and where γ_{ij} is a traceless transverse tensor, given by

$$
\gamma_{ij} = -\frac{4}{5}\left[9\left(\frac{1}{3} - \frac{j_1(k\tau)}{k\tau}\right)\partial^{-2} + \frac{1}{5a^2H^2}\right] P^{\mathrm{TT}}_{ij\,kl}\,(\partial_k\zeta_0\partial_l\zeta_0) \,. \qquad (2.107)
$$

Note that at early times, when all the modes are out of the Hubble radius so that we can neglect gradients, the above metric is just

$$
ds^2 = -dt^2 + a^2(t)e^{2\zeta_0(\vec{x})}d\vec{x}^2 \,. \qquad (2.108)
$$

This is the form of the metric in comoving gauge after all the modes have left the Hubble radius during inflation. This metric is known to remain the same during the subsequent evolution of the Universe until the modes re-enter the Hubble radius. This shows that our metric nicely matches the initial conditions set by inflation.

As shown by eq. (2.106) above, we have found that ζ is not conserved at second order during the matter dominated era. Indeed, super-Hubble modes can evolve under the presence of short sub-Hubble perturbations. We commented on a reinterpretation of this result as suggested in [63]. They point out that "not all gauges are created equal" in the sense that one should chose a gauge in which the metric at large scales is consistently defined when averaging over small scales. While the comoving gauge does

not satisfy this condition the Poisson gauge does, and one can define an equivalent of ζ by using the linear relations between the potentials in the Poisson gauge and in comoving gauge, then averaging this quantity with a filter of the size of the horizon, and taking into account the renormalization of the background induced by small-scale fluctuations. Such a quantity is consistently defined even when taking into account non-linear structures at small scales, and is furthermore not expected to evolve outside the horizon. However, we have concluded that a discussion on the physical significance of this phenomena should be done by using observable quantities, and we hinted at how this might solve the apparent paradox (*i.e.* there is no paradox when considering the temperature fluctuations at the CMB), but a rigorous discussion is left for future work.

Our method can be generalized to compute the metric during a cosmological era dominated by a fluid with constant equation of state $w \neq 0$, like during radiation dominance, by choosing the appropriate w in the linear evolution equation for ζ, eq. (2.31), and in the third order action (2.48). The only difference is that in this case ζ is not constant inside the Hubble (see eq. (2.31)), so that the second order solution will involve an integral over conformal time. Furthermore, one can generalize our method to study the evolution of perturbations in a Universe filled with several coupled fluids, by introducing a correspondent number of derivatively coupled scalar fields in the action, with possible couplings between them. As a future study, this method could be used to study the generation of gravitational waves from second order perturbations during radiation and matter dominated era [64; 65].

In chapter 3 we will use the metric computed here in order to compute the equivalent of the Sachs-Wolfe effect at second-order, thus giving an estimate for the contribution of gravitational non-linearities to the non-Gaussianity of the CMB.

3

Sachs-Wolfe at second order: the CMB bispectrum on large angular scales

3.1 Introduction

The linear approximation to cosmological perturbations has been so far sufficient and extremely fruitful, at least on large scales, before non-linearities induced by gravity become significant. However, the accuracy of observations is now reaching a level such that all second-order effects, naively of magnitude $\sim (10^{-5})^2$, may become relevant. This is particularly important in the context of primordial non-Gaussianities: second-order effects are in fact expected to give a signal of order $f_{\rm NL} \sim$ few, which is not far from the present experimental limits [11; 44; 66]. A large amount of work has been done to study Cosmic Microwave Background (CMB) fluctuations beyond the linear approximation, in order to make predictions for the temperature bispectrum. As a complete calculation of the bispectrum is a daunting task, people concentrated on specific effects which are expected to dominate in particular limits. The bispectrum generated by the correlation between lensing and the Integrated Sachs-Wolfe (ISW) effect has been studied in [67; 68]. The one coming from lensing and the Sunyaev-Zel'dovich effect has been studied in [68]. In [38; 39; 69] the bispectrum generated by perturbations in the recombination history has been calculated. Refs [40; 41; 70] (see also [71]) focused on very short angular scales where the signal is dominated by the non-linearity induced by dark matter clustering. A systematic control of all second-order effects in the Boltzmann equations is currently under study: see [41; 42; 72] and references therein.

In this chapter we calculate the CMB bispectrum in the limit of large angles, i.e. on angular scales larger than the one subtended by the Hubble radius at recombination ($\theta \gtrsim 1^\circ$); we do this assuming perfect matter dominance. Important, although, as we

will see, partial results were obtained in this regime in [50; 73; 74].

Our calculation can be seen as the extension to second order of the classic Sachs-Wolfe formula [48]

$$\frac{\delta T}{T} = \frac{\Phi_e}{3}\,, \tag{3.1}$$

where Φ_e is the Newtonian potential at recombination, which gives the large-angle prediction for the spectrum of the CMB fluctuations. As it is well known, this formula describes the angular variation of the temperature without considering the dynamics of the photon/baryon plasma, but only the gravitational redshift of photons from the last scattering to us. Therefore, it describes correctly the CMB anisotropies only in the limit where the scales considered are well out of the Hubble radius at recombination: the same restriction will apply to our calculation. The Sachs-Wolfe formula further assumes that decoupling took place when the universe was matter dominated – neglecting the transition between radiation and matter domination – and that the universe is still matter dominated nowadays, neglecting the present acceleration. At linear order this simplification is very convenient as the gravitational potential stays constant during matter dominance. At second-order the gravitational potential is no longer constant but the second-order metric during matter dominance is known [49] and can be written analytically as a function of the large-scale inflationary perturbations [1; 50].

Clearly, these approximations do not hold in the real universe. However, our calculations give the exact bispectrum in the same limit in which the Sachs-Wolfe formula becomes exact: zero cosmological constant, recombination that happens much after equality and in the limit in which all scales are much larger than the horizon at recombination. This last limit can be imagined by thinking about an experimentalist making measurements in the far future, when the angle subtended by the Hubble radius at recombination is minuscule. The fact that our results become exact in a well defined physical limit is quite important, as on large angular scales the separation among different effects is in general gauge dependent. Therefore, one has to be careful in making approximations because neglecting some effects leads, in general, to a gauge dependent result. Besides its theoretical interest, we expect our result to represent a fair approximation to the real universe on large angular scales and it can be taken as a starting point for more elaborate calculations.

Motivated by inflation, we assume that there are no vector or tensor perturbations in the initial conditions on super-Hubble scales. We perform the calculation of the CMB anisotropies by integrating the photon geodesic equation during matter dominance using the so called generalized Poisson gauge, which generalizes at second order the standard Newtonian gauge. Besides the Newtonian and curvature potential, at second order new terms are present in the metric, generated by the product of linear fluctuations: a vector mode in the $\mathrm{d}x^i\mathrm{d}t$ entry of the metric, and a tensor mode in the spatial part.

All these terms contribute to the final CMB anisotropy. The time independent parts of the gravitational potentials give rise to second-order terms evaluated at last scattering, in analogy with eq. (3.1); their contribution was calculated in [73]. However,

at second-order there are also terms integrated along the photon trajectory, similarly to what happens at first order when we depart from matter dominance with the ISW effect. The time-evolution at second order of the gravitational potential on sub-Hubble scales generates the well-known Rees-Sciama effect [75; 76]. But also the vector and tensor part of the metric contribute with two integrated terms.[1] All these terms contain a number of spatial gradients higher than the intrinsic terms, so that one may think that they can be neglected on large scales as suppressed by positive powers of $k/(aH)$ at recombination. However, this conclusion is too hasty: these terms are integrated along the photon trajectory while modes progressively reenter the Hubble radius. Thus the ratio $k/(aH)$ should not be evaluated at recombination but when the terms contribute to the time integral. We will see that all the integrated pieces give a contribution of the same order as the intrinsic terms in the equilateral limit. Actually the separation between intrinsic and integrated effects has no physical, gauge invariant meaning. For example, a part of the integrated vector contribution will turn out to be a boundary term.

Another integrated contribution is gravitational lensing, due to the gravitational deflection of the photon trajectory with respect to the line of observation. Although the effect of lensing on the bispectrum through its correlation with the ISW effect is well known [67; 68] (but absent in our calculation as we are assuming perfect matter dominance), we will see that lensing is important also when correlated with intrinsic contributions at last scattering. In particular, we will find that lensing gives a squeezed limit contribution of the same order as the one due to intrinsic effects, but which depends on the angle between the long and the short modes. The effect of lensing on the bispectrum was studied in [77] with the conclusion that its effect is suppressed in the squeezed limit by the tilt of the spectrum. We will see that this conclusion is not correct.

In computing the CMB bispectrum we will employ the flat-sky approximation, which is valid for small angles of view. Given that at the same time we are interested in angles which are much larger than the Hubble radius at recombination, there is a quite narrow range of scales, $1 \ll l \ll l_{\text{1st peak}}$, where our approximations hold. However, the flat-sky approximation greatly simplifies the algebra and makes the result much more transparent. The results will be given by 2-dimensional kernels $B(\vec{l}_1, \vec{l}_2, \vec{l}_3)$, which can be thought of as the 2d observable analogue of the kernels used (in 3 dimensions) to describe the shape of the primordial non-Gaussianity [34].

This chapter is organized as follows. In the next section we give the second-order metric in matter dominance in the generalized Poisson gauge as a function of the inflationary initial conditions and we calculate the temperature anisotropy at second order integrating the photon geodesic. In section 3.3 we make a general discussion about the bispectrum of the temperature anisotropy in the flat-sky approximation and we calculate this quantity induced by a primordial non-Gaussianity of the local and equilateral kind. These are useful for comparison with our results. In section

[1] The integrated tensor contribution has been taken into account for the large scale anisotropies in [50]. The vector contribution has, to our knowledge, always been ignored.

3.4 we calculate the bispectrum using the results of section 3.2. The calculation is split (for convenience, not because the effects are physically distinguishable) in various pieces: intrinsic effects at last scattering, integrated vector contribution, integrated tensor contribution and lensing. The resulting total bispectrum is discussed in section 3.5. An argument for the form of the bispectrum in the squeezed limit is given in section 3.6 and conclusions are drawn in section 3.7. The flat-sky approximation is discussed in appendix 3.A, while the details of the calculation of the Rees-Sciama effect are presented in appendix 3.B.

3.2 Second-order temperature anisotropies

In this section we calculate the CMB temperature anisotropy at second order in perturbations, in the large angular scale limit and for matter dominance, as a function of the angle of observation. On large angular scales, the effect of second-order perturbations on the CMB fluctuations have been studied more generally in [78; 79]. Although we will later use the flat-sky approximation, the results of this section hold also in a full-sky treatment.

We are interested in the CMB temperature fluctuations,

$$\frac{\delta T}{T}(\hat{n}) \equiv \frac{T_o(\hat{n}) - \bar{T}_o}{\bar{T}_o} , \tag{3.2}$$

where $T_o(\hat{n})$ is the observed photon temperature in the angular direction $\hat{n}$ ($\hat{n}^2 = 1$) and $\bar{T}_o$ is its average over the sky. For a black-body spectrum the observed temperature $T_o(\hat{n})$ is related to the one of emission $T_e(\vec{x}_e)$ by Liouville's theorem: as phase space density is conserved in the propagation of photons (assuming there is no further scattering), the phase space density received in a given direction $\hat{n}$ is the same as the one at emission but with a temperature [48; 80]

$$T_o(\hat{n}) = \frac{\omega_o}{\omega_e} T_e(\vec{x}_e) , \tag{3.3}$$

where ω_e and ω_o are the frequencies at emission and observation of a given photon. Notice that this statement is exact and therefore holds at any order in perturbation theory. In general, also the temperature at emission will not be isotropic, but will depend on the angle of emission. This dependence can be however neglected in our case, as we are interested in perturbations which are much longer than the horizon at recombination.

We work in the so called generalized Poisson gauge and use conformal time τ. In this gauge, the metric reads [49]

$$\mathrm{d}s^2 = a^2(\tau) \left\{ -(1+2\Phi)\mathrm{d}\tau^2 + 2\omega_i \mathrm{d}x^i \mathrm{d}\tau + [(1-2\Psi)\delta_{ij} + \gamma_{ij}] \mathrm{d}x^i \mathrm{d}x^j \right\} , \tag{3.4}$$

where ω_i is transverse, $\omega_{i,i} = 0$, and γ_{ij} is transverse and traceless, $\gamma_{ij,i} = 0 = \gamma_{ii}$. In the matter dominated era, assuming that the amount of primordial gravitational waves

is negligible, the components of this metric are [1; 49; 50]

$$\Phi = \phi + \left[\phi^2 + \partial^{-2}(\partial_j\phi)^2 - 3\partial^{-4}\partial_i\partial_j(\partial_i\phi\partial_j\phi)\right] + \frac{2}{21a^2H^2}\partial^{-2}\left[2(\partial_i\partial_j\phi)^2 + 5(\partial^2\phi)^2 + 7\partial_i\phi\partial_i\partial^2\phi\right] , \tag{3.5}$$

$$\Psi = \phi - \left[\phi^2 + \frac{2}{3}\partial^{-2}(\partial_i\phi)^2 - 2\partial^{-4}\partial_i\partial_j(\partial_i\phi\partial_j\phi)\right] + \frac{2}{21a^2H^2}\partial^{-2}\left[2(\partial_i\partial_j\phi)^2 + 5(\partial^2\phi)^2 + 7\partial_i\phi\partial_i\partial^2\phi\right] , \tag{3.6}$$

$$\omega_i = -\frac{8}{3aH}\partial^{-2}\left[\partial^2\phi\partial_i\phi - \partial^{-2}\partial_i\partial_j(\partial^2\phi\partial_j\phi)\right] , \tag{3.7}$$

$$\gamma_{ij} = -20\left(\frac{1}{3} - \frac{j_1(k\tau)}{k\tau}\right)\partial^{-2}P^{\mathrm{TT}}_{ij\,kl}\left(\partial_k\phi\partial_l\phi\right) . \tag{3.8}$$

The scalar quantities Φ and Ψ are the Newtonian and curvature potentials, respectively, while we will refer to ω_i and γ_{ij} as the vector and tensor components of the metric. The metric is expressed in terms of ϕ, the time-independent quantity representing the initial curvature perturbation generated during inflation. Indeed, ϕ is simply proportional to the (non-linear) curvature perturbation on uniform density hypersurfaces ζ: on super-Hubble scales, where ζ is constant,

$$\phi = -\frac{3}{5}\zeta \qquad (k \ll aH) . \tag{3.9}$$

In the following we are going to assume that ζ on large scales, and therefore ϕ, is perfectly Gaussian, which is a very good approximation for example in minimal single field inflationary models [4; 27]. In the expression for tensor modes, the spherical Bessel function $j_1(x)$ is given by $j_1(x) = \sin(x)/x^2 - \cos(x)/x$, while $P^{\mathrm{TT}}_{ij\,kl}$ is a transverse traceless projector defined as

$$P^{\mathrm{TT}}_{ij\,kl} \equiv \frac{1}{2}\left(P_{ik}P_{jl} + P_{jk}P_{il} - P_{ij}P_{kl}\right) , \tag{3.10}$$

where P_{ij} is a symmetric transverse projector given by

$$P_{ij} \equiv \delta_{ij} - \frac{\partial_i\partial_j}{\partial^2} . \tag{3.11}$$

It can be expanded to give

$$P^{\mathrm{TT}}_{ij\,kl}\left(\partial_k\phi\partial_l\phi\right) = -\partial^{-2}\left[\partial^2\Theta_0\delta_{ij} + \partial_i\partial_j\Theta_0 + 2(\partial^2\phi\partial_i\partial_j\phi - \partial_i\partial_k\phi\partial_j\partial_k\phi)\right] , \tag{3.12}$$

with

$$\Theta_0 = -\frac{1}{2}\partial^{-2}\left[(\partial^2\phi)^2 - (\partial_i\partial_j\phi)^2\right] . \tag{3.13}$$

In order to study the photon redshift we must solve the photon geodesic equation from last scattering to us, taking into account the perturbations of the metric above. The photon geodesic equation can be written as

$$\frac{\mathrm{d}P_\mu}{\mathrm{d}\lambda} = \frac{1}{2}\partial_\mu g_{\alpha\beta}P^\alpha P^\beta , \tag{3.14}$$

where $P^\mu = dx^\mu/d\lambda$ is the four-momentum of the photon, $P^\mu P_\mu = 0$. The frequency of a photon with four-momentum P^μ as measured by an observer with four-velocity u^μ, is given by $\omega = -P_\mu u^\mu$. For simplicity, we choose the observer today to have zero spatial velocity, $u^i_o = 0$. Indeed, any peculiar motion of the observer leads to a dipole anisotropy that can easily be subtracted. Furthermore, since we are interested in the large angular scales, we neglect also the Doppler effect due to the velocity of the photon/baryon fluid at recombination, which vanishes on super-Hubble scales at decoupling. Thus, we choose also the emitter to have zero spatial velocity, $u^i_e = 0$, so that we have $\omega = -P_0 u^0$ both for the observer and the emitter. Making use of the normalization condition of the four-velocity, $u^\mu u_\mu = -1$, one obtains $\omega = -P_0/\sqrt{-g_{00}}$, and thus

$$\frac{\omega_o}{\omega_e} = \frac{P_0(\tau_o)}{P_0(\tau_e)} \frac{\sqrt{-g_{00}}|_e}{\sqrt{-g_{00}}|_o} \,. \tag{3.15}$$

In order to compute P_0 we need to solve the time component of eq. (3.14). Using that $P^0 = \mathrm{d}\tau/\mathrm{d}\lambda$ and plugging the metric (3.4) into eq. (3.14) yields

$$P^0 \frac{\mathrm{d}P_0}{\mathrm{d}\tau} = \mathcal{H} g_{\alpha\beta} P^\alpha P^\beta - a^2 \Phi' (P^0)^2 + a^2 \omega_i' P^0 P^i + a^2 \Big(- \Psi' \delta_{ij} + \frac{1}{2} \gamma_{ij}' \Big) P^i P^j \,, \tag{3.16}$$

where by a prime we denote the partial derivative with respect to conformal time, $' \equiv \partial/\partial\tau$, and $\mathcal{H}$ is the conformal Hubble rate, $\mathcal{H} \equiv a'/a$. One can immediately notice that the first term on the right hand side vanishes because of the massless condition $P^\mu P_\mu = 0$. Note also that, as we are studying perfect matter dominance, the two potentials Φ and Ψ are constant at linear order, see eqs. (3.5) and (3.6). Thus, their time derivatives Φ', Ψ', together with ω_i' and γ_{ij}', are all second-order quantities. One can therefore replace the zeroth-order expression $P^i = -P^0 \hat{n}^i$ into this equation; furthermore, using the background relation $P^0 = -P_0/a^2$, the geodesic equation can be finally rewritten as

$$\frac{1}{P_0} \frac{\mathrm{d}P_0}{\mathrm{d}\tau} = \Phi' + \Psi' + \omega_i' \hat{n}^i - \frac{1}{2} \gamma_{ij}' \hat{n}^i \hat{n}^j, \tag{3.17}$$

that upon integration yields

$$\frac{P_0(\tau_o)}{P_0(\tau_e)} = 1 + \int_{\tau_e}^{\tau_o} \mathrm{d}\tau \left(\Phi' + \Psi' + \omega_i' \hat{n}^i - \frac{1}{2} \gamma_{ij}' \hat{n}^i \hat{n}^j \right). \tag{3.18}$$

Plugging this expression into eq. (3.15), one obtains the photon redshift up to second-order as a function of the metric perturbations,

$$\frac{\omega_o}{\omega_e} = \frac{a_e}{a_o} \sqrt{\frac{1 + 2\Phi_e}{1 + 2\Phi_o}} \left[1 + \int_{\tau_e}^{\tau_o} \mathrm{d}\tau \left(\Phi' + \Psi' + \omega_i' \hat{n}^i - \frac{1}{2} \gamma_{ij}' \hat{n}^i \hat{n}^j \right) \right] . \tag{3.19}$$

Now we need to relate $T_e(\vec{x}_e)$ on the right hand side of eq. (3.3) to the metric perturbations at decoupling. Since we concentrate on large angular scales, we only

need the super-Hubble relation. We will use adiabatic initial conditions. In this case the dark matter energy density ρ_m simply scales as the third power of the temperature,

$$\rho_m \propto T_e^3 \,. \tag{3.20}$$

In the matter dominated era, the energy density of dark matter is related to the metric perturbations through the Einstein equations, in particular through the energy constraint equation. On super-Hubble scales, i.e. neglecting spatial gradients, and using the fact that the potentials Φ and Ψ at first order are time-independent in the matter dominated era this reads, up to second order, (see for instance eq. (196) of [81])

$$3H^2\left(1-2\Phi_e+4\Phi_e^2\right)=8\pi G\rho_m \,, \tag{3.21}$$

where H is the Hubble rate. Using the background Friedmann equation and eq. (3.20) above, this equation can be rewritten as

$$T_e=\left(1-2\Phi_e+4\Phi_e^2\right)^{1/3}\bar{T}_e \,, \tag{3.22}$$

where $\bar{T}_e$ is the average temperature at emission, which simply scales as the inverse of the background scale factor.

This equation can be derived in a simpler way [74] taking into account that, at recombination, all the modes that we are considering are much longer than the horizon and adiabatic. This means that each local observer will see a completely unperturbed history at any order in perturbations. Indeed, the vector and tensor components of the metric, eqs. (3.7) and (3.8), are suppressed by powers of $k/(aH)$ and can be neglected at recombination.[1] The same holds for the time dependent part of Φ and Ψ, i.e. the second lines of eq. (3.5) and (3.6). This means that the metric on large scales takes the form

$$\mathrm{d}s^2=a^2(\tau)\left\{-(1+2\Phi)\mathrm{d}\tau^2+(1-2\Psi)\delta_{ij}\mathrm{d}x^i\mathrm{d}x^j\right\} \,, \qquad (k\ll aH) \,, \tag{3.23}$$

where Φ and Ψ are now time independent and slowly varying in space. Locally, i.e. on scales of order of the horizon at recombination, this metric describes an unperturbed universe as the terms with Φ and Ψ can be taken to be constant in space and reabsorbed with a change of coordinates. In particular, the evolution is unperturbed in terms of a new conformal time $\tilde{\tau}$ which satisfies (in matter dominance $a\propto\tau^2$)

$$\tau^4(1+2\Phi)d\tau^2=\tilde{\tau}^4d\tilde{\tau}^2 \,, \tag{3.24}$$

i.e. $\tilde{\tau}=\tau(1+2\Phi)^{1/6}$. The temperature on a $\tau=$ const. surface will be perturbed, because the same value of τ corresponds to different moments along the unperturbed evolution, i.e. to different values of $\tilde{\tau}$. As $T\propto 1/\tilde{\tau}^2$ we have

$$T_e=(1+2\Phi_e)^{-1/3}\bar{T}_e \,, \tag{3.25}$$

[1] Notice that in eq. (3.8) the prefactor in parentheses, $1/3-j_1(k\tau)/(k\tau)$, goes to zero for $k\tau\to 0$, i.e. when the γ mode is out of the horizon.

which coincides, at second order, with eq. (3.22).

Now, let us plug both this equation and eq. (3.19) into eq. (3.3) and write the observed CMB temperature up to second-order as a function of the metric perturbations,

$$T_o(\hat{n}) = \frac{a_e}{a_o}\bar{T}_e\sqrt{\frac{1+2\Phi_e}{1+2\Phi_o}}\,(1+2\Phi_e)^{-1/3}\left[1+\int_{\tau_e}^{\tau_o}\mathrm{d}\tau\left(\Phi'+\Psi'+\omega_i'\hat{n}^i-\frac{1}{2}\gamma_{ij}'\hat{n}^i\hat{n}^j\right)\right]. \tag{3.26}$$

Note that on the right hand side of this equation, the gravitational potential at the observer, Φ_o, does not depend on the direction of observation. Thus, its dependence can be simply reabsorbed into the definition of $\bar{T}_o$. Expanding this equation up to second order in the perturbation and plugging the right hand side in eq. (3.2) we finally obtain the CMB temperature anisotropies,

$$\frac{\delta T_o}{T_o}(\hat{n}) = \frac{1}{3}\Phi_e - \frac{5}{18}\Phi_e^2 + \int_{\tau_e}^{\tau_o}\mathrm{d}\tau\left(\Phi'+\Psi'+\omega_i'\hat{n}^i-\frac{1}{2}\gamma_{ij}'\hat{n}^i\hat{n}^j\right). \tag{3.27}$$

The first two terms on the right hand side of this equation have to be evaluated at the position of the emitted photon, $\vec{x}_e$. Since the second term is second-order, it can be simply evaluated at the background position $\hat{n}D_e$, with $D_e \equiv \tau_o - \tau_e$. Also the integral is second-order; thus it can be computed along the background photon trajectory, i.e. $\vec{x}(\tau) = \hat{n}\,D(\tau)$, $D(\tau) \equiv \tau_o - \tau$. However, the first term on the right hand side is a first-order quantity. Thus, at second order it must be evaluated at the perturbed position of the photon at emission. Expanding around the background position $\hat{n}D_e$ we can write it as

$$\Phi(\vec{x}_e) = \Phi(\hat{n}D_e) + \delta\vec{x}_e\cdot\vec{\nabla}\phi(\hat{n}D_e)\,, \tag{3.28}$$

where $\delta\vec{x}_e \equiv \vec{x}_e - \hat{n}D_e$ is the deviation from the background trajectory and we have used that $\Phi = \phi$ at first order.

In order to find $\delta\vec{x}_e$ we must solve the spatial component of the geodesic equation. Since $\vec{\nabla}\phi$ is already first-order we need to compute $\delta\vec{x}_e$ at first-order only. Thus, equation (3.14) gives

$$P^0\frac{\mathrm{d}P_i}{\mathrm{d}\tau} = -2a^2\partial_i\phi(P^0)^2\,, \tag{3.29}$$

where we have used that $\Phi+\Psi = 2\phi$ at first order. This equation can be integrated using the background relation $P^0 \propto 1/a^2$. The spatial gradient can be decomposed along and orthogonally to the background photon trajectory. Since ϕ is time-independent, the component along the photon trajectory is a total derivative. Furthermore, raising the spatial index with the first order metric and then using $P^0 \propto (1-2\phi)/a^2$ one obtains

$$\frac{dx^i}{d\tau} = \frac{P^i}{P^0} = -\hat{n}^i(1+2\phi) + 2\int_{\tau}^{\tau_o}\mathrm{d}\tau'\nabla_{\parallel}^i\phi\,, \tag{3.30}$$

where we have defined $\nabla_{\parallel}^i \equiv (\delta^{ij}-\hat{n}^i\hat{n}^j)\partial_j$ as the spatial gradient orthogonal to the line of sight[1] and we have absorbed the dependence on ϕ_o in the first-order definition

[1]Notice that the direction perpendicular to the photon trajectory is parallel to the flat sky, so that, in our notation, the gradient is *parallel to the sky*.

of $\hat{n}$, $\hat{n}^i \equiv -P_o^i/P_o^0(1+2\phi_o)$. Integrating this equation and subtracting the background value $\hat{n}D_e$, after an integration by parts in the second integral one obtains the geodesic deviation

$$\delta\vec{x}_e = 2\hat{n}\int_{\tau_e}^{\tau_o} \mathrm{d}\tau\phi - 2\int_{\tau_e}^{\tau_o} \mathrm{d}\tau(\tau-\tau_e)\vec{\nabla}_{\parallel}\phi\,. \tag{3.31}$$

The first term on the right hand side, longitudinal to the line of sight, is the so-called Shapiro time-delay. This effect was discussed in [82] and we will discard it from the following discussion where we will concentrate on modes much shorter than the present Hubble radius, where the flat-sky approximation is valid. Indeed, since the integral of ϕ tends to average to zero unless the mode wave-vector is orthogonal to the line of sight, it gives a negligible contribution to the CMB anisotropy for $l \gg 1$. The second term is the transverse deviation from the background trajectory, responsible for the lensing effect [83].

Including the lensing effect by re-expressing Φ_e using eq. (3.28) and re-writing Φ in terms of ϕ using the large-scale limit of eq. (3.5), i.e., $\Phi = \phi + \phi^2 + \partial^{-2}(\partial_j\phi)^2 - 3\partial^{-4}\partial_i\partial_j(\partial_i\phi\partial_j\phi)$, eq. (3.27) can be finally written as

$$\begin{aligned}\frac{\delta T}{T}(\hat{n}) = &\left[\frac{1}{3}\phi + \frac{1}{18}\phi^2 + \frac{1}{3}\partial^{-2}\left((\partial_i\phi)^2 - 3\partial^{-2}\partial_i\partial_j(\partial_i\phi\partial_j\phi)\right)\right]_e \\ &+ \int_{\tau_e}^{\tau_o} \mathrm{d}\tau\left(\Phi' + \Psi' + \omega_i'\hat{n}^i - \frac{1}{2}\gamma_{ij}'\hat{n}^i\hat{n}^j\right) + \frac{1}{3}\vec{\alpha}\cdot\vec{\nabla}_{\hat{n}}\phi_e\,,\end{aligned} \tag{3.32}$$

where $\vec{\alpha}$ is the deviation angle given by eq. (3.31) as

$$\vec{\alpha} \equiv -2\int_{\tau_e}^{\tau_o} \mathrm{d}\tau\frac{\tau-\tau_e}{\tau_o-\tau_e}\vec{\nabla}_{\parallel}\phi\,. \tag{3.33}$$

On the right hand side of eq. (3.32), the subscript "e" means at the background position of the emitted photon, $\hat{n}D_e$. The first line of eq. (3.32) was found in [73]. It represents an intrinsic effect due to the combination of the Doppler effect and the adiabatic temperature fluctuation of the plasma at recombination. The second line contains the Rees-Sciama effect, due to the second-order time evolution of the scalar potentials, and the effect of the time dependence of the vector and tensor components of the metric. Finally, the last term in the second line of eq. (3.32) represents the lensing effect. All these effects were discussed for a more general metric in [78; 79].

There is a nice way to check the factor $\phi_e^2/18$ in the expression (3.32) which, as we will see, is important for the squeezed limit of the bispectrum [73]. Let us take the limit in which one of the two Fourier modes of the initial conditions ϕ_e becomes infinitely long. This mode is still out of the horizon today and therefore cannot affect any physical observable. Let us check that this is indeed the case. When one of the wavevectors goes to zero, all the terms containing spatial derivatives in the expression above vanish, as it is clear from the explicit form of the metric eqs. (3.5)–(3.8). One is left only with the first two terms which, up to second order, it is useful to rewrite in

an exponential form [74] as

$$\frac{\delta T}{T}(\hat{n}) = \left[\frac{1}{3}\phi + \frac{1}{18}\phi^2\right]_e \simeq e^{\phi_e/3} - 1 \,. \tag{3.34}$$

At first sight it looks as if the constant mode could affect observations through the second order term, which mixes a short mode with the constant one. This actually is not the case as the constant mode also affects the average measured temperature. Indeed the well defined measurable quantity is given by

$$\frac{T_o(\hat{n}) - \bar{T}_o}{\bar{T}_o} = \frac{e^{\phi_e/3}}{\langle e^{\phi_e/3}\rangle} - 1 \,. \tag{3.35}$$

Now we see that indeed a constant contribution to ϕ_e cancels out: the quadratic term cancels with the redefinition of the average temperature. Notice that this is only possible because of the exact numerical coefficient $1/18$ in front of the quadratic term. For the calculation of the bispectrum we are only interested in modes inside the Hubble radius at present time, thus it is not necessary to modify eq. (3.32) to take into account the correct average temperature as in eq. (3.35).

In this way we also understand why the argument presented in [77] for the squeezed limit of the 3-point function is not correct. In that reference it is argued that a term like $\phi_e^2/18$, which induces a correlation between short and long modes, cannot exist, as it would imply – as in eq. (3.34) – that a mode which is still out of the horizon gives a measurable effect. What was neglected is that the same mode would change the average of the measured temperature.

3.3 The CMB bispectrum and its shape

In this section we will discuss the CMB bispectrum and its shape dependence. We will use the flat-sky approximation. Even though this approximation is not very good for the lowest multipoles, the expressions that we will derive are much more transparent than using a full-sky treatment.

In the flat-sky approximation (see appendix 3.A) the Fourier transform in the sky of the temperature anisotropies is

$$a_{\vec{l}} = \int \mathrm{d}^2 m \, \frac{\delta T}{T}(\hat{n})\, e^{-i\vec{l}\cdot\vec{m}} \,, \tag{3.36}$$

where we have decomposed $\hat{n}$ into a part orthogonal and parallel to the line of sight as $\hat{n} \simeq (\vec{m}, 1)$ (see appendix 3.A). The spectrum of the 2-point function is defined as

$$\langle a_{\vec{l}} a_{\vec{l}'} \rangle = (2\pi)^2 \delta(\vec{l} + \vec{l}') C_l \,. \tag{3.37}$$

We can rewrite the standard linear Sachs-Wolfe term in eq. (3.32) in Fourier space,

$$\frac{\delta T}{T}(\hat{n}) = \int \frac{\mathrm{d}^3 k}{(2\pi)^3} \, \frac{1}{3}\phi_{\vec{k}}\, e^{i\vec{k}\cdot\hat{n}D_e} \,. \tag{3.38}$$

As explained more accurately in appendix 3.A, it is convenient to separate $\vec{k}$ as the sum of a 2-dimensional vector parallel to the flat sky and a component orthogonal to it,

$$\vec{k} \equiv (\vec{k}^{\parallel}, k^{\perp}) \,. \tag{3.39}$$

Using this decomposition and inserting eq. (3.38) in eq. (3.36) one obtains

$$a_{\vec{l}} = \int \frac{\mathrm{d}^3 k}{(2\pi)^3} \frac{1}{3} \phi_{\vec{k}} \, e^{ik^{\perp} D_e} \, (2\pi)^2 \delta(\vec{l} - \vec{k}^{\parallel} D_e) \,. \tag{3.40}$$

From this expression the power spectrum defined in eq. (3.37) reads,

$$C_l = \frac{A}{9\pi l^2} \,, \tag{3.41}$$

where for simplicity we have used a scale invariant power spectrum for the gravitational potential ϕ,

$$\langle \phi_{\vec{k}} \, \phi_{\vec{k}'} \rangle \equiv (2\pi)^3 \delta(\vec{k} + \vec{k}') \frac{A}{k^3} \,. \tag{3.42}$$

We are interested in the ensemble average of the product of three $a_{\vec{l}}$. Thus, we define the CMB bispectrum $B(\vec{l}_1, \vec{l}_2, \vec{l}_3)$ as

$$\langle a_{\vec{l}_1} a_{\vec{l}_2} a_{\vec{l}_3} \rangle = (2\pi)^2 \delta(\vec{l}_1 + \vec{l}_2 + \vec{l}_3) B(\vec{l}_1, \vec{l}_2, \vec{l}_3) \,. \tag{3.43}$$

Translational and rotational invariance reduce the number of degrees of freedom of B to three independent variables only, for instance l_1, l_2, l_3. This is completely general, but in the particular limit that we are studying (large scales and perfect matter dominance) we will also see that the leading contributions to the bispectrum are scale invariant, i.e. the amount of non-Gaussianity is the same at long and short scales. Mathematically this implies that the function B is a homogeneous function of degree -4,

$$B(\lambda \vec{l}_1, \lambda \vec{l}_2, \lambda \vec{l}_3) = \lambda^{-4} B(\vec{l}_1, \vec{l}_2, \vec{l}_3) \,, \tag{3.44}$$

which further reduces the number of degrees of freedom to two, for instance the ratios $r_2 \equiv l_2/l_1$ and $r_3 \equiv l_3/l_1$. Without loss of generality we can assume $0 \leq r_2 \leq r_3 \leq 1$; the triangle inequality implies $r_2 \geq 1 - r_3$. This is very similar to what happens when one studies the shape dependence of the primordial 3-point function of the curvature perturbation [34], with the difference that here we are in two and not three dimensions.

We are interested in the dependence of B on the two ratios r_2 and r_3, which describes how the bispectrum changes as we change the shape of the triangle in Fourier space. The possibility to measure a bispectrum depends on its signal to noise ratio S/N, which is given in flat-sky approximation by [84]

$$(S/N)^2 = \frac{1}{\pi} \int \frac{\mathrm{d}^2 l_2 \mathrm{d}^2 l_3}{(2\pi)^2} \frac{B(\vec{l}_1, \vec{l}_2, \vec{l}_3)^2}{6 C_{l_1} C_{l_2} C_{l_3}} \,. \tag{3.45}$$

The overall scaling in l is fixed by eq. (3.44) and (3.41): the integrand scales as l^{-2}. To study the shape dependence one can look at the quantity

$$r_2\, r_3\, B(1, r_2, r_3)\ . \tag{3.46}$$

The square of this quantity is in fact proportional to the integrand in the expression above and thus quantifies the contribution to $(S/N)^2$ of triangles with a given shape. To be more precise one could rewrite the expression (3.45) for $(S/N)^2$ as an integral over the two ratios r_2 and r_3

$$(S/N)^2 \propto \int \mathrm{d}r_2 \mathrm{d}r_3 \left[\frac{r_2^{3/2} r_3^{3/2}}{(2r_2^2 + 2r_3^2 + 2r_2^2 r_3^2 - 1 - r_2^4 - r_3^4)^{1/4}} B(1, r_2, r_3) \right]^2 . \tag{3.47}$$

Therefore it would seem appropriate to consider the function in brackets as a measure of the S/N contribution; in this way in fact the integral of the square of the function over an r_2, r_3 region would directly give the contribution of those shape configurations to $(S/N)^2$. This would exactly parallel what is done in [34] to study the shape dependence of the primordial 3-point function. However in this way we would introduce a spurious divergence in the plots for flattened configurations when all the sides of the triangle are aligned: indeed, the denominator of the expression above blows up in this limit. This is just a consequence of describing the triangle shape in terms of r_2 and r_3 and it does not imply that flattened triangles are indeed more important. For this reason we prefer to plot $r_2\, r_3\, B(1, r_2, r_3)$ in the following.

For comparison with the results that we will derive later, it is interesting to study the function (3.46) when the CMB bispectrum is dominated by a primordial contribution. Two interesting cases are given by the so-called local and equilateral shapes [34].

3.3.1 The local shape

A popular shape, usually used in data analysis, is the one obtained when the potential ϕ contains a non-linear correction in coordinate space,

$$\phi(\vec{x}) = \phi_g(\vec{x}) - f_{\rm NL}^{\rm local}(\phi_g^2(\vec{x}) - \langle \phi_g^2 \rangle)\ . \tag{3.48}$$

(We are using the same sign convention for $f_{\rm NL}^{\rm local}$ as Komatsu et al. [85].) In this case, the 3-point function of the gravitational potential ϕ is

$$\langle \phi_{\vec{k}_1} \phi_{\vec{k}_2} \phi_{\vec{k}_3} \rangle = (2\pi)^3 \delta(\vec{k}_1 + \vec{k}_2 + \vec{k}_3)(-2 f_{\rm NL}^{\rm local} A^2) \left(\frac{1}{k_1^3 k_2^3} + \frac{1}{k_1^3 k_3^3} + \frac{1}{k_2^3 k_3^3} \right) . \tag{3.49}$$

If the non-linear correction (3.48) dominates over those computed in the previous section, then $a_{\vec{l}}$ can be simply computed using eq. (3.40). By taking the ensemble average of the product of three $a_{\vec{l}}$ and using eq. (3.49), the CMB bispectrum induced by local non-linear corrections reads

$$B^{\rm local} = -\frac{2 f_{\rm NL}^{\rm local} A^2}{27\pi^2} \left(\frac{1}{l_1^2 l_2^2} + \frac{1}{l_1^2 l_3^2} + \frac{1}{l_2^2 l_3^2} \right) . \tag{3.50}$$

Note that by rescaling l_2 and l_3 we can pull out an overall factor $l_{\rm I}^{-4}$ and rewrite this bispectrum in terms of the two independent variables r_2 and r_3,

$$B^{\rm local} = -\frac{2f_{\rm NL}^{\rm local}A^2}{27\pi^2 l_1^4}\left(\frac{1}{r_2^2}+\frac{1}{r_3^2}+\frac{1}{r_2^2 r_3^2}\right) . \tag{3.51}$$

In the following we will always use this trick and plot the bispectrum as a function of r_2 and r_3 setting $l_1 = 1$ and $A = 1$. The shape corresponding to eq. (3.51) is plotted in figure 3.1.

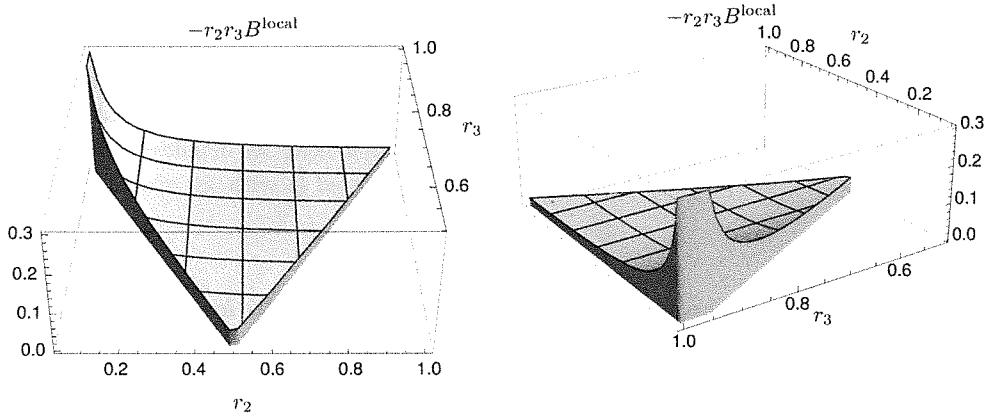


Figure 3.1: The CMB bispectrum on large angular scales induced by primordial non-Gaussianities of the local form for $f_{\rm NL}^{\rm local} = 1$. According to its definition, the bispectrum is negative for positive $f_{\rm NL}^{\rm local}$; thus, we have plotted it with an overall minus sign.

3.3.2 The equilateral shape

Another theoretically motivated shape for the primordial 3-point function is the so-called equilateral shape, that can be described by [34]

$$\langle \phi_{\vec{k}_1}\phi_{\vec{k}_2}\phi_{\vec{k}_3}\rangle = (2\pi)^3\delta(\vec{k}_1+\vec{k}_2+\vec{k}_3)(-6f_{\rm NL}^{\rm equil}A^2) \times\left(-\frac{1}{2k_1^3k_2^3}-\frac{1}{3k_1^2k_2^2k_3^2}+\frac{1}{k_1k_2^2k_3^3}+5\text{ perms.}\right) . \tag{3.52}$$

Notice that the divergence in the squeezed limit is in this case milder than for the local shape, due to a cancellation among the various terms. We can compute the CMB bispectrum similarly to what is done in the local case. It is convenient to define

$$y_1 \equiv k_1^{\perp}(D_e/l_1) , \qquad y_2 \equiv k_2^{\perp}(D_e/l_2) . \tag{3.53}$$

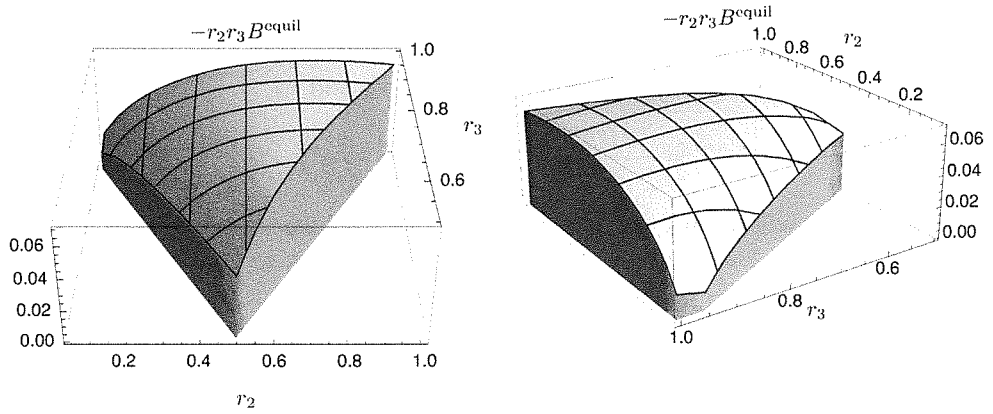


Figure 3.2: The CMB bispectrum on large angular scales induced by primordial non-Gaussianities of the equilateral form for $f_{\rm NL}^{\rm equil} = 1$. According to its definition, the bispectrum is negative for positive $f_{\rm NL}^{\rm equil}$; thus, we have plotted it with an overall minus sign.

With such a definition, using eq. (3.40) for the $a_{\vec{l}}$ and eq. (3.52) for the expectation value of three gravitational potentials, one finally obtains

$$B^{\rm equil} = \frac{2f_{\rm NL}^{\rm equil}A^2}{9(2\pi)^2 l_1^4}\int_{-\infty}^{+\infty} {\rm d}y_1{\rm d}y_2 \left(\frac{1}{2(y_2^2+r_2^2)^{3/2}(y_1^2+r_1^2)^{3/2}} + \frac{1}{3(y_1^2+r_1^2)(y_2^2+r_2^2)\big((y_1+y_2)^2+r_3^2\big)} - \frac{1}{(y_1^2+r_1^2)^{1/2}(y_2^2+r_2^2)\big((y_1+y_2)^2+r_3^2\big)^{3/2}} + 5\,\text{perms.} \right) . \quad (3.54)$$

Here and in the following we sum over all permutations of (r_1, r_2, r_3) and we subsequently set $r_1 = 1$. The integrals cannot be done analytically but the result is plotted in figure 3.2.

From figures 3.1 and 3.2 we see that the CMB bispectra preserve in 2d the qualitative features of the primordial 3-point functions: the signal is peaked on squeezed and equilateral configurations respectively.

3.4 Computing the CMB bispectrum

In this section we compute the CMB bispectra due to the different second-order contributions in eq. (3.32). For comparison, we will use the two typical primordial shapes, local and equilateral, discussed above. We are only interested in computing the CMB non-Gaussianities generated in the Sachs-Wolfe limit; thus, as already mentioned, we

will assume that there is no primordial non-Gaussianity, i.e. that the curvature perturbation on uniform-density hypersurfaces, ζ, is Gaussian on super-Hubble scales. Consequently, from eq. (3.9) it follows that ϕ is Gaussian.

3.4.1 Intrinsic contributions at last scattering

Let us start by computing the CMB non-Gaussianity due to the second-order effects in the first line of eq. (3.32), i.e.,

$$\frac{\delta T}{T}(\hat{n}) \supset \left[\frac{1}{18}\phi^2 + \frac{1}{3}\partial^{-2}\left((\partial_i\phi)^2 - 3\partial^{-2}\partial_i\partial_j(\partial_i\phi\partial_j\phi)\right)\right]_e . \tag{3.55}$$

This contribution has been first derived in [73] and its bispectrum and detectability have been studied in [86]. Note that, although we have dubbed it "intrinsic", this contribution is not physically separable from the other second-order contributions integrated along the photon path that we will study below.

The momentum-independent quadratic term, $\phi_e^2/18$, gives a contribution to the bispectrum exactly of the local shape, equivalent to $f_{\rm NL}^{\rm local} = -1/6$ [73], in eq. (3.51). Its contribution does not vanish in the equilateral limit. We can compare it to an equilateral contribution by evaluating its bispectrum in the equilateral configuration. We find

$$\frac{B^{-1/6}(1,1,1)}{B^{\rm equil}(1,1,1)} \simeq -0.24 , \tag{3.56}$$

where we have evaluated $B^{\rm equil}(1,1,1)$ for $f_{\rm NL}^{\rm equil} = 1$. We conclude that this contribution is equivalent to $f_{\rm NL}^{\rm equil} \simeq -0.24$ in the equilateral limit.

In order to compute the contribution from the momentum-dependent term, we rewrite it as

$$\begin{aligned}\frac{1}{3}\partial^{-2}(\partial_i\phi_e)^2 - \partial^{-4}\partial_i\partial_j(\partial_i\phi_e\partial_j\phi_e) \\ = \int \frac{{\rm d}^3p_1}{(2\pi)^3}\frac{{\rm d}^3p_2}{(2\pi)^3} f^{\rm intr}(\vec{p}_1,\vec{p}_2)\,\phi_{\vec{p}_1}\phi_{\vec{p}_2} e^{i(\vec{p}_1+\vec{p}_2)\cdot\hat{n}\,D_e} ,\end{aligned} \tag{3.57}$$

where $f^{\rm intr}(\vec{p}_1,\vec{p}_2)$ is a kernel defined as

$$f^{\rm intr}(\vec{p}_1,\vec{p}_2) \equiv \frac{1}{3}\frac{\vec{p}_1\cdot\vec{p}_2}{(\vec{p}_1+\vec{p}_2)^2} - \frac{p_1^2\,p_2^2 + (p_1^2+p_2^2)(\vec{p}_1\cdot\vec{p}_2) + (\vec{p}_1\cdot\vec{p}_2)^2}{(\vec{p}_1+\vec{p}_2)^4} . \tag{3.58}$$

Note that this kernel vanishes in the limit of either p_1 or p_2 going to zero. Thus, we expect this contribution to be suppressed with respect to the local shape in the squeezed limit.

The Fourier transform in the sky of this contribution is

$$a_{\vec{l}} = \int \frac{{\rm d}^3p_1}{(2\pi)^3}\frac{{\rm d}^3p_2}{(2\pi)^3} f^{\rm intr}(\vec{p}_1,\vec{p}_2)\,\phi_{\vec{p}_1}\phi_{\vec{p}_2} e^{i(p_1^\perp+p_2^\perp)D_e}(2\pi)^2\delta(\vec{l} - (\vec{p}_1^{\,\parallel} + \vec{p}_2^{\,\parallel})D_e) . \tag{3.59}$$

To compute the bispectrum we can contract this contribution, which is quadratic in ϕ, with the product of two linear Sachs-Wolfe effects, whose $a_{\vec{l}}$ are given by eq. (3.40). By doing so, evaluating the 4-point function of ϕ using Wick's theorem and the definition of the power spectrum, eq. (3.42), summing over all permutations, and using the definition of the bispectrum, eq. (3.43), one obtains

$$B^{\rm intr} = \frac{2A^2}{9(2\pi)^2 l_1^4}\int_{-\infty}^{+\infty} dy_1 dy_2 \Bigg\{ \frac{1}{(y_1^2+r_1^2)^{3/2}(y_2^2+r_2^2)^{3/2}} \Bigg[\frac{2y_1y_2+r_3^2-r_1^2-r_2^2}{6\big((y_1+y_2)^2+r_3^2\big)} \\ - \frac{1}{4\big((y_1+y_2)^2+r_3^2\big)^2}\Big(4(y_1^2+r_1^2)(y_2^2+r_2^2) + (2y_1y_2+r_3^2-r_1^2-r_2^2)^2 \\ + 2(y_1^2+r_1^2+y_2^2+r_2^2)(2y_1y_2+r_3^2-r_1^2-r_2^2)\Big)\Bigg] + 2\,\text{cyclic}\Bigg\} . \qquad (3.60)$$

The integrals in the expression above can be integrated numerically. The final result for the bispectrum coming from this contribution is plotted in figure 3.3. Its contribution is equivalent to $f_{\rm NL}^{\rm equil} \simeq 1.21$.

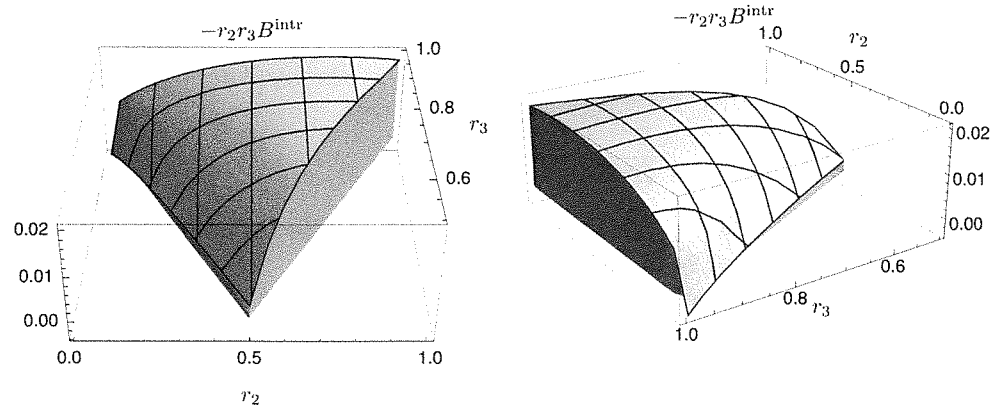


Figure 3.3: The CMB bispectrum induced by the momentum dependent intrinsic contribution in eq. (3.55).

Notice that this bispectrum is suppressed in the squeezed limit with respect to the local case in figure 3.1. This, as discussed, is a consequence of the derivatives in eq. (3.55). Notice also that the suppression, in the limit $r_2 \to 0$, is linear in r_2 as there is one derivative acting on each ϕ in eq. (3.55). Thus, in the plots (which include a measure r_2r_3) the function goes to a constant. This constant depends on the orientation between the long wavelength mode and the short ones as it is clear from eq. (3.58): indeed, in the figure we see that the limit $r_2 \to 0$ depends on the direction from which the limit is approached. Notice that this behaviour is different from the case of primordial equilateral non-Gaussianity where there is a suppression going like

r_2^2 in the squeezed limit – $B^{\rm equil}$ diverges logarithmically for $r_2 \to 0$, see eq. (3.54) – so that the plot in figure 3.2 goes to zero. Indeed, in this limit the 3d kernel (3.52) is suppressed by two powers of k_3 with respect to the local shape and this behavior is typical of all equilateral models [30; 34].

3.4.2 Contribution from the Rees-Sciama effect

At second-order in the perturbations, the Newtonian and curvature potentials Φ and Ψ have a constant and a time-dependent part. While the constant part given in the first line of eqs. (3.5) and (3.6) dominates on large scales, on sub-Hubble scales one recovers the standard Newtonian limit [87], i.e. the two potentials become equal, $\Phi = \Psi$, and grow as the scale factor, $\Phi \propto (aH)^{-2} \propto a$, where we have used $a \propto \tau^2$. Thus, we expect the photon frequency to be affected by an integrated effect. This is the so-called Rees-Sciama effect [88], given by

$$\frac{\delta T}{T}(\hat{n}) \supset \int_{\tau_e}^{\tau_o} \mathrm{d}\tau \, \left(\Phi' + \Psi'\right) . \tag{3.61}$$

Its contribution to the CMB bispectrum has already been considered in [75; 76] although these analysis were restricted only to the diagonal terms of the bispectrum. More generally, the bispectrum from the Rees-Sciama effect has been studied in [89].[1]

Symmetrizing over the momenta, we can rewrite the integrand in eq. (3.61) as

$$\Phi' + \Psi' = \frac{1}{D_e} \int \frac{\mathrm{d}^3 p_1}{(2\pi)^3} \frac{\mathrm{d}^3 p_2}{(2\pi)^3} f^{\rm RS}(\vec{p}_1, \vec{p}_2)\, \phi_{\vec{p}_1} \phi_{\vec{p}_2} e^{i(\vec{p}_1+\vec{p}_2)\cdot\hat{n} D(\tau)} , \tag{3.62}$$

where $f^{\rm RS}$ is an explicitly time-dependent kernel derived from eqs. (3.5) and (3.6) defined as

$$f^{\rm RS}(\vec{p}_1, \vec{p}_2) \equiv -\tau D_e \frac{4(\vec{p}_1 \cdot \vec{p}_2)^2 + 10 p_1^2 p_2^2 + 7(p_1^2 + p_2^2)(\vec{p}_1 \cdot \vec{p}_2)}{21(\vec{p}_1 + \vec{p}_2)^2} . \tag{3.63}$$

Note that we have multiplied it by D_e to make it dimensionless. The Fourier transform in the sky of this contribution is given by

$$a_{\vec{l}} = \int_{\tau_e}^{\tau_o} \frac{d\tau}{D_e} \int \frac{\mathrm{d}^3 p_1}{(2\pi)^3} \frac{\mathrm{d}^3 p_2}{(2\pi)^3} (2\pi)^2 \delta(\vec{l} - (\vec{p}_1^{\,\parallel} + \vec{p}_2^{\,\parallel}) D(\tau)) \times f^{\rm RS}(\vec{p}_1, \vec{p}_2)\, \phi_{\vec{p}_1} \phi_{\vec{p}_2} e^{i(p_1^\perp + p_2^\perp) D(\tau)} . \tag{3.64}$$

As done for the intrinsic contribution, in order to compute the bispectrum we need to contract $a_{\vec{l}}$ in the above equation with the product of two linear Sachs-Wolfe effects, whose $a_{\vec{l}}$ are given by eq. (3.40). Note, however, that the Rees-Sciama kernel $f^{\rm RS}$ in eq. (3.63) is higher order in the spatial gradients with respect to the intrinsic kernel

[1] As there is an error in the derivation of eq. (23) of [89], our results cannot be compared with that reference.

$f^{\rm intr}$ of eq. (3.58), so that one may think that its contribution to the bispectrum will be relevant only on short scales. Indeed, since we are correlating the Rees-Sciama effect with the linear Sachs-Wolfe effect, which takes place at the last scattering surface, one may naively conclude that its contribution to the bispectrum is suppressed in the limit of large angles, i.e. in the limit where gradients are much smaller than the Hubble rate at decoupling. However, this is not the case. Indeed, the correlation with what happens at the last scattering surface does not vanish immediately for $\tau > \tau_e$, but for a given mode l, it remains large for $\tau \lesssim \tau_* \equiv D_e/l$ and after that decays exponentially. In other words the correlation decays when the distance from the last scattering surface is of the order of the typical wavelength. In appendix 3.A we explain better this point with a simple example. Now, since the Rees-Sciama effect grows with τ, the contribution to the bispectrum will be maximal for $\tau \approx \tau_*$. Using that $k \sim l/D_e$ one has that the maximal contribution comes for $k\tau_* \sim 1$ so that the gradients are not suppressed at τ_* and one expects the Rees-Sciama contribution to the bispectrum to be of the same order as one of the intrinsic terms. Notice also that, as for the intrinsic kernel, also the kernel (3.63) vanishes in the limit of either p_1 or p_2 going to zero; we thus expect the Rees-Sciama bispectrum to be suppressed in the squeezed limit with respect to the local shape.

Let us move to the explicit calculation. It is convenient to define

$$x \equiv (\tau - \tau_e)(l_1/D_e) \,. \tag{3.65}$$

By contracting $a_{\vec{l}}$ given by eq. (3.64) with the product of two linear contributions given by eq. (3.40), using Wick's theorem and the definition of the power spectrum, eq. (3.42), to rewrite the 4-point function of ϕ, and summing over all permutations one obtains, by using the variables y_1 and y_2,

$$\begin{aligned} B^{\rm RS} = -\frac{2A^2}{189(2\pi)^2}\frac{1}{l_1^4}\int_0^{l_1} {\rm d}x\,(x+\tau_e(l_1/D_e))\int_{-\infty}^{+\infty} {\rm d}y_1{\rm d}y_2\, e^{i(y_1+y_2)x} \\ \times\Bigg[\frac{1}{(y_1^2+r_1^2)^{3/2}(y_2^2+r_2^2)^{3/2}}\Bigg(\frac{3}{2}r_1^2+\frac{3}{2}r_2^2+r_3^2+2y_1y_2 \\ +\frac{5}{2}(y_1^2+y_2^2)-\frac{5}{2}\frac{(y_1^2-y_2^2+r_1^2-r_2^2)^2}{(y_1+y_2)^2+r_3^2}\Bigg)+2\ \text{cyclic}\Bigg] \,. \end{aligned} \tag{3.66}$$

Actually the result of the calculation is not proportional to $(2\pi)^2\delta(\vec{l}_1+\vec{l}_2+\vec{l}_3)$ as in the definition of eq. (3.43), but to $(2\pi)^2\delta\left((\vec{l}_1+\vec{l}_2)\frac{\tau_o-\tau}{\tau_o-\tau_e}+\vec{l}_3\right)$ and permutations, as a consequence of the fact that we are correlating effects at different conformal times τ. This is a bit surprising as the delta function $\delta(\vec{l}_1+\vec{l}_2+\vec{l}_3)$ is just a consequence of translational invariance. However, the discussion above implies that the bispectrum is exponentially suppressed when the triangle in Fourier space does not close, i.e. when $\frac{\tau_o-\tau}{\tau_o-\tau_e}l_3 \sim 1$. This can be checked explicitly in the expression (3.66). In appendix 3.A we discuss a simple example in which this issue is made more transparent.

The above integrals are particularly challenging even numerically. However, some simplifications can be made. Since the integrand is exponentially suppressed for $x \gg 1$

by the rapid oscillations of $e^{i(y_1+y_2)x}$, one can push the upper limit of the integral in x to ∞. Another simplification consists in neglecting $\tau_e(l_1/D_e)$ in the first integral of eq. (3.66), which is justified by the fact that we consider only modes well outside the Hubble radius at recombination and thus $\tau_e(l_1/D_e) \sim \tau_e k \ll 1$. With these approximations eq. (3.66) can be rewritten as

$$B^{\rm RS} = -\frac{2A^2}{189(2\pi)^2}\frac{1}{l_1^4}\int_0^\infty \mathrm{d}x\, x \int_{-\infty}^{+\infty} \mathrm{d}y_1 \mathrm{d}y_2\, e^{i(y_1+y_2)x} \times \left[\frac{1}{(y_1^2+r_1^2)^{3/2}(y_2^2+r_2^2)^{3/2}}\left(\frac{3}{2}r_1^2+\frac{3}{2}r_2^2+r_3^2+2y_1y_2 + \frac{5}{2}(y_1^2+y_2^2) - \frac{5}{2}\frac{(y_1^2-y_2^2+r_1^2-r_2^2)^2}{(y_1+y_2)^2+r_3^2}\right) + 2 \text{ cyclic}\right] . \tag{3.67}$$

We see that the bispectrum induced by the Rees-Sciama effect goes as l^{-4} and it is parametrically similar to the intrinsic contribution discussed in the previous section. The analytical and numerical study of this expression is postponed to appendix 3.B. The final result for the bispectrum is given in figure 3.4.

As for the intrinsic contribution (3.60), in the squeezed limit $r_2 \to 0$ the Rees-Sciama bispectrum is suppressed when compared with the local shape by r_2, with a coefficient which depends on the angle. We show this analytically in appendix 3.B. By comparing the Rees-Sciama bispectrum to the equilateral contribution, as we did for the intrinsic one, we find that the Rees-Sciama contribution is equivalent to $f_{\rm NL}^{\rm equil} \simeq 0.74$.

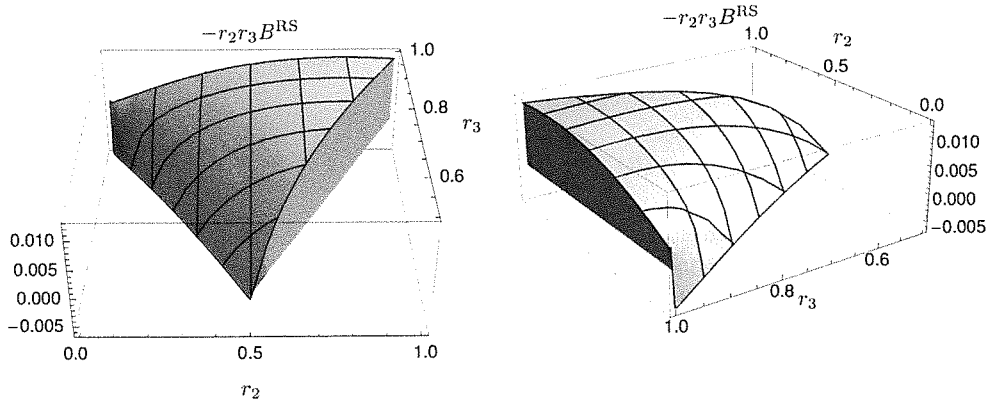


Figure 3.4: The CMB bispectrum induced by the Rees-Sciama effect, eq. (3.61).

3.4.3 Integrated vector contribution

At second order, the non-diagonal part of the metric $g_{0i} \equiv a^2\omega_i$ becomes non-vanishing and time dependent on sub-Hubble scales. Similarly to the time-dependent part of the

gravitational potentials, it induces an integrated effect on the photon redshift, given in eq. (3.32) by

$$\frac{\delta T}{T}(\hat{n}) \supset \int_{\tau_e}^{\tau_o} \mathrm{d}\tau \; \omega_i' \hat{n}^i \,. \tag{3.68}$$

As ω_i is transverse we refer to this effect as the integrated vector contribution. As for the Rees-Sciama, to compute the bispectrum we need to correlate this integrated effect with the intrinsic temperature fluctuation at last scattering. Even though this effect is suppressed at last scattering, when modes are still out of the Hubble radius, it will give us a contribution to $f_{\rm NL}$ of order unity, similarly to what happens for the Rees-Sciama effect.

From eq. (3.7) we can rewrite the integrand as

$$\omega_i' \hat{n}^i = \frac{1}{D_e} \int \frac{\mathrm{d}^3 p_1}{(2\pi)^3} \frac{\mathrm{d}^3 p_2}{(2\pi)^3} f^{\rm V}(\vec{p}_1, \vec{p}_2)\, \phi_{\vec{p}_1} \phi_{\vec{p}_2} e^{i(\vec{p}_1+\vec{p}_2)\cdot\hat{n} D(\tau)} \,, \tag{3.69}$$

where $f^{\rm V}$ is a kernel defined as

$$f^{\rm V}(\vec{p}_1, \vec{p}_2) = -\frac{2iD_e}{3} \left[\frac{p_1^2(\hat{n}\cdot\vec{p}_2) + p_2^2(\hat{n}\cdot\vec{p}_1)}{(\vec{p}_1+\vec{p}_2)^2} - \hat{n}\cdot(\vec{p}_1+\vec{p}_2) \frac{2p_1^2 p_2^2 + (p_1^2+p_2^2)(\vec{p}_1\cdot\vec{p}_2)}{(\vec{p}_1+\vec{p}_2)^4} \right] . \tag{3.70}$$

Note that the second term in the kernel (3.70) is proportional to $\hat{n}\cdot(\vec{p}_1+\vec{p}_2)$. Thus, it is a time total derivative which can be trivially integrated in τ in eq. (3.68). Therefore we have another term evaluated at last scattering, analogous to the intrinsic contributions studied in section 3.4.1, of the form

$$\frac{\delta T}{T}(\hat{n}) \supset \frac{4}{3} \left[\partial^{-4} \partial_j (\partial^2 \phi \partial_j \phi) \right]_e \,. \tag{3.71}$$

This shows clearly that there is nothing really intrinsic about the contributions discussed in section 3.4.1: the splitting among the various effects is gauge dependent and only the total sum has a well defined gauge invariant meaning.

One can then split the rest of the kernel orthogonally to and along the line of sight. Indeed, decomposing $\hat{n}$ into the parts orthogonal and parallel to the line of sight as $\hat{n} = (\vec{m}, 1)$, the first term in eq. (3.70) can be rewritten as

$$-\frac{2iD_e}{3} \left[\frac{\vec{m}\cdot(\vec{p}_2^{\,\parallel} p_1^2 + \vec{p}_1^{\,\parallel} p_2^2)}{(\vec{p}_1+\vec{p}_2)^2} + \frac{p_1^2 p_2^\perp + p_2^2 p_1^\perp}{(\vec{p}_1+\vec{p}_2)^2} \right] . \tag{3.72}$$

The first term of this expression is proportional to $\vec{m}$. Thus, it is higher order in $1/l$ with respect to the second term and therefore negligible in the flat-sky approximation. Thus, the Fourier transform on the sky of the contribution (3.72) can be approximated

with

$$a_{\vec{l}} = -\frac{2iD_e}{3}\int_{\tau_e}^{\tau_o}\frac{d\tau}{D_e}\int\frac{\mathrm{d}^3p_1}{(2\pi)^3}\frac{\mathrm{d}^3p_2}{(2\pi)^3}(2\pi)^2\delta(\vec{l}-(\vec{p}_1^{\,\parallel}+\vec{p}_2^{\,\parallel})D(\tau)) \times e^{i(p_1^\perp+p_2^\perp)D(\tau)}\frac{p_1^2p_2^\perp+p_2^2p_1^\perp}{(\vec{p}_1+\vec{p}_2)^2}\phi_{\vec{p}_1}\phi_{\vec{p}_2}\,. \quad (3.73)$$

Proceeding as in the case of the intrinsic and Rees-Sciama contributions, the total contribution from vectors can be written, using the variables y_1 and y_2, as

$$\begin{aligned}B^{\mathrm{V}} = \frac{4A^2}{27(2\pi)^2l_1^4}\int_{-\infty}^{\infty}\mathrm{d}y_1\mathrm{d}y_2\Bigg[&\frac{1}{(y_1^2+r_1^2)^{3/2}(y_2^2+r_2^2)^{3/2}}\\ &\times\Bigg(\frac{1}{2\big((y_1+y_2)^2+r_3^2\big)^2}\Big((y_1^2+r_1^2)\big(2y_2(y_1+y_2)-(r_1^2-r_2^2-r_3^2)\big)\\ &+(y_2^2+r_2^2)\big(2y_1(y_1+y_2)-(r_2^2-r_1^2-r_3^2)\big)\Big)\\ &+i\int_0^\infty\mathrm{d}x\,e^{i(y_1+y_2)x}\frac{(y_1^2+r_1^2)y_2+(y_2^2+r_2^2)y_1}{(y_1+y_2)^2+r_3^2}\Bigg)+2\text{ cyclic}\Bigg]\,. \end{aligned} \quad (3.74)$$

The first piece, which is not integrated in x, comes from eq. (3.71), while the other term describes the contribution integrated along the line of sight. The integral over time can be dealt with as in the Rees-Sciama case: see appendix 3.B. The final result for this bispectrum is given in figure 3.5. Again, the result is suppressed with respect to the local shape in the squeezed limit because the kernel (3.70) vanishes when either p_1 or p_2 go to zero. The behaviour in this limit is qualitatively the same as in the Rees-Sciama case. This vector contribution is equivalent to $f_{\mathrm{NL}}^{\mathrm{equil}} \simeq -0.84$ in the equilateral configuration.

3.4.4 Integrated tensor contribution

At second order, even in the absence of primordial gravitational waves, the part of the spatial metric not proportional to the identity matrix, $a^2\gamma_{ij}$, is non-vanishing and time dependent on sub-Hubble scales. Thus, it induces an integrated effect given by

$$\frac{\delta T}{T}(\hat{n}) \supset -\int_{\tau_e}^{\tau_o}\mathrm{d}\tau\,\frac{1}{2}\gamma'_{ij}\hat{n}^i\hat{n}^j\,, \quad (3.75)$$

which we expect to contribute to the bispectrum similarly to what happens for the vectors. As γ_{ij} is transverse and traceless, we refer to this effect as the tensor contribution. From eq. (3.8) and using eqs. (3.12) and (3.13) to rewrite the transverse traceless projector, the integrand is

$$-\frac{1}{2}\gamma'_{ij}\hat{n}^i\hat{n}^j = \frac{1}{D_e}\int\frac{\mathrm{d}^3p_1}{(2\pi)^3}\frac{\mathrm{d}^3p_2}{(2\pi)^3}f^{\mathrm{T}}(\vec{p}_1,\vec{p}_2)\,\phi_{\vec{p}_1}\phi_{\vec{p}_2}e^{i(\vec{p}_1+\vec{p}_2)\cdot\hat{n}\,D(\tau)}\,, \quad (3.76)$$

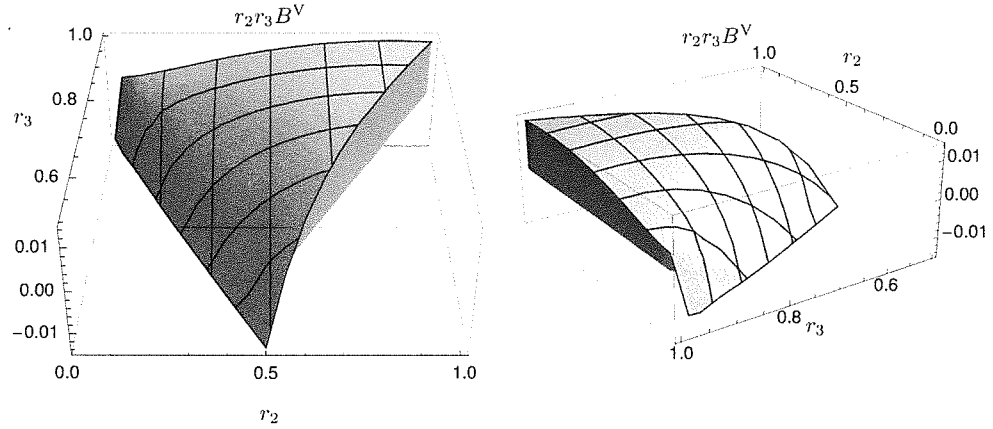


Figure 3.5: The CMB bispectrum induced by the vector contribution, eq. (3.68).

where the kernel $f^{\rm T}$ is defined as

$$f^{\rm T}(\vec{p}_1,\vec{p}_2) = -j_2(|\vec{p}_1+\vec{p}_2|\tau)\frac{5D_e}{\tau}\left[\frac{(\vec{p}_1\cdot\vec{p}_2)^2-p_1^2p_2^2}{(\vec{p}_1+\vec{p}_2)^4}\left(1+\frac{(\hat{n}\cdot(\vec{p}_1+\vec{p}_2))^2}{(\vec{p}_1+\vec{p}_2)^2}\right)\right.$$
$$\left.+\frac{2p_1^2(\hat{n}\cdot\vec{p}_2)^2+2p_2^2(\hat{n}\cdot\vec{p}_1)^2-4(\vec{p}_1\cdot\vec{p}_2)(\hat{n}\cdot\vec{p}_1)(\hat{n}\cdot\vec{p}_2)}{(\vec{p}_1+\vec{p}_2)^4}\right] , \quad (3.77)$$

and j_2 is a spherical Bessel function that appears from taking the time derivative of γ_{ij},

$$\left(\frac{j_1(k\tau)}{k\tau}\right)' = -\frac{j_2(k\tau)}{\tau} . \quad (3.78)$$

As we did for the vector kernel $f^{\rm V}$, $f^{\rm T}$ can be decomposed into a part parallel and orthogonal to the sky. The parallel part is higher order in $1/l$ and thus negligible in the flat-sky approximation. Thus, the kernel can be approximated as

$$f^{\rm T}(\vec{p}_1,\vec{p}_2) \simeq -j_2(|\vec{p}_1+\vec{p}_2|\tau)\frac{5D_e}{\tau}\left[\frac{(\vec{p}_1\cdot\vec{p}_2)^2-p_1^2p_2^2}{(\vec{p}_1+\vec{p}_2)^4}\left(1+\frac{(p_1^\perp+p_2^\perp)^2}{(\vec{p}_1+\vec{p}_2)^2}\right)\right.$$
$$\left.+\frac{2p_1^2(p_2^\perp)^2+2p_2^2(p_1^\perp)^2-4(\vec{p}_1\cdot\vec{p}_2)p_1^\perp p_2^\perp}{(\vec{p}_1+\vec{p}_2)^4}\right] . \quad (3.79)$$

The Fourier transform on the sky of this contribution is given by

$$a_{\vec{l}} = \int_{\tau_e}^{\tau_o}\frac{{\rm d}\tau}{D_e}\int\frac{{\rm d}^3p_1}{(2\pi)^3}\frac{{\rm d}^3p_2}{(2\pi)^3}(2\pi)^2\delta(\vec{l}-(\vec{p}_1^{\,\parallel}+\vec{p}_2^{\,\parallel})D(\tau))$$
$$\times f^{\rm T}(\vec{p}_1,\vec{p}_2)\,\phi_{\vec{p}_1}\phi_{\vec{p}_2}e^{i(p_1^\perp+p_2^\perp)D(\tau)} . \quad (3.80)$$

With this simplification the time integral can be analytically computed and yields, expressing it in terms of the variables x, y_1 and y_2,

$$\int_0^\infty dx\, \frac{j_2(\sqrt{(y_1+y_2)^2+r_3^2}\,x)}{x} e^{i(y_1+y_2)x} = \frac{(2r_3^2-(y_1+y_2)^2)}{6((y_1+y_2)^2+r_3^2)} - \frac{(y_1+y_2)r_3^2 \coth^{-1}\left(\frac{\sqrt{(y_1+y_2)^2+r_3^2}}{y_1+y_2}\right)}{2((y_1+y_2)^2+r_3^2)^{3/2}}, \quad (3.81)$$

plus an imaginary term odd under $(y_1, y_2) \to (-y_1, -y_2)$ which does not contribute to the integral. This gives for the bispectrum

$$\begin{aligned} B^{\rm T} = \frac{10A^2}{9(2\pi)^2 l_1^4} \int_{-\infty}^{+\infty} {\rm d}y_1 {\rm d}y_2 \Bigg\{ & \frac{1}{(y_1^2+r_1^2)^{3/2}(y_2^2+r_2^2)^{3/2}((y_1+y_2)^2+r_3^2)^2} \\ & \times \left[\frac{(2r_3^2-(y_1+y_2)^2)}{6((y_1+y_2)^2+r_3^2)} - \frac{(y_1+y_2)r_3^2 \coth^{-1}\left(\frac{\sqrt{(y_1+y_2)^2+r_3^2}}{y_1+y_2}\right)}{2((y_1+y_2)^2+r_3^2)^{3/2}} \right] \\ & \times \Bigg[\frac{1}{4}\left(1+\frac{(y_1+y_2)^2}{(y_1+y_2)^2+r_3^2}\right)\left(4(y_1^2+r_1^2)(y_2^2+r_2^2) - (2y_1y_2+r_3^2-r_1^2-r_2^2)^2\right) \\ & -2\left(y_2^2(y_1^2+r_1^2)+y_1^2(y_2^2+r_2^2) - y_1y_2(2y_1y_2+r_3^2-r_1^2-r_2^2)\right) \Bigg] \\ & + 2 \text{ cyclic} \Bigg\}, \end{aligned} \quad (3.82)$$

and the final result is plotted in figure 3.6. Again, given that the kernel (3.79) goes to zero when either p_1 or p_2 go to zero, this shape is suppressed with respect to the local one in the squeezed limit. From figure 3.6 we see that the integrated tensor contribution is qualitatively similar to the intrinsic kernel, Rees-Sciama and vector contributions discussed previously. This contribution is equivalent to $f_{\rm NL}^{\rm equil} \simeq -0.61$ for an equilateral configuration.

3.4.5 Lensing

The deflection angle of a light ray as it propagates from the last scattering surface to us is given by eq. (3.33) (for a review of lensing effects on the CMB see [90]). For convenience we reproduce it here,

$$\vec{\alpha} = -2 \int_{\tau_e}^{\tau_o} {\rm d}\tau\, \frac{\tau-\tau_e}{\tau_o-\tau_e} \vec{\nabla}_\parallel \phi\,. \quad (3.83)$$

The geometrical weight $(\tau-\tau_e)/(\tau_o-\tau_e)$ tells us that the effect is suppressed close to the last scattering surface. For this reason, usually the main contribution to the 3-point

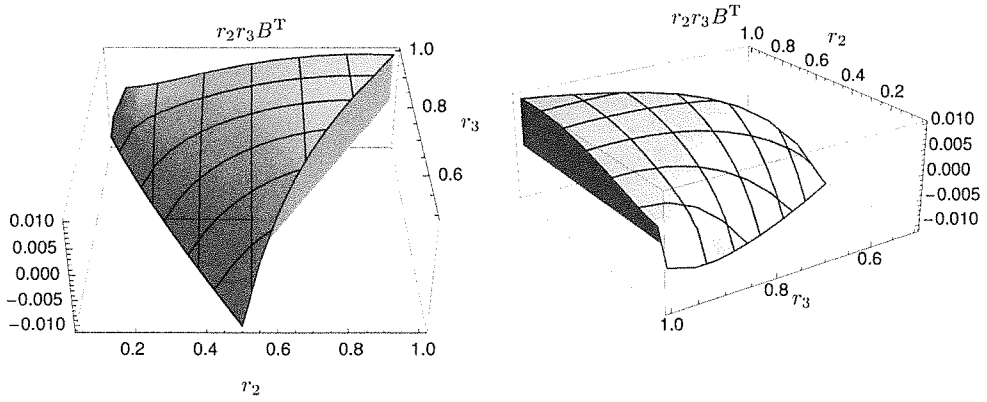


Figure 3.6: The CMB bispectrum induced by the tensor contribution, eq. (3.75).

function due to lensing comes from the correlation of the photon deflection with the ISW [67; 68]. This effect is absent in our case as we are studying a universe with only matter. However, there is still the correlation of the intrinsic temperature fluctuation at last scattering with the lensing contribution given by

$$\frac{\delta T}{T}(\hat{n}) \supset \frac{1}{3}\vec{\alpha}\cdot\vec{\nabla}_{\hat{n}}\phi_e\,. \tag{3.84}$$

Similarly to the other integrated effects also this will give an effective $f_{\rm NL}\sim 1$.

Let us compute the contribution to the bispectrum. Inserting the deviation angle (3.83) into eq. (3.84) and using $\vec{\nabla}_{\hat{n}} = D_e\vec{\nabla}_{\parallel}$, the lensing contribution to the temperature fluctuation can be written as

$$\frac{\delta T}{T}(\hat{n}) \supset \frac{2}{3}\int_{\tau_e}^{\tau_o} {\rm d}\tau\,(\tau-\tau_e)\int \frac{{\rm d}^3p_1}{(2\pi)^3}\frac{{\rm d}^3p_2}{(2\pi)^3}(\vec{p}_1^{\,\parallel}\cdot\vec{p}_2^{\,\parallel})\phi_{\vec{p}_1}\phi_{\vec{p}_2}e^{i\vec{p}_1\cdot\hat{n}D(\tau)}e^{i\vec{p}_2\cdot\hat{n}D_e}\,. \tag{3.85}$$

Taking the Fourier transform on the sky yields

$$a_{\vec{l}} = \frac{2}{3}\int_{\tau_e}^{\tau_o} {\rm d}\tau\,(\tau-\tau_e)\int \frac{{\rm d}^3p_1}{(2\pi)^3}\frac{{\rm d}^3p_2}{(2\pi)^3}(2\pi)^2\delta(\vec{l}-\vec{p}_1^{\,\parallel}D(\tau)-\vec{p}_2^{\,\parallel}D_e) \times(\vec{p}_1^{\,\parallel}\cdot\vec{p}_2^{\,\parallel})\phi_{\vec{p}_1}\phi_{\vec{p}_2}e^{ip_1^{\perp}D(\tau)}e^{ip_2^{\perp}D_e}\,. \tag{3.86}$$

As usual, we can compute the bispectrum by correlating this effect with the intrinsic temperature at last scattering. By doing so, we obtain

$$B^{\rm lens} = -\frac{A^2}{27(2\pi)^2 l_1^4}\int_0^\infty {\rm d}x\,x\int_{-\infty}^{\infty}{\rm d}y_1{\rm d}y_2\,e^{iy_1x}\frac{r_3^2-r_1^2-r_2^2}{(y_1^2+r_1^2)^{3/2}(y_2^2+r_2^2)^{3/2}} + 5\text{ perms.}. \tag{3.87}$$

Note that here one must sum over all permutations of (r_1, r_2, r_3), including the anticyclic ones. The integrals above can be computed analytically, yielding

$$B^{\rm lens} = \frac{8A^2}{27(2\pi)^2 l_1^4} \frac{r_3^2 - r_1^2 - r_2^2}{r_1^4 r_2^2} + 5 \text{ perms.} \tag{3.88}$$

This result is plotted in figure 3.7. Alternatively, this equation can be written as

$$B^{\rm lens} = \frac{16A^2}{27(2\pi)^2} \frac{\vec{l}_1 \cdot \vec{l}_2}{l_1^4 l_2^2} + 5 \text{ perms.} \tag{3.89}$$

Another method to derive the lensing CMB bispectrum is through the lensing potential

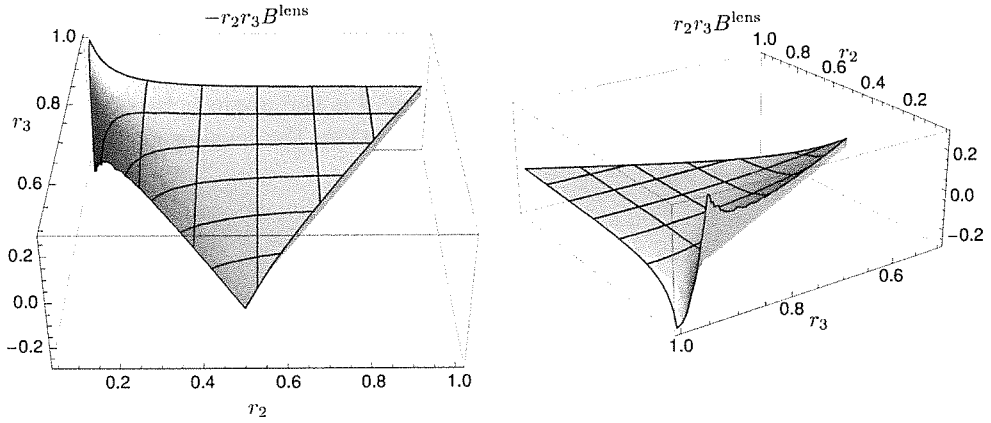


Figure 3.7: The CMB bispectrum induced by the lensing contribution, eq. (3.84).

ψ defined as (see for example [90])

$$\psi(\hat{n}) \equiv -2 \int_{\tau_e}^{\tau_o} {\rm d}\tau \, \frac{\tau - \tau_e}{(\tau_o - \tau_e)(\tau_o - \tau)} \, \phi\left(\hat{n}(\tau_o - \tau), \tau\right) \ . \tag{3.90}$$

The deflection angle (3.83) is obtained by taking the flat-sky gradient of this expression $\vec{\alpha} = \vec{\nabla}_{\hat{n}} \psi$. The correlation between the temperature at the last scattering surface and the lensing potential is given by[1]

$$\langle \psi_{\vec{l}_1} a_{\vec{l}_2} \rangle = -\frac{8\pi A}{3 D_e^2 l_2} \int_{\tau_e}^{\tau_o} {\rm d}\tau \, \frac{(\tau - \tau_e)^2}{(\tau_o - \tau)} \, K_1(l_2(\tau - \tau_e)/D_e) \, \delta(\vec{l}_1 + \vec{l}_2(\tau_o - \tau)/D_e) \ . \tag{3.91}$$

The temperature fluctuation is localized at τ_e while the lensing becomes more and more important at later times. It is easy to see that the correlation is maximal at

[1] As explained in [90], the divergence of the lensing potential at τ_o affects only the monopole, which can always be subtracted.

$\tau_* \sim \tau_o/l$, similarly to what was discussed for all integrated effects in section 3.4.2 (see also appendix 3.A).

As the integral is dominated by $\tau \ll \tau_0$, we can approximate the δ function with $\delta(\vec{l}_1 + \vec{l}_2)$ to get

$$\langle \psi_{\vec{l}_1} a_{\vec{l}_2} \rangle = (2\pi)^2 \delta(\vec{l}_1 + \vec{l}_2) C_{l_1}^{T,\psi} , \qquad C_l^{T,\psi} = -\frac{4A}{3\pi l^4}. \tag{3.92}$$

The bispectrum can be written as [84; 91]

$$B^{\rm lens} = -\vec{l}_1 \cdot \vec{l}_2 \left(C_{l_1} C_{l_2}^{T,\psi} + C_{l_2} C_{l_1}^{T,\psi} \right) + 2 \text{ perms} , \tag{3.93}$$

which coincides with eq. (3.89).

Note that the expression in eq. (3.89) diverges in the squeezed limit. However, the form of the divergence depends on the direction one approaches the limit. One can compare the expression resulting from taking $r_2 \to 0$ with the local form. This gives a contribution equivalent to $f_{\rm NL}^{\rm local} = -\cos(2\theta)$ where θ is the angle between $\vec{l}_2$ and $\vec{l}_1$ when one takes the limit. In the equilateral configuration, the lensing gives a sizable contribution, equivalent to $f_{\rm NL}^{\rm equil} \simeq 2.87$.

3.5 The total CMB bispectrum

In the previous section we have separated the calculation of the CMB bispectrum generated in the Sachs-Wolfe limit into five contributions: an intrinsic contribution expressed in terms of the Newtonian potential evaluated at last scattering, in eq. (3.60), the Rees-Sciama effect, in eq. (3.67), a contribution from the time dependence of the vector and tensor components of the metric, respectively in eqs. (3.74) and (3.82), and finally the lensing effect, in eq. (3.88). However, it is important to stress that only the sum of these contributions has a physical, gauge invariant, meaning. In this section we turn to discuss this sum, i.e. the total bispectrum. This is plotted in figure 3.8.

By comparing this with figure 3.7 one can appreciate that the lensing effect largely dominates the total bispectrum [1]. Let us see this more quantitatively.

In the squeezed limit the bispectrum is dominated by the intrinsic contribution and the lensing. In this limit we can compare the total bispectrum to the local bispectrum (3.51) taken with $f_{\rm NL}^{\rm local} = 1$. This yields

$$\frac{B^{\rm total}(1, r_2 \to 0, r_3 \to 1)}{B^{\rm local}(1, r_2 \to 0, r_3 \to 1)} = -1/6 - \cos(2\theta) , \tag{3.94}$$

[1] As already stressed, the separation among different effects is gauge-dependent. Here we use the standard terminology in calling "lensing" the deflection of photons in Newtonian gauge. For a discussion about the gauge-dependence of lensing, see [90].

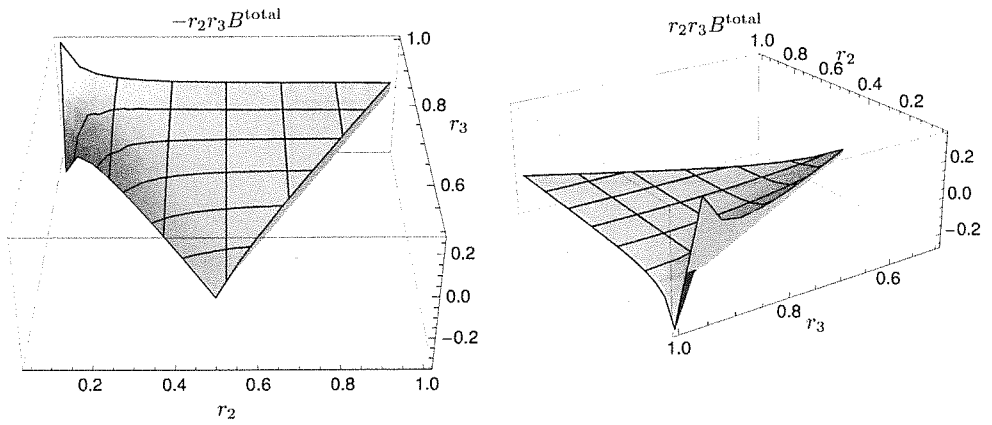


Figure 3.8: The total CMB bispectrum.

where θ represents the angle between the short and long wavelength modes $\vec{l}_1$ and $\vec{l}_2$. Thus, the total bispectrum corresponds to $f_{\rm NL}^{\rm local} = -1/6 - \cos(2\theta)$.

Note that this result can be obtained by simple arguments. As explained in section 3.2, the factor $-1/6$ can be inferred using the fact that a mode still out of the Hubble radius today cannot affect a physical measurement. The angular dependent factor $-\cos(2\theta)$ can be inferred by looking at the effect of a long wavelength lensing mode on the power spectrum, as explained in section 3.4.5.

A remark on the angular dependence in eq. (3.94) is in order here. Although it is non-vanishing in the squeezed limit, the lensing contribution (3.88) is not of the local form (3.51). In particular, as the angular dependence averages to zero, a non-Gaussianity test based on a local estimator of the form (3.51) would be almost blind to the lensing signal. A quantitative way to measure how a signal overlaps with another is provided by the cosine between two bispectra, defined as [34]

$$\cos(B_1, B_2) \equiv \frac{B_1 \cdot B_2}{\sqrt{B_1 \cdot B_1}\sqrt{B_2 \cdot B_2}}\,, \tag{3.95}$$

where $B_1 \cdot B_2$ is the scalar product between two bispectra, given by

$$B_1 \cdot B_2 \equiv \frac{1}{\pi}\int \frac{{\rm d}^2 l_2 {\rm d}^2 l_3}{(2\pi)^2}\frac{B_1(\vec{l}_1,\vec{l}_2,\vec{l}_3)\,B_2(\vec{l}_1,\vec{l}_2,\vec{l}_3)}{6C_{l_1}C_{l_2}C_{l_3}} \tag{3.96}$$

$$\propto \int {\rm d}r_2 {\rm d}r_3 \frac{r_2^3 r_3^3 B_1(1,r_2,r_3)\,B_2(1,r_2,r_3)}{\left(2r_2^2+2r_3^2+2r_2^2r_3^2-1-r_2^4-r_3^4\right)^{1/2}}\,. \tag{3.97}$$

Indeed, we find that the cosine between the lensing bispectrum (3.88) and the local bispectrum (3.51) is $\cos(B_{\rm lens}, B_{\rm local}) = 0.03$.[1] For instance, one can compare

[1] Note that the scalar product with the local bispectrum is logarithmically divergent for $r_2 \to 1$ or $r_3 \to 1$. Thus, in order to evaluate it we have put the cutoff $r_{\rm max} = 0.999$.

Shape:	total	local	equil	lens
total	1.00	-0.17	0.41	0.98
local		1.00	0.30	0.03
equil			1.00	0.47
lens				1.00

Table 3.1: Cosines between different shapes of bispectra.

this to the cosine between the local and equilateral bispectra, which is much larger, $\cos(B_{\rm equil}, B_{\rm local}) = 0.30$. Thus, due to the angular dependence of the squeezed limit, the lensing signal is orthogonal to the local one. We can now compare the total bispectrum to the local one. The cosine is $\cos(B_{\rm total}, B_{\rm local}) = -0.17$. Thus, as it is dominated by lensing, the total bispectrum is almost orthogonal to the local signal. However, due to the term $-1/6$ in eq. (3.94) the orthogonality is not complete and the total bispectrum slightly overlaps with the local one.

In the equilateral limit all the five contributions to the total bispectrum become important. However, the lensing numerically dominates. In this limit we can compare the total bispectrum to the equilateral bispectrum (3.54) taken with $f_{\rm NL}^{\rm equil} = 1$. This yields

$$\frac{B^{\rm total}(1,1,1)}{B^{\rm equil}(1,1,1)} = 3.13\,. \tag{3.98}$$

Thus, the total bispectrum corresponds to $f_{\rm NL}^{\rm equil} = 3.13$. As it is not vanishing in the squeezed limit, its cosine with the equilateral shape will be smaller than unity. Indeed we find $\cos(B_{\rm total}, B_{\rm equil}) = 0.41$. Note that this value is larger than the cosine between local and equilateral shapes, i.e. 0.30. Thus, the total bispectrum is "more equilateral" than the local one. Finally, to have a confirmation that the lensing effect dominates the total bispectrum, we can compute the cosine between the total signal and the lensing. This is $\cos(B_{\rm total}, B_{\rm lens}) = 0.98$, which is very close to one, as expected. A summary of the cosines is given in table 3.1.

It is important to stress that the shape associated with lensing, with an angle dependent squeezed limit, represents another interesting template for the bispectrum besides the local, the equilateral and the ones studied in [36; 92]. As it is rather orthogonal to the standard local and equilateral templates, in the future it would be interesting to put limits on it, even independently of lensing.

3.6 The form of the bispectrum in the squeezed limit.

In the previous sections we have computed the physical limit in which an observer far in the future in a completely matter dominated universe observes the CMB and uses the flat-sky approximation to compute the contribution of second-order gravitational effects

to the bispectrum. We will now try to argue that the form of the bispectrum which we get in the squeezed limit will be the same even when considering the contribution of plasma physics. This should be a quantity that can be compared with numerical codes currently being developed by other groups (see for example [41; 42; 72]).

Here, we wish to present a simple argument supporting the result that, in the squeezed limit during pure matter domination, the non-gaussianity induced by second order effects will be given by

$$f_{\rm NL}^{\rm loc} \approx -\frac{1}{6} - \cos(2\theta)\,. \tag{3.99}$$

We will do this by analyzing the effect that the presence of a long-wavelength mode has on the two-point function of two short-wavelength modes.

The question can be stated in the following way: How do we connect the temperature two-point function we ($\mathcal{O}_o$) observe today to the temperature two-point function observed by an hypothetical observer ($\mathcal{O}_f$) near the moment of recombination (*i.e.* before the long-wavelength mode enters the horizon)? If we work under the likely assumption that temperature fluctuations as observed by us and by $\mathcal{O}_f$ do not change, there will be two effects: The mean temperature measured at $\mathcal{O}_o$ will be different from the one measured at $\mathcal{O}_f$ since it will be modified by the long-wavelength mode of the potential:

$$\bar{T}_o = \left(1 - \frac{2}{3}\phi_L\right)\bar{T}_f\,. \tag{3.100}$$

Furthermore, the trajectory of the photons will be changed by the long-wavelength potential, which induces a lensing effect:

$$\vec{\alpha} = -2\int_{\tau_e}^{\tau_o} \frac{\tau - \tau_e}{\tau_o - \tau_e}\nabla_{\parallel}\phi_L\,. \tag{3.101}$$

We will ignore for now this last effect and consider it separately at the end of this section.

To compute the modification of the two-point function in the presence of the long-wavelength mode we use equation (3.100) and the redshift of the temperature $T_o/T_e = (\Phi_e - \Phi_o)$ to write

$$\left\langle \frac{\delta T}{\bar{T}_o}\frac{\delta T}{\bar{T}_o}\right\rangle = \left(1 + \frac{2}{3}\phi_L\right)\left\langle \frac{\delta T}{\bar{T}_f}\frac{\delta T}{\bar{T}_f}\right\rangle. \tag{3.102}$$

We can now relate two short-wavelength modes with a long-wavelength mode in order to compute the contribution of second order gravitational effects to the observed local non-gaussianity:

$$\begin{aligned}\left\langle \frac{\delta T}{\bar{T}}\bigg|_L \frac{\delta T}{\bar{T}}\bigg|_S \frac{\delta T}{\bar{T}}\bigg|_S\right\rangle &= \frac{2}{3}\left\langle \frac{\delta T}{\bar{T}}\bigg|_L \phi_L\right\rangle\left\langle \frac{\delta T}{\bar{T}_f}\frac{\delta T}{\bar{T}_f}\right\rangle \\ &= 2\left\langle \frac{\delta T}{\bar{T}}\bigg|_L \frac{\delta T}{\bar{T}}\bigg|_L\right\rangle\left\langle \frac{\delta T}{\bar{T}}\bigg|_S \frac{\delta T}{\bar{T}}\bigg|_S\right\rangle.\end{aligned} \tag{3.103}$$

In the second line we used the leading order expression for each quantity.

In order to prove that the equivalent non-gaussianity parameter is $f_{\rm NL}^{\rm local} = -1/6$ compare it with $\delta T/\bar{T} = (1/3)(\phi - f_{\rm NL}^{\rm equiv}\phi^2)$ which gives

$$\left\langle \frac{\delta T}{\bar{T}}\Big|_L \frac{\delta T}{\bar{T}}\Big|_S \frac{\delta T}{\bar{T}}\Big|_S \right\rangle = -4\left\langle \frac{1}{3}\phi_L \frac{1}{3} f_{\rm NL}^{\rm equiv}\phi_L\phi_S \frac{1}{3}\phi_S \right\rangle$$
$$= -12 f_{\rm NL}^{\rm equiv}\left\langle \frac{\delta T}{\bar{T}}\Big|_L \frac{\delta T}{\bar{T}}\Big|_L \right\rangle \left\langle \frac{\delta T}{\bar{T}}\Big|_S \frac{\delta T}{\bar{T}}\Big|_S \right\rangle . \qquad (3.104)$$

Let us now turn to the lensing 3-point function in the squeezed limit. We are going to calculate the 3-point function by first taking the long wavelength mode fixed and then studying its lensing effect on the short scale 2-point function.[1] At the end we average over the long wavelength mode.[2] Consider the 2-point correlation function of the temperature fluctuations in two different directions $\hat{n}_1$ and $\hat{n}_2$. In the presence of a long wavelength mode the real space 2-point function is lensed

$$\left\langle \frac{\delta T}{T}(\hat{n}_1)\frac{\delta T}{T}(\hat{n}_2)\right\rangle_{\rm lens} = \left\langle \frac{\delta T}{T}\frac{\delta T}{T}\right\rangle [\hat{n}_1 + \vec{\alpha}(\hat{n}_1) - \hat{n}_2 - \vec{\alpha}(\hat{n}_2)] , \qquad (3.105)$$

where we used the fact that the unlensed 2-point function just depends on the distance between the points. Obviously there is no effect if the two lensing angles are the same: the 2-point function is just translated. Expanding at first order and defining by $\vec{m}_1$ and $\vec{m}_2$ the components of $\hat{n}_1$ and $\hat{n}_2$ parallel to the (flat) sky we have

$$\left\langle \frac{\delta T}{T}(\hat{n}_1)\frac{\delta T}{T}(\hat{n}_2)\right\rangle_{\rm lens} = \left\langle \frac{\delta T}{T}\frac{\delta T}{T}\right\rangle [\vec{m}_1 - \vec{m}_2]$$
$$+ \nabla_i \left\langle \frac{\delta T}{T}\frac{\delta T}{T}\right\rangle \nabla_j\alpha_i \left[\frac{\vec{m}_1 + \vec{m}_2}{2}\right] \cdot (\vec{m}_1 - \vec{m}_2)_j . \qquad (3.106)$$

By assumption the lensing wave is of long wavelength so that we can evaluate the gradient of the lensing angle at the midpoint $(\vec{m}_1 + \vec{m}_2)/2$. If we call $\vec{m} \equiv \vec{m}_1 - \vec{m}_2$, we have

$$\left\langle \frac{\delta T}{T}(\hat{n}_1)\frac{\delta T}{T}(\hat{n}_2)\right\rangle_{\rm lens} = \left\langle \frac{\delta T}{T}\frac{\delta T}{T}\right\rangle [m]$$
$$+ \frac{\rm d}{{\rm d}\log m}\left\langle \frac{\delta T}{T}\frac{\delta T}{T}\right\rangle [m]\, \frac{m_j}{m}\frac{m_i}{m}\, \nabla_j\alpha_i \left[\frac{\vec{m}_1 + \vec{m}_2}{2}\right] . \qquad (3.107)$$

[1] It is easy to argue that the leading contribution in the squeezed limit is obtained when the lensing mode is of long wavelength. Indeed, lensing is effective far from the last scattering surface, but as we get far from it the correlation with the temperature fluctuation rapidly decreases. The loss of correlation happens at $\tau_* \sim \tau_o/l$, i.e. it is faster at high l, that is why the squeezed limit is dominated by a long lensing wave.

[2] This discussion is inspired by the derivation of the consistency relation for the squeezed limit of the primordial 3-point function in single field inflation [4; 28; 29]. In particular we will parallel the explicit derivation done in sec. 2 of [29].

We can now Fourier transform to $\vec{l}_1$ and $\vec{l}_2$. The result can be expressed in terms of $\vec{l}_S = (\vec{l}_1 - \vec{l}_2)/2$ and $\vec{l}_L = \vec{l}_1 + \vec{l}_2$, where $_L$ and $_S$ stand for long and short wavelength,

$$\langle a_{\vec{l}_1} a_{\vec{l}_2} \rangle_{\rm lens} = C_{l_S} + i l_{Lj} \alpha_i(\vec{l}_L) \int {\rm d}^2 m \frac{\rm d}{{\rm d}\log m} \left\langle \frac{\delta T}{T} \frac{\delta T}{T} \right\rangle [m] \, \frac{m_j}{m} \frac{m_i}{m} \, e^{-i\vec{l}_S \vec{m}} \, . \quad (3.108)$$

The 3-point function is obtained multiplying the above expression by $\delta T/T$ of the long wavelength mode and averaging,

$$\begin{aligned} \langle a_{\vec{l}_1} a_{\vec{l}_2} a_{\vec{l}_3} \rangle = (2\pi)^2 \delta(\vec{l}_1 + \vec{l}_2 + \vec{l}_3) \cdot i l_{Lj} \langle \frac{\delta T}{T} \alpha_i \rangle'(l_L) \\ \times \int {\rm d}^2 m \frac{\rm d}{{\rm d}\log m} \left\langle \frac{\delta T}{T} \frac{\delta T}{T} \right\rangle [m] \, \frac{m_j}{m} \frac{m_i}{m} \, e^{-i\vec{l}_S \vec{m}} \, . \end{aligned} \quad (3.109)$$

The prime in the correlation between lensing and the temperature means that we have to remove the momentum conservation factor $(2\pi)^2\delta$, which has been factored out.

Let us evaluate the integral over $\vec{m}$, which describes the effect of lensing on the 2-point function. One may naïvely think that for a scale invariant 2-point function, which is the case that we are studying in this chapter, the effect of lensing vanishes. Indeed, the calculations above are very similar to the ones leading to the consistency relation for the squeezed limit of the primordial 3-point function [4; 28; 29]. In that case, however, the integral over $\vec{m}$ does not contain the angular weight $m^j m^i/m^2$. Without this terms the integral vanishes for a scale invariant spectrum: indeed the 2-point function in real space is a logarithm of the distance, so that its log-derivative is a constant. The Fourier transform of a constant is $\delta(\vec{l}_S)$ which vanishes for any non-zero $\vec{l}_S$.

The situation is different in the presence of the angular weight $m^j m^i/m^2$. To be more explicit, let us introduce a scale dependence in the 2-point function and evaluate the integral in eq. (3.109) for a power spectrum of the form $C_l = C \cdot l^{-2+(n_s-1)}$, which corresponds to a 2-point function going as $m^{-(n_s-1)}$, to see that the result does not vanish for $n_s \to 1$. The integral can be written as

$$\begin{aligned} \int {\rm d}^2 m \frac{\rm d}{{\rm d}\log m} \left\langle \frac{\delta T}{T} \frac{\delta T}{T} \right\rangle [m] \, \frac{m_j}{m} \frac{m_i}{m} \, e^{-i\vec{l}_S \vec{m}} \\ &= -(n_s - 1) \frac{\partial_{l_i} \partial_{l_j}}{\nabla^2} \int {\rm d}^2 m \left\langle \frac{\delta T}{T} \frac{\delta T}{T} \right\rangle e^{-i\vec{l}_S \vec{m}} \\ &= -(n_s - 1) \frac{\partial_{l_i} \partial_{l_j}}{\nabla^2} C \cdot l^{-2+(n_s-1)} \\ &= -(n_s - 1) \, \partial_{l_i} \partial_{l_j} C \cdot \frac{l^{n_s - 1}}{(n_s - 1)^2} \\ &= -C \cdot l^{-2+(n_s-1)} \left[\delta_{ij} + (n_s - 3) \frac{l_i l_j}{l^2} \right] \, . \end{aligned} \quad (3.110)$$

We see that the result does not vanish for $n_s = 1$. What vanishes for $n_s = 1$ is the trace of this tensor. This means that for a scale invariant spectrum, the isotropic rescaling due to lensing does not contribute to the 3-point function. This makes sense

in light of the discussion above: for the isotropic part there is no angular weight so that everything works as for the consistency relation for the squeezed limit of the primordial 3-point function [4; 28; 29]. On the other hand, the anisotropic case is similar to what happens when one calculates the primordial 3-point function of a graviton and two scalar modes, in the limit when the graviton wavelength becomes very long. The gravitational wave induces an anisotropic rescaling of the scalar 2-point function and the result does not vanish for a scale invariant spectrum [4]. An analogous effect is found when computing the contribution to the scalar trispectrum from graviton exchange [93]. In the limit where the graviton wavelength is very long, the non-isotropic rescaling induces a correlation between a pair of scalar 2-point functions. This effect has the same spin-2 angular dependence as the lensing.

Let us go back to eq. (3.109). In our case the normalization of the spectrum is given by $C = A/(9\pi)$, so that the expression of the 3-point function in the squeezed limit gives

$$\langle a_{\vec{l}_1} a_{\vec{l}_2} a_{\vec{l}_3} \rangle = (2\pi)^2 \delta(\vec{l}_1 + \vec{l}_2 + \vec{l}_3) \cdot i l_{Lj} \langle \frac{\delta T}{T} \alpha_i \rangle'(l_L) \left(-\frac{A}{9\pi} \right) \frac{1}{l_S^2} \left(\delta_{ij} - 2 \frac{l_{Si} l_{Sj}}{l_S^2} \right) . \quad (3.111)$$

The correlation between the temperature and the deflection angle is given by

$$\begin{aligned} \langle \frac{\delta T}{T} \alpha_i \rangle'(l_L) &= -\frac{2}{3} \frac{1}{D_e^2} \int \frac{\mathrm{d}k_\perp}{2\pi} \int_{\tau_e}^{\tau_o} \mathrm{d}\tau \, \frac{\tau - \tau_e}{D_e} \frac{i l_{Li}}{D_e} \frac{A}{(k_\perp^2 + l_L^2/D_e^2)^{3/2}} e^{i k_\perp (\tau - \tau_e)} \\ &= -\frac{1}{3\pi} \cdot \frac{4A \, i l_{Li}}{l_L^4} . \end{aligned} \quad (3.112)$$

Thus we have

$$B^{\mathrm{lens}} = -\frac{4A^2}{27\pi^2} \frac{l_{Li} l_{Lj}}{l_L^4 l_S^2} \left(\delta_{ij} - 2 \frac{l_{Si} l_{Sj}}{l_S^2} \right) . \quad (3.113)$$

In the limit $\vec{l}_2 \to 0$, the explicit expression (3.89) gives, taking into account the permutation $l_1 \leftrightarrow l_3$,

$$B^{\mathrm{lens}} = \frac{4A^2}{27\pi^2} \frac{1}{l_2^4} \left[\frac{\vec{l}_1 \cdot \vec{l}_2}{l_1^2} - \frac{\vec{l}_2 \cdot (\vec{l}_1 + \vec{l}_2)}{(\vec{l}_1 + \vec{l}_2)^2} \right] \simeq -\frac{4A^2}{27\pi^2} \frac{l_{2i} l_{2j}}{l_2^4} \frac{\mathrm{d}}{\mathrm{d}l_{1j}} \frac{l_{1i}}{l_1^2} , \quad (3.114)$$

which coincides with the expression of section 3.4.5.

3.7 Conclusion

In this chapter we have calculated, assuming perfect matter dominance, the complete CMB bispectrum on large angular scales, larger than the Hubble radius at recombination, considering for the first time all the relevant effects. Although our results give the exact bispectrum in a well defined physical limit, there are many ways to improve our calculations to make them closer to the real universe. One should include the recent dark energy domination and the early transition from radiation to matter dominance

along the lines of [50]. This will give qualitative new phenomena, like the rather large ISW-lensing correlation [37]. Given that we are on large angular scales, a full-sky treatment would be more precise than our flat-sky expressions, although the results for the bispectrum will be much more complicated and difficult to understand. Finally, the small deviation from a scale invariant spectrum should be included. Taking all this into account would give the correct prediction for our universe of the large angle bispectrum. This is clearly far from the complete answer. The modes on scales larger than the horizon at recombination are quite few and most of the bispectrum signal comes from triangles with modes on sub-Hubble scales. Entering in a sub-Hubble regime requires the whole machinery of second-order Boltzmann equations that we have not touched in this chapter.

The calculated bispectrum is rather small: the final bispectrum is dominated by the lensing contribution, which gives $f_{\rm NL}^{\rm local} = -\cos(2\theta)$, with θ the angle between long and short modes. Even if we could use our results on arbitrarily short scales, this would be below Planck sensitivity, limited to $f_{\rm NL}^{\rm local} \sim 5$. This means that the bispectrum in the Sachs-Wolfe limit does not represent a relevant contamination for the forthcoming searches for primordial non-Gaussianities.

A way to go beyond the large angle regime is to correct the results of [77] to get the full bispectrum in the squeezed limit, with one (but not necessarily all) of the modes on scales larger than the horizon at recombination. We leave all these directions for future work, though a first step in this direction was given in section 3.6.

3.A Appendix: Flat-sky and integrated effects

At first order, the gravitational contribution to the temperature anisotropies in matter domination is the Sachs-Wolfe effect,

$$\frac{\delta T}{T}(\hat{n}) = \frac{1}{3}\phi(\hat{n}D_e) = \frac{1}{3}\int \frac{\mathrm{d}^3k}{(2\pi)^3} e^{i\vec{k}\cdot\hat{n}D_e}\phi_{\vec{k}}\,, \tag{3.115}$$

where $\hat{n}$ is the unit vector specifying the line of sight direction, $D_e = \tau_o - \tau_e$ is the (background) conformal distance to the last scattering surface and ϕ is the first order Newtonian potential. In the flat-sky formalism [83; 84], one chooses a fiducial direction $\hat{z}$ and expands at the lowest order in the angle θ between $\hat{z}$ and $\hat{n}$:

$$\hat{n} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta) \simeq (m_x, m_y, 1)\,, \tag{3.116}$$

$\vec{m}$ being a 2-dimensional vector normal to $\hat{z}$. The multipole is simply the 2-dimensional Fourier transform with respect to $\vec{m}$:

$$a_{\vec{l}} = \int \mathrm{d}^2m\; e^{-i\vec{l}\cdot\vec{m}}\frac{\delta T}{T}(\hat{n}) = \frac{1}{3}\int \frac{\mathrm{d}^3k}{2\pi}\delta(\vec{l} - \vec{k}_\parallel D_e)e^{ik_\perp D_e}\phi_{\vec{k}}\,. \tag{3.117}$$

One can show that the flat-sky multipole corresponds to the large l limit of the full-sky one. The two are related by [84]

$$a_{\vec{l}} = \sqrt{\frac{4\pi}{2l+1}} \sum_m i^{-m} a_{lm} e^{im\varphi_l} \,, \tag{3.118}$$

$$a_{lm} = \sqrt{\frac{2l+1}{4\pi}} i^m \int \frac{\mathrm{d}\varphi_l}{2\pi} e^{-im\varphi_l} a_{\vec{l}} \,. \tag{3.119}$$

Similar expressions hold also for the power spectrum and the bispectrum. The power spectrum is defined as $\langle a_{\vec{l}_1} a_{\vec{l}_2} \rangle \equiv (2\pi)^2 \delta(\vec{l}_1 + \vec{l}_2) C_{l_1}^{\mathrm{flat}}$ in flat-sky approximation, and as $\langle a_{l_1 m_1} a_{l_2 m_2} \rangle \equiv \delta_{m_1 m_2} \delta_{l_1 l_2} C_{l_1}^{\mathrm{full}}$ in full sky; the two expressions are related by $C_l^{\mathrm{full}} \approx C_l^{\mathrm{flat}}$ for large l. The bispectrum in the full and flat sky are defined respectively as

$$\langle a_{\vec{l}_1} a_{\vec{l}_2} a_{\vec{l}_3} \rangle \equiv (2\pi)^2 \delta(\vec{l}_1 + \vec{l}_2 + \vec{l}_3) B(\vec{l}_1, \vec{l}_2, \vec{l}_3) \,, \tag{3.120}$$

$$\langle a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3} \rangle \equiv \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} B_{l_1 l_2 l_3} \,, \tag{3.121}$$

where $\left(\begin{smallmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{smallmatrix}\right)$ is the Wigner 3-j symbol. The two expressions are related by:

$$B_{l_1 l_2 l_3} \approx \begin{pmatrix} l_1 & l_2 & l_3 \\ 0 & 0 & 0 \end{pmatrix} \sqrt{\frac{(2l_1+1)(2l_2+1)(2l_3+1)}{4\pi}} B(\vec{l}_1, \vec{l}_2, \vec{l}_3) \,. \tag{3.122}$$

The derivation of these expressions can be found in [84].

To better understand what happens when we correlate effects which are important at different times, we can do a simple exercise[1]: we calculate the 2-point function of two integrated effects which peak at different times τ_1 and τ_2. We will see that the correlation decays exponentially when $\tau \gtrsim (\tau_2 - \tau_1)/l$, and that the power spectrum is proportional to $\delta(\vec{l}_1 + \vec{l}_2)$ up to exponentially small terms. Consider a generic integrated effect at first order:

$$a_{\vec{l}} = \int_{\tau_e}^{\tau_o} \mathrm{d}\tau \int \frac{\mathrm{d}^3 k}{2\pi} \delta(\vec{l} - \vec{k}_{\parallel}(\tau_o - \tau)) e^{ik_\perp(\tau_o - \tau)} g'(\tau) \phi_{\vec{k}} \,, \tag{3.123}$$

where $g(\tau)$ is a growth function. Now we correlate two such effects, with different growth functions $g(\tau)$ and $f(\tau)$:

$$\begin{aligned} \langle a^f_{\vec{l}_1} a^g_{\vec{l}_2} \rangle &= 4\pi \int_{\tau_e}^{\tau_o} \mathrm{d}\tau_a \, f'(\tau_a) \int_{\tau_e}^{\tau_o} \mathrm{d}\tau_b \, g'(\tau_b) \int \mathrm{d}^3 k \delta(\vec{l}_1 - \vec{k}_{\parallel}(\tau_o - \tau_a)) \\ &\quad \times \delta(\vec{l}_2 + \vec{k}_{\parallel}(\tau_o - \tau_b)) e^{ik_\perp(\tau_b - \tau_a)} \frac{A}{k^3} \\ &= 4\pi \int_{\tau_e}^{\tau_o} \mathrm{d}\tau_a \frac{f'(\tau_a)}{(\tau_o - \tau_a)^2} \int_{\tau_e}^{\tau_o} \mathrm{d}\tau_b \, g'(\tau_b) \int \mathrm{d}k_\perp \delta\left(\vec{l}_1 + \vec{l}_2 + \vec{l}_1 \frac{\tau_a - \tau_b}{\tau_o - \tau_a}\right) \\ &\quad \times e^{ik_\perp(\tau_b - \tau_a)} \frac{A}{(k_\perp^2 + l_1^2/(\tau_o - \tau_a)^2)^{3/2}} \,. \end{aligned} \tag{3.124}$$

[1] We thank F. Bernardeau for suggesting this example.

For simplicity, we approximate the growth functions with step functions, such that

$$f'(\tau) \sim \delta(\tau - \tau_1)\,, \qquad g'(\tau) \sim \delta(\tau - \tau_2)\,, \tag{3.125}$$

where we consider $\tau_e \leq \tau_1 \leq \tau_2 < \tau_o$. Thus we find

$$\langle a^f_{\vec{l}_1} a^g_{\vec{l}_2} \rangle = \frac{4\pi A}{(\tau_o - \tau_1)^2} \delta\left(\vec{l}_1 + \vec{l}_2 + \vec{l}_1 \frac{\tau_1 - \tau_2}{\tau_o - \tau_1}\right) \times \int \mathrm{d}k_\perp e^{ik_\perp(\tau_2 - \tau_1)} (k_\perp^2 + l_1^2/(\tau_o - \tau_1)^2)^{-3/2}\,. \tag{3.126}$$

The integration over $k_\perp$ can be done analytically, yielding

$$\langle a^f_{\vec{l}_1} a^g_{\vec{l}_2} \rangle = (2\pi)^2 \delta\left(\vec{l}_1 + \vec{l}_2 + \vec{l}_1 \frac{\tau_1 - \tau_2}{\tau_o - \tau_1}\right) \frac{2}{\pi} \frac{|\tau_2 - \tau_1|}{(\tau_o - \tau_1)} \frac{A}{l_1} K_1\left(l_1 \frac{|\tau_2 - \tau_1|}{\tau_o - \tau_1}\right)\,, \tag{3.127}$$

where K_1 is the modified Bessel function, with asymptotic behaviours $K_1(x) \to 1/x$ for $x \ll \sqrt{2}$ and $K_1(x) \to \sqrt{\pi/2x}\, e^{-x}$ for $x \gg 3/4$.

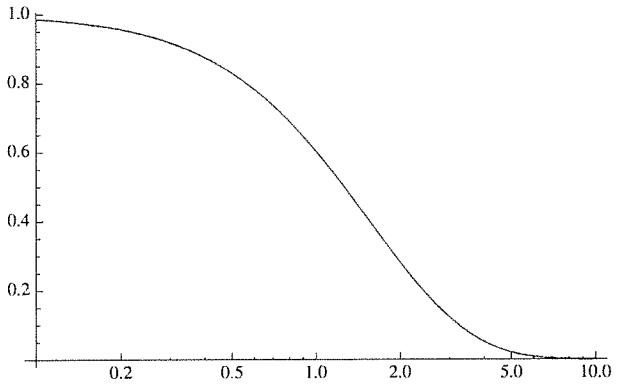


Figure 3.9: The function $xK_1(x)$ with the x axis in logarithmic scale.

We can see that, when $l_1 \frac{\tau_2 - \tau_1}{\tau_o - \tau_1} \gtrsim 1$, the correlation is exponentially suppressed. Then, in the limit $l_1 \ll (\tau_o - \tau_1)/(\tau_2 - \tau_1)$ we find

$$\langle a^f_{\vec{l}_1} a^g_{\vec{l}_2} \rangle = (2\pi)^2 \delta(\vec{l}_1 + \vec{l}_2) \frac{2}{\pi} \frac{A}{l_1^2}\,. \tag{3.128}$$

In general, the translational invariance in 2d is only approximate; however, the approximation is very good since for large multipoles the correlations are exponentially suppressed if the sum $\sum_i \vec{l}_i \neq 0$.

3.B Appendix: Detailed calculation of the Rees-Sciama effect

In this appendix we compute the Rees-Sciama bispectrum. We start from eq. (3.67) before summing over cyclic permutations:

$$B^{\rm RS} = -\frac{A^2}{378\pi^2}\frac{1}{l_1^4}\int_0^\infty \mathrm{d}x\, x \int_{-\infty}^{+\infty} \mathrm{d}y_1\mathrm{d}y_2\, e^{i(y_1+y_2)x}(y_1^2+r_1^2)^{-3/2}(y_2^2+r_2^2)^{-3/2} \times \left[\frac{3}{2}r_1^2+\frac{3}{2}r_2^2+r_3^2+2y_1y_2+\frac{5}{2}(y_1^2+y_2^2)-\frac{5}{2}\frac{(y_1^2-y_2^2+r_1^2-r_2^2)^2}{(y_1+y_2)^2+r_3^2}\right]. \quad (3.129)$$

Since there are some pieces in the kernel that can be integrated analytically, we now compute them as a check of our numerical integration. To proceed, we make use of the following known integrals:

$$\int_{-\infty}^{+\infty}\mathrm{d}y\frac{e^{iyx}}{(y^2+a^2)^{\frac{3}{2}}} = \frac{2x}{a}K_1(ax)\,, \quad (3.130)$$

$$\int_{-\infty}^{+\infty}\mathrm{d}y e^{iyx}\frac{y}{(y^2+a^2)^{\frac{3}{2}}} = -i\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{2x}{a}K_1(ax)\right) = 2ixK_0(ax)\,, \quad (3.131)$$

$$\int_{-\infty}^{+\infty}\mathrm{d}y e^{iyx}\frac{y^2}{(y^2+a^2)^{\frac{3}{2}}} = -i\frac{\mathrm{d}}{\mathrm{d}x}\left(2ixK_0(ax)\right) = 2\left[K_0(ax)-axK_1(ax)\right], \quad (3.132)$$

where the K_i are Bessel modified functions. We then split the bispectrum into four pieces, three of which are integrated analytically:

$$B^{(1)} = -\frac{2A^2}{189\pi^2}\frac{1}{l_1^4}\left(\frac{3}{2}r_1^2+\frac{3}{2}r_2^2+r_3^2\right)\frac{1}{r_1r_2}\int_0^\infty \mathrm{d}x\, x^3K_1(r_1x)K_1(r_2x)\,, \quad (3.133)$$

$$B^{(2)} = \frac{4A^2}{189\pi^2}\frac{1}{l_1^4}\int_0^\infty \mathrm{d}x\, x^3K_0(r_1x)K_0(r_2x)\,, \quad (3.134)$$

$$B^{(3)} = -\frac{5A^2}{189\pi^2}\frac{1}{l_1^4}\int_0^\infty \mathrm{d}x\, x\left[x\left(\frac{1}{r_1}K_1(r_1x)K_0(r_2x)+\frac{1}{r_2}K_0(r_1x)K_1(r_2x)\right) - x^2\left(\frac{r_2}{r_1}+\frac{r_1}{r_2}\right)K_1(r_1x)K_1(r_2x)\right], \quad (3.135)$$

$$B^{(4)} = \frac{5A^2}{756\pi^2}\frac{1}{l_1^4}\int_0^\infty \mathrm{d}x\, x\int_{-\infty}^{+\infty}\mathrm{d}y_1\mathrm{d}y_2\, e^{i(y_1+y_2)x} \times (y_1^2+r_1^2)^{-\frac{3}{2}}(y_2^2+r_2^2)^{-\frac{3}{2}}\frac{(y_1^2-y_2^2+r_1^2-r_2^2)^2}{(y_1+y_2)^2+r_3^2}\,. \quad (3.136)$$

The first three pieces can be integrated in x, giving:

$$B^{(1)} = -\frac{4A^2}{189\pi^2}\frac{1}{l_1^4}\left(\frac{3}{2}r_1^2+\frac{3}{2}r_2^2+r_3^2\right) \times \frac{1}{r_1^2r_2^2(r_2^2-r_1^2)^3}\left[r_2^4-r_1^4-4r_1^2r_2^2\ln\frac{r_2}{r_1}\right], \quad (3.137)$$

$$B^{(2)} = \frac{16A^2}{189\pi^2}\frac{1}{l_1^4}\frac{1}{(r_2^2-r_1^2)^3}\left[r_1^2 - r_2^2 - (r_1^2+r_2^2)\ln\frac{r_1}{r_2}\right], \tag{3.138}$$

$$B^{(3)} = -\frac{5A^2}{189\pi^2}\frac{1}{l_1^4}\frac{1}{(r_1^2-r_2^2)^3}\left[5(r_2^2-r_1^2) + \frac{r_2^4}{r_1^2} - \frac{r_1^4}{r_2^2} + 2(r_1^2-r_2^2)\ln\frac{r_1}{r_2} - 2(3r_1^2+5r_2^2)\ln\frac{r_2}{r_1}\right]. \tag{3.139}$$

The fourth piece (3.136) cannot be integrated analytically, making the numerical integration necessary. However, comparison between the numerical integration of the other three pieces and the analytical expressions (3.137) to (3.139) gives consistent results. This provides a check of the validity of our computation.

Now we turn to the numerical integration of eq. (3.129). To do it we first have to perform analytically the x integral, which is ill-defined. In order to overcome this problem we first change variables from y_1, y_2 to $y_+ \equiv y_1 + y_2$ and $y_- \equiv y_1 - y_2$, and then regularize the integral in the following way:

$$\begin{aligned}\int_0^\infty \mathrm{d}x \int_{-\infty}^{+\infty} \mathrm{d}y_+ \int_{-\infty}^{+\infty} \mathrm{d}y_-\, x e^{iy_+x} f(y_+,y_-) &= -i\int_{-\infty}^{+\infty}\mathrm{d}y_+\int_{-\infty}^{+\infty}\mathrm{d}y_-\int_0^\infty \mathrm{d}x f(y_+,y_-)\frac{\partial}{\partial y_+}e^{iy_+x}\\ &= i\int_{-\infty}^{+\infty}\mathrm{d}y_+\int_{-\infty}^{+\infty}\mathrm{d}y_-\frac{\partial}{\partial y_+}f(y_+,y_-)\int_0^\infty \mathrm{d}x\, e^{iy_+x}\\ &= -\int_{-\infty}^{+\infty}\mathrm{d}y_+\int_{-\infty}^{+\infty}\mathrm{d}y_-\frac{1}{y_+}\frac{\partial}{\partial y_+}f(y_+,y_-)\,. \end{aligned}\tag{3.140}$$

The last integral follows from the prescription

$$\int_0^\infty \mathrm{d}x e^{iy_+x}e^{-\varepsilon x} = \frac{1}{-iy_+ + \varepsilon}\,. \tag{3.141}$$

After integrating in x, we obtain:

$$B^{\mathrm{RS}} = \frac{A^2}{378\pi^2}\frac{1}{l_1^4}\int_{-\infty}^{+\infty}\mathrm{d}y_1\mathrm{d}y_2\,\frac{1}{y_+}\frac{\partial}{\partial y_+}\Bigg\{(y_1^2+r_1^2)^{-3/2}(y_2^2+r_2^2)^{-3/2} \times\left[\frac{3}{2}r_1^2+\frac{3}{2}r_2^2+r_3^2+2y_1y_2+\frac{5}{2}(y_1^2+y_2^2)-\frac{5}{2}\frac{(y_1^2-y_2^2+r_1^2-r_2^2)^2}{(y_1+y_2)^2+r_3^2}\right]\Bigg\}, \tag{3.142}$$

which, after changing variables from (y_1, y_2) to (y_+, y_-), and performing the derivative, gives a form which can be integrated numerically. The final results of the integration, after summing over cyclic permutations and setting $r_1 = 1$, are presented in figure 3.4.

We can compare the Rees-Sciama contribution with the local shape in the squeezed limit. We will see that while the local shape diverges as $1/r^2$ in this limit, the Rees-Sciama only diverges as $1/r$. Going back to eq. (3.129), we can study the behavior when one of the r goes to zero; notice that the expression must be symmetrized so that we have to study both the limits $r_2 \to 0$ and $r_3 \to 0$ in eq. (3.129). For $r_2 \to 0$ we have an infrared divergence in the y_2 integral coming from the power spectrum which goes as y_2^{-3} for $r_2 = 0$. This would give a divergence r_2^{-2} as in the local model. However, for $r_2 = 0$ and $r_1 = r_3 = 1$ the expression in brackets in the second line of (3.129) goes as y_2 for $y_2 \to 0$, but its integral vanishes due to parity, leaving only terms which are at most logarithmic divergent and thus suppressed with respect to the local shape. An additional divergence comes from the limit $r_3 \to 0$ in eq. (3.129); in this case the integral diverges in the limit $y_1 + y_2 \to 0$. Notice that in this case one also has to take into account the integral over x which diverges for $y_1 + y_2 = 0$. To study the behavior for $r_3 \to 0$ one must first integrate in x using the prescription (3.140). One can see that the leading divergence in the resulting expression comes from a term of the form r_3^2/y_+^4, which gives a $1/r_3$ divergence. This is dominant compared to the divergence in r_2, but it is still subdominant compared to the local case. We conclude that the Rees-Sciama result is subdominant compared to the local shape in the squeezed limit. This analysis is a good check of the numerics, which indeed shows a $1/r$ divergence in the squeezed limit.

4

An Improved Calculation of the Non-Gaussian Halo Mass Function

4.1 Introduction

As explained in the introduction, the primordial curvature inhomogeneities, generated by the inflationary mechanism, obey a statistics which is nearly Gaussian. The deviations from Gaussianity, while expected to be small, provide a unique window into the physics of inflation. For example, single-field slow-roll models of inflation lead to a small level of non-Gaussianity (NG), so that an observation of a large NG would indicate a deviation from this paradigm.

Until a few years ago, the main tool to constrain NG was considered to be the statistics of the cosmic microwave background (CMB) temperature field, since inhomogeneities at the CMB epoch are small and the physics can be described by a perturbative treatment. In recent years, however, thanks to observations and developments in the theory, the large-scale structure (LSS) of the universe has emerged as a complementary probe to constrain primordial NG. While it is true that the n-point functions of the density field on small scales are dominated by the recent gravitational evolution, and do not reflect anymore the statistics of primordial perturbations, it turns out that the abundance of very massive objects, which form out of high peaks of the density perturbations, is a powerful probe of primordial NG. In this context, much attention has been given recently to three possible methods of constraining the magnitude and shape of the primordial NG with the LSS: the galaxy power spectrum, the galaxy bispectrum and the mass function. It was pointed out in Refs. [94; 95] that a NG of a local type induces a scale dependence on the galaxy power spectrum, thus making it a sensitive probe of the magnitude of local NG f_{NL}^{loc}. From Ref. [44] one finds the following constraints: $-29 < f_{NL}^{loc} < +69$, already comparable with those obtained from CMB

measurements in Ref. [11]: $-10 < f_{NL}^{loc} < +74$. The future is even more promising, with precisions of $\Delta f_{NL}^{loc} \sim 10$ [96] and $\Delta f_{NL}^{loc} \sim 1$ [97] being claimed for future surveys. The galaxy bispectrum is also a promising probe of NG as it could be more sensitive to other triangle configurations [98]. The mass function – which is the focus of this work, and which we discuss in detail below – has been used for example in Ref. [96] together with the scale dependent bias to produce forecasts for future surveys, and in Ref. [99] in an attempt to explain the presence of a very massive cluster at a large redshift as an indication of a large NG. For more references and information we refer the reader to reviews summarizing recent results on these topics [5; 100].

The formation of bound dark matter halos from initially small density perturbations, as seen in numerical simulations, is a complicated and violent process. Some insight into the physics involved has been gained from the study of analytical models. The quantity of interest is the halo mass function, defined as the number density of dark matter halos with a mass between M and $M + dM$,

$$\frac{dn}{dM} = \frac{\bar{\rho}}{M^2} f(\sigma) \left| \frac{d \ln \sigma}{d \ln M} \right| , \tag{4.1}$$

where $\bar{\rho}$ is the average density of the universe, $\sigma(M)$ is the variance of the density contrast δ_R filtered on some comoving scale R corresponding to the mass M, and the function $f(\sigma)$ is to be computed. Throughout this work, we will refer to $f(\sigma)$ itself as the mass function. A very useful tool in the analysis is the spherical collapse model [101], which predicts that the value of the linearly extrapolated density contrast of a spherical halo, at the time when the halo collapses, is $\delta_c \simeq 1.686$, with a weak cosmology dependence. This value serves as a collapse threshold for determining which inhomogeneous regions will end up as collapsed objects. Using this idea, Press & Schechter [102] (PS) first computed the mass function $f(\sigma)$ in the case of Gaussian initial conditions. Their calculation however suffered from a problem of undercounting which affects the overall normalization – their approach does not count underdense regions embedded in larger overdense regions as eventually collapsed objects. To account for this discrepancy, PS introduced an ad-hoc factor of 2 by demanding that the mass function be correctly normalized, such that all the mass in the universe must be contained in collapsed objects. In the excursion set approach, Bond *et al.* [103] resolved this issue and derived a correctly normalized mass function, for Gaussian initial conditions. They argued that the filtered density contrast δ_R follows a random walk as a function of the filtering scale, and the problem of computing $f(\sigma)$ is translated into the problem of finding the rate of "first crossing" of the barrier δ_c, whose solution is well-known. We will study this formalism in detail in section 4.3 for the more general non-Gaussian case.

Turning to non-Gaussianities, the most popular non-Gaussian mass functions are those due to Matarrese, Verde and Jimenez [104] (MVJ) and LoVerde *et al.* [105] (LMSV). Both groups used the PS approach, by modifying the probability density function for the (linearized) density contrast to describe non-Gaussian initial conditions. In their prescription, the relevant object is the *ratio* $R_{\rm ng}$ of non-Gaussian to Gaussian mass functions. The full mass function is usually taken as the product of $R_{\rm ng}$ and an

appropriate Gaussian mass function as given by N-body simulations, e.g. the Sheth & Tormen mass function [106]. It is not clear however that this is the correct way to proceed. Indeed, in a series of papers [107; 108; 109], Maggiore & Riotto (MR) presented a rigorous approach to the first-passage problem in terms of path integrals, and in Ref. [109] they pointed out that a PS-like prescription in fact misses some important non-Gaussian effects stemming from 3-point correlations between *different* scales (so-called "unequal time" correlators).

On the other hand, MR treated non-Gaussian contributions to $f(\sigma)$ by simply linearizing in the 3-point function of δ_R, i.e. by linearizing in the non-Gaussian parameter f_{NL}. Since the NG are assumed to be small, in the sense that the parameter $\epsilon = \langle\delta^3\rangle/\sigma^3$ satisfies $\epsilon \ll 1$, one might expect that such a perturbative treatment is valid. However, another crucial ingredient in the problem is that the length scales of interest are large, which leads to a *second* small parameter ν^{-1} where $\nu = \delta_c/\sigma$. This is evident in the calculations of MR, who crucially use $\nu^{-2} \propto \sigma^2$ as a small parameter. Any perturbative treatment now depends not only on the smallness of ϵ and ν^{-1} individually, but also on the specific combinations of these parameters which appear in the calculation. It is known (and we will explicitly see below) that a natural combination that appears is $\epsilon\nu^3$, which can become of order unity on scales of interest. The mass functions given by LMSV and MR therefore break down as valid series expansions when this occurs. Interestingly, MVJ's PS-like treatment on the other hand involved a saddle point approximation, allowing them to *non-perturbatively* account for the $\epsilon\nu^3$ term (which appears in an exponential in their approach). For a discussion, see Ref. [110].

It appears to us therefore, that there is considerable room for improvement in the theoretical calculation of the mass function. The goal of this chapter is twofold. Firstly, we present a rigorous calculation of the mass function in the following way : (a) we use the techniques developed by MR in Refs. [107; 108; 109], which allow us to track the complex multi-scale correlations involved in the calculation, and (b) we demonstrate that MR's approach can be combined with saddle point techniques (used by MVJ), to non-perturbatively handle terms which can become of order unity. This leads to an expression for the mass function which is valid on much larger scales than those presented by MR and LMSV. Secondly, by keeping track of the terms ignored, we calculate theoretical error bars on the expressions for $f(\sigma)$ resulting not only from our own calculations, but also for those of the other authors [104; 105; 109]. Since the terms ignored depend on ν in general, these error bars are clearly scale dependent. This allows us to estimate the validity of each of the expressions for the mass function at different scales, but importantly it also allows us to analytically compare between different expressions. In this chapter we will not explicitly account for effects of the ellipsoidal collapse model [111; 112], since these are expected to be negligible on the very large scales which are of interest to us. For a recent treatment of ellipsoidal collapse effects in the presence of non-Gaussianities on scales where $\epsilon\nu^3 \ll 1$, see Lam & Sheth [113].

This chapter is organized as follows. In section 4.2 we fix some notation and briefly introduce the two most popular shapes of primordial NG, i.e. the local and equilateral

ones. In section 4.3 we present our calculation of the mass function. In section 4.4 we discuss certain subtleties regarding the truncation of the perturbative series, and also compare with the other expressions for $f(\sigma)$ mentioned above. In section 4.5 we discuss the effects induced by some additional complications introduced in the problem due to the specific choice of the filter function [107], which we take to be a top-hat in real space, and due to the inclusion of stochasticity in the value of the collapse threshold δ_c (which is also expected to partially account for effects of ellipsoidal collapse) [108] . In section 4.6 we compare our final result Eqn. (4.68) with those of other authors, including theoretical errors for each, and conclude with a brief discussion of the results and directions for future work. Some technical asides have been relegated to the Appendices.

4.2 Models of non-Gaussianity

In this section we will repeat some of the concepts introduced in previous chapter, and though it is somewhat repetitive, it is useful to do so in order to introduce some notation that we will use in the following sections.

We need to relate the linearly evolved density field to the primordial curvature perturbation, which carries the information of the non-linearities produced during and after inflation. We start from the Bardeen potential Φ on subhorizon scales, given by

$$\Phi(\vec{k},z) = -\frac{3}{5}T(k)\frac{D(z)}{a}\zeta(k)\,, \tag{4.2}$$

where $\zeta(\vec{k})$ is the (comoving) curvature perturbation, which stays constant on super-horizon scales; $T(k)$ is the transfer function of perturbations, normalized to unity as $k \to 0$, which describes the suppression of power for modes that entered the horizon before the matter-radiation equality; and $D(z)$ is the linear growth factor of density fluctuations, normalized such that $D(z) = (1+z)^{-1}$ in the matter dominated era. Then, the density contrast field is related to the potential by the Poisson equation, which in Fourier space reads

$$\begin{aligned}\delta(\vec{k},z) &= -\frac{2ak^2}{3\Omega_m H_0^2}\Phi(\vec{k},z) = \frac{2k^2}{5\Omega_m H_0^2}T(k)D(z)\zeta(k)\\ &\equiv \mathcal{M}(k,z)\zeta(k)\,,\end{aligned} \tag{4.3}$$

where we substituted Eqn. (4.2). Here, Ω_m is the present time fractional density of matter (cold dark matter and baryons), and $H_0 = 100h\,\mathrm{km\,s^{-1}Mpc^{-1}}$ is the present time Hubble constant. The redshift dependence is trivially accounted for by the linear growth factor $D(z)$ and in the following, for notational simplicity, we will often suppress it. All our calculations will use a reference ΛCDM cosmology compatible with WMAP7 data [11], using parameters $h = 0.702$, $\Omega_m = 0.272$, present baryon density $\Omega_b = 0.0455$, scalar spectral index $n_s = 0.961$ and $\sigma_8 = 0.809$, where σ_8^2 is the variance of the density field smoothed on a length scale of $8h^{-1}$Mpc. For simplicity, for the transfer function

$T(k)$ we use the BBKS form, proposed in Bardeen *et al.* [111]:

$$T_{\rm BBKS}(x) \equiv \frac{1}{2.34x}\ln\left(1+2.34x\right) \times \left(1+3.89x+(16.1x)^2+(5.46x)^3+(6.71x)^4\right)^{-1/4}\,, \quad (4.4)$$

where $x \equiv k(h{\rm Mpc}^{-1})/\Gamma$ with a shape parameter, which accounts for baryonic effects as described in Ref. [114], defined as $\Gamma = \Omega_m h \exp\left[-\Omega_b(1+\sqrt{2h}/\Omega_m)\right]$. For more accurate results, one could use a numerical transfer function, as obtained by codes like CMBFAST [22] or CAMB [21]; the results are not expected to be qualitatively different.

In order to study halos, which form where an extended region of space has an average overdensity which is above threshold, it is useful to introduce a filter function $W_R(|\vec{x}|)$, and consider the smoothed density field (around one point, which we take as the origin),

$$\delta_R = \int \frac{d^3k}{(2\pi)^3}\widetilde{W}(kR)\delta(\vec{k})\,, \quad (4.5)$$

where $\widetilde{W}(kR)$ is the Fourier transform of the filter function. For all numerical calculations we will use the spherical top-hat filter in real space, whose Fourier transform $\widetilde{W}(kR)$ is given by

$$\widetilde{W}(y) = \frac{3}{y^3}\left(\sin y - y\cos y\right)\,. \quad (4.6)$$

This choice allows us to have a well-defined relation between length scales and masses, namely $M = (4\pi/3)\Omega_m\rho_c R^3$ with $\rho_c = 3H_0^2/(8\pi G) = 2.75\cdot 10^{11}h^{-1}M_{\rm sol}(h^{-1}{\rm Mpc})^{-3}$. However it introduces some complexities in the analysis, which we will comment on later. By using Eqns. (4.5) and (4.3) we have, for the 3-point function,

$$\langle\,\delta_{R_1}\delta_{R_2}\delta_{R_3}\,\rangle_c = \int \frac{d^3k_1}{(2\pi)^3}\frac{d^3k_2}{(2\pi)^3}\frac{d^3k_3}{(2\pi)^3}\widetilde{W}(k_1R_1)\widetilde{W}(k_2R_2)\widetilde{W}(k_3R_3) \times \mathcal{M}(k_1)\mathcal{M}(k_2)\mathcal{M}(k_3)\langle\,\zeta(\vec{k}_1)\zeta(\vec{k}_2)\zeta(\vec{k}_3)\,\rangle_c\,, \quad (4.7)$$

where the subscript c denotes the connected part, and analogous formulae are valid for the higher order correlations.

4.2.1 Shapes of non-Gaussianity

The function $\langle\,\zeta(\vec{k}_1)\zeta(\vec{k}_2)\zeta(\vec{k}_3)\,\rangle_c$ encodes information about the physics of the inflationary epoch. By translational invariance, it is proportional to a momentum-conserving delta function:

$$\langle\,\zeta(\vec{k}_1)\zeta(\vec{k}_2)\zeta(\vec{k}_3)\,\rangle_c = (2\pi)^3\delta_D(\vec{k}_1+\vec{k}_2+\vec{k}_3)B_\zeta(k_1,k_2,k_3)\,, \quad (4.8)$$

where the (reduced) bispectrum $B_\zeta(k_1,k_2,k_3)$ depends only on the magnitude of the k's by rotational invariance. According to the particular model of inflation, the bispectrum

will be peaked about a particular shape of the triangle. The two most common cases are the squeezed (or local) NG, peaked on squeezed triangles $k_1 \ll k_2 \simeq k_3$, and the equilateral NG, peaked on equilateral triangles $k_1 \simeq k_2 \simeq k_3$. Indeed, one can define a scalar product of bispectra, which describes how sensitive one is to a NG of a given type if the analysis is performed using some template form for the bispectrum. As expected, the local and equilateral shapes are approximately orthogonal with respect to this scalar product [34]. We will now describe these two models in more detail.

The local model:

The local bispectrum is produced when the NG is generated outside the horizon, for instance in the curvaton model [32; 115] or in the inhomogeneous reheating scenario [116]. In these models, the curvature perturbation can be written in the following form,

$$\zeta(\vec{x}) = \zeta_g(\vec{x}) + \frac{3}{5} f_{NL}^{\rm loc} \left(\zeta_g^2(\vec{x}) - \langle \zeta_g^2 \rangle\right) + \frac{9}{25} g_{\rm NL} \zeta_g^3(\vec{x}) \,, \tag{4.9}$$

where ζ_g is the linear, Gaussian field. We have included also a cubic term, which will generate the trispectrum at leading order. The bispectrum is given by

$$B_\zeta(k_1, k_2, k_3) = \frac{6}{5} f_{NL}^{\rm loc} \left[P_\zeta(k_1) P_\zeta(k_2) + {\rm cycl.}\right] \,, \tag{4.10}$$

where "cycl." denotes the 2 cyclic permutations of the wavenumbers, and $P_\zeta(k)$ is the power spectrum given by $P_\zeta(k) = A k^{n_s - 4}$. The trispectrum is given by

$$\begin{aligned} \langle \zeta(\vec{k}_1)\zeta(\vec{k}_2)\zeta(\vec{k}_3)\zeta(\vec{k}_4) \rangle_c = (2\pi)^3 \delta_D(\vec{k}_1 + \vec{k}_2 + \vec{k}_3 + \vec{k}_4) \\ \times \Bigg[\frac{36}{25} f_{NL}^2 \sum_{\substack{b<c \\ a \neq b,c}} P_\zeta(|\vec{k}_a + \vec{k}_b|) P_\zeta(k_b) P_\zeta(k_c) \\ + \frac{54}{25} g_{\rm NL} \sum_{a<b<c} P_\zeta(k_a) P_\zeta(k_b) P_\zeta(k_c) \Bigg] \,. \end{aligned} \tag{4.11}$$

The equilateral model:

Models with derivative interactions of the inflaton field [30; 117; 118] give a bispectrum which is peaked around equilateral configurations, whose specific functional form is model dependent. Moreover, the form of the bispectrum is usually not convenient to use in numerical analyses. This is why, when dealing with equilateral NG, it is convenient to use the following parametrization, given in Ref. [119],

$$\begin{aligned} B_\zeta(k_1, k_2, k_3) = \frac{18}{5} f_{NL}^{\rm equil} A^2 \Bigg[\frac{1}{2 k_1^{4-n_s} k_2^{4-n_s}} + \frac{1}{3(k_1 k_2 k_3)^{2(4-n_s)/3}} \\ - \frac{1}{(k_1 k_2^2 k_3^3)^{(4-n_s)/3}} + 5 \text{ perms.} \Bigg] \,. \end{aligned} \tag{4.12}$$

This is peaked on equilateral configurations, and its scalar product with the bispectra produced by the realistic models cited above is very close to one. Therefore, being a sum of factorizable functions, it is the standard template used in data analyses.

4.3 Random walks and the halo mass function

We now turn to the main calculation of the chapter. The non-Gaussian halo mass function can be obtained by calculating the barrier first crossing rate $\mathcal{F}$ of a random walk with non-Gaussian noise, in the presence of an absorbing barrier. This can be done perturbatively, starting from a path integral approach as prescribed by MR [107; 109] and the mass function can be shown to be $f(\sigma) = 2\sigma^2\mathcal{F}(\sigma)$. As discussed by MR, the calculation of f involves certain assumptions regarding the type of filter used and also the location of the barrier. In particular, the formalism is simplest for a sharp filter in k-space, and using the spherical top-hat of Eqn. (4.6) introduces complications in the form of non-Markovian effects. Further, in order to make the spherical collapse ansatz more realistic and obtain better agreement with N-body simulations, MR show that it is useful to treat the location of the barrier δ_c as a stochastic variable itself, and allow it to diffuse. For the time being, we will ignore these complications, and will return to their effects in section 4.5.

To make the chapter self-contained, we begin with a brief review of the path integral approach to the calculation of the mass function. The reader is referred to Ref. [107] for a more pedagogical introduction. In the path integral approach, one treats the variance $\sigma_R^2 \equiv \langle\, \hat{\delta}_R^2 \,\rangle$ as a "time" parameter, $s \equiv \sigma_R^2$, and considers the random walk followed by the smoothed density field $\hat{\delta}_R$ as this "time" is increased in discrete steps starting from small values (equivalently, as R is decreased from very large values). Here $\hat{\delta}_R(\vec{x})$ is a stochastic quantity in real space due to the stochasticity inherent in the initial conditions. We use the notation $\hat{\delta}_R$ to distinguish the stochastic variable from the values it takes, which will be noted by δ_i below. We probe this stochasticity by changing the smoothing scale at a fixed location $\vec{x} = 0$, thus making the variable perform a random walk, which obeys a Langevin equation

$$\frac{\partial\hat{\delta}}{\partial s} = \hat{\eta}\,, \tag{4.13}$$

with a stochastic noise $\hat{\eta}$ whose statistical properties depend on the choice of filter used. In particular, for a top hat filter in k-space, the noise is white, i.e. its 2-point function is a Dirac delta [103],

$$\langle\, \hat{\eta}(s_1)\hat{\eta}(s_2) \,\rangle = \delta_{\rm D}(s_1 - s_2) \tag{4.14}$$

The random walk can be described as a trajectory $\{\delta_0, \delta_1, \ldots, \delta_n\}$ which starts with $\hat{\delta}$ taking the value $\delta_0 = 0$ at $s = 0$ (or $R \to \infty$ which is the homogeneous limit), then taking values δ_i at times s_i, finally arriving at δ_n at time t_n, with a discrete timestep $\Delta s = s_{k+1} - s_k = s_n/n$. The probability $\mathcal{P}(s)$ that the trajectory crosses the barrier at δ_c at a time larger than some s (i.e. at scales smaller than the corresponding R or M), is the same as the probability that the trajectory *did not* cross the barrier at any time smaller than s, so that

$$\mathcal{P}(s) = \int_{-\infty}^{\delta_c} {\rm d}\delta_1 \ldots {\rm d}\delta_n W(\{\delta_j\}; s)\,, \tag{4.15}$$

where the probability density over the space of trajectories, $W(\{\delta_j\}; s)$ is defined as

$$W(\{\delta_j\}; s) \equiv \langle \delta_{\rm D}(\hat{\delta}(s_1) - \delta_1) \dots \delta_{\rm D}(\hat{\delta}(s_n) - \delta_n) \rangle \,, \tag{4.16}$$

where $\delta_{\rm D}$ is the Dirac delta distribution. The first crossing *rate* is given by the negative time derivative of $\mathcal{P}$, $\mathcal{F} = -\partial_s \mathcal{P}$, and the mass function is then $f = 2s\mathcal{F}(s)$. In Eqn. (4.16) one can write the Dirac deltas using the integral representation $\delta_{\rm D}(x) = \int_{-\infty}^{\infty} d\lambda e^{-i\lambda x}/2\pi$, to obtain

$$W(\{\delta_j\}; s) = \int_{-\infty}^{\infty} \frac{d\lambda_1}{2\pi} \dots \frac{d\lambda_n}{2\pi} \langle e^{-i\sum_j \lambda_j \hat{\delta}(s_j)} \rangle e^{i\sum_j \lambda_j \delta_j} \,. \tag{4.17}$$

The object $\langle e^{-i\sum_j \lambda_j \hat{\delta}_j} \rangle$ is the exponential of the generating functional of the connected Green's functions, and can be shown to reduce to [120]

$$\langle e^{-i\sum_j \lambda_j \hat{\delta}_j} \rangle = \exp\left[\sum_{p=2}^{\infty} \frac{(-i)^p}{p!} \sum_{j_1,..,j_p=1}^{n} \lambda_{j_1} \dots \lambda_{j_p} \langle \hat{\delta}_{j_1} \dots \hat{\delta}_{j_p} \rangle_c \right] , \tag{4.18}$$

where $\langle \hat{\delta}_{j_1} \dots \hat{\delta}_{j_p} \rangle_c$ is the connected p-point function of $\hat{\delta}$, with the short-hand notation $\hat{\delta}_j = \hat{\delta}(s_j)$.

4.3.1 Halo mass function: Gaussian case, sharp-k filter

In the Gaussian case, all connected n-point correlators vanish except for $n = 2$, and in the Markovian (sharp-k filter) case which we are considering, the 2-point function becomes $\langle \hat{\delta}_j \hat{\delta}_k \rangle = \min(s_j, s_k)$, where $\min(s_j, s_k)$ is the minimum of s_j and s_k. The resulting n-dimensional Gaussian integral can be handled in a straightforward way to obtain

$$W^{\rm gm} = \prod_{k=0}^{n-1} \Psi_{\Delta s}(\delta_{k+1} - \delta_k) \; ; \quad \Psi_{\Delta s}(x) = (2\pi\Delta s)^{-1/2} e^{-x^2/(2\Delta s)} \,, \tag{4.19}$$

where we follow MR's notation and use the superscript "gm" to denote "Gaussian Markovian". As MR have shown [107], the resulting expression for $\mathcal{P}_{\rm gauss}(t)$ in the continuum limit $\Delta s \to 0$ is simply

$$\mathcal{P}_{\rm gauss} = \int_{-\infty}^{\delta_c} d\delta_1 \dots d\delta_n W^{\rm gm} = {\rm erf}\left(\frac{\nu}{\sqrt{2}} \right) \,, \tag{4.20}$$

where we use the notation $\nu \equiv \delta_c/\sigma$. (This in principle also includes the redshift dependence of the collapse threshold δ_c, see below.) This expression for the continuum limit probability $\mathcal{P}_{\rm gauss}$ is of course a well-known result going back to Chandrasekhar [121]. This leads to the standard excursion set result for the Gaussian mass function $f_{\rm PS} = -2s\partial_s|_{\delta_c} \mathcal{P}_{\rm gauss}$,

$$f_{\rm PS}(\nu) = \sqrt{\frac{2}{\pi}} \nu \, e^{-\nu^2/2} \,, \tag{4.21}$$

where we use the subscript PS (for Press-Schechter), to conform with the conventional notation for this object.

4.3.2 Halo mass function: non-Gaussian case, sharp-k filter

In the non-Gaussian case (but still retaining the sharp-k filter), the probability density $W(\{\delta_j\}; s)$ also gets contributions from connected n-point correlators with $n \geq 3$, since these in general do not vanish. These can be handled by using the relation $\lambda_k e^{i\sum_j \lambda_j \delta_j} = -i\partial_k e^{i\sum_j \lambda_j \delta_j}$, with $\partial_j \equiv \partial/\partial\delta_j$. A straightforward calculation then shows the mass function to be

$$f = -2s\,\frac{\partial}{\partial s}\bigg|_{\delta_c} \int_{-\infty}^{\delta_c} \mathrm{d}\delta_1 \ldots \mathrm{d}\delta_n \exp\Big[-\frac{1}{3!}\sum_{j,k,l=1}^{n} \langle\, \hat{\delta}_j \hat{\delta}_k \hat{\delta}_l \,\rangle_c \partial_j\partial_k\partial_l + \frac{1}{4!}\sum_{j,k,l,m=1}^{n} \langle\, \hat{\delta}_j \hat{\delta}_k \hat{\delta}_l \hat{\delta}_m \,\rangle_c \partial_j\partial_k\partial_l\partial_m + \ldots \Big] W^{\mathrm{gm}}\,, \tag{4.22}$$

where it is understood that one takes the continuum limit $\Delta s \to 0$ before computing the overall derivative with respect to s. We will find it useful to change variables from $(\delta_c,\, s)$ to $(\nu,\, s)$, in which case the partial derivative becomes

$$-2s(\partial/\partial s)|_{\delta_c} = \nu(\partial/\partial\nu)|_s - 2s(\partial/\partial s)|_\nu \equiv \nu\partial_\nu - 2s\partial_s\,. \tag{4.23}$$

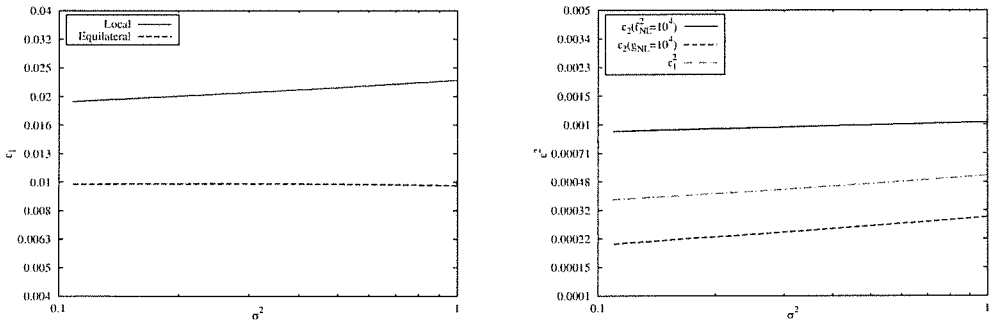
It is also useful at this stage to take a small detour and introduce some notation which we will use throughout the rest of the chapter. We define the scale dependent "equal time" functions

$$\varepsilon_{n-2} \equiv \frac{\langle\, \hat{\delta}_R^n \,\rangle_c}{\sigma_R^n} \quad ; \quad n \geq 3\,, \tag{4.24}$$

which as we will see, remain approximately constant over the scales of interest. We assume the ordering $\varepsilon_{n-2} \sim \mathcal{O}(\epsilon^{n-2})$ with $\epsilon \ll 1$, which can be motivated from their origin in inflationary physics, where one finds $\varepsilon_1 \sim f_{NL}A^{1/2}$, $\varepsilon_2 \sim g_{\mathrm{NL}}A$, etc[1]. Typically we expect $\epsilon \lesssim 10^{-2}$ for $f_{NL} \lesssim 100$. Fig. 4.1 shows the behaviour of ε_1 and ε_2 in the local and equilateral models, as a function of $t = \sigma_R^2$. We see e.g. that ε_2 in the local model is comparable to ε_1^2. In the literature one usually encounters the reduced cumulants $\mathcal{S}_n$, which are related to the ε_{n-2} by $\varepsilon_1 = \sigma\mathcal{S}_3$, $\varepsilon_2 = \sigma^2\mathcal{S}_4$ and so on. The motivation for using the $\mathcal{S}_n$ comes from the study of NG induced by nonlinear gravitational effects. However, as we see from Fig. 4.1, when studying *primordial* NG it is more meaningful to consider the ε_n which are approximately scale-independent and perturbatively ordered.

We will soon see that a natural expansion parameter that arises in the calculation has the form $\sim \epsilon\nu$, and we therefore require that the mass scales under scrutiny are not large enough to spoil the relation $\epsilon\nu \ll 1$. It turns out that observationally interesting mass scales can nevertheless be large enough to satisfy $\epsilon\nu^3 \sim \mathcal{O}(1)$. Fig. 4.2 shows the behaviour of $\varepsilon_1\nu^3$ and $\varepsilon_1\nu$ at different redshifts, as a function of mass, in our reference ΛCDM model for local type NG, with $f_{NL}^{\mathrm{loc}} = 100$. The behaviour for the equilateral NG is qualitatively similar. The redshift dependence of these quantities comes from

[1] Notationally we distinguish the order parameter ϵ from the specific NG functions ε_1 and ε_2.



(a) (b)

Figure 4.1: Scale dependence of the ε_n. Panel (a) : Behaviour of ε_1 vs. σ^2 in the local and equilateral models, for $f_{NL} = 100$ in each case. Panel (b) : Behaviour of ε_2 for the local model with $f_{NL} = 100$ and $g_{\rm NL} = 10^4$. The terms proportional to f_{NL}^2 and $g_{\rm NL}$ are shown separately. Also shown is ε_1^2 for the same model. The axes are logscale.

the definition of ν,

$$\nu(M,z) \equiv \sqrt{a}\frac{\delta_{c0}}{\sigma(M)}\frac{D(0)}{D(z)} \equiv \frac{\delta_c(z)}{\sigma(M)}\,, \tag{4.25}$$

where we denote the usual spherical collapse threshold as $\delta_{c0} = (3/5)(3\pi/2)^{2/3} \simeq 1.686$, reserving δ_c for the full, redshift dependent quantity, and a is a parameter accounting for deviations from the simplest collapse model. In the standard spherical collapse picture we have $a = 1$. A value of a different from unity (specifically $\sqrt{a} \simeq 0.89$) can be motivated by allowing the collapse threshold to vary stochastically [108], as we will discuss in section 4.5. We will soon see that the object $\epsilon\nu^3$ appears naturally in the calculation, and to be definite we will assume $\epsilon\nu^3 \sim \mathcal{O}(1)$ for now. In section 4.4 we will discuss the effects of relaxing this condition and probing smaller length scales.

We now turn to the "unequal time" correlators appearing in Eqn. (4.22). Since we are concerned with large scales, we are in the small t limit, and following MR we expand the n-point correlators around the "final time" t. We can define the Taylor coefficients

$$\mathcal{G}_3^{(p,q,r)}(s) \equiv \left[\frac{d^p}{ds_j^p}\frac{d^q}{ds_k^q}\frac{d^r}{ds_l^r}\langle\,\hat{\delta}(s_j)\hat{\delta}(s_k)\hat{\delta}(s_l)\,\rangle_c\right]_{s_j=s_k=s_l=s}\,, \tag{4.26}$$

and then expand

$$\langle\,\hat{\delta}_j\hat{\delta}_k\hat{\delta}_l\,\rangle_c = \sum_{p,q,r=0}^{\infty}\frac{(-1)^{p+q+r}}{p!q!r!}\mathcal{G}_3^{(p,q,r)}(s)(s-s_j)^p(s-s_k)^q(s-s_l)^r\,. \tag{4.27}$$

For the 4-point function we will have an analogous expression involving coefficients $\mathcal{G}_4^{(p,q,r,l)}$.

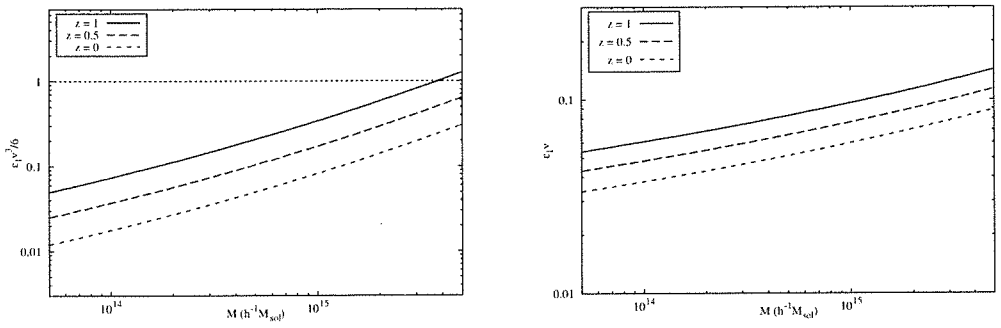


Figure 4.2: Panel (a) : Behaviour of $\varepsilon_1\nu^3/6$ vs. mass in the local non-Gaussian model, for $f_{NL} = 100$. The three curves correspond to different redshifts. The horizontal line corresponds to $\varepsilon_1\nu^3/6 = 1$. Panel (b) : Behaviour of $\varepsilon_1\nu$ with the same setup as in panel (a).

Since calculations involving a general set of coefficients $\mathcal{G}_3$, $\mathcal{G}_4$, etc. are algebraically rather involved, we find it useful to first consider a toy example in which these coefficients take simple forms. In this model we assume that the ε_n are exactly constant, and moreover that the n-point correlators take the form[1]

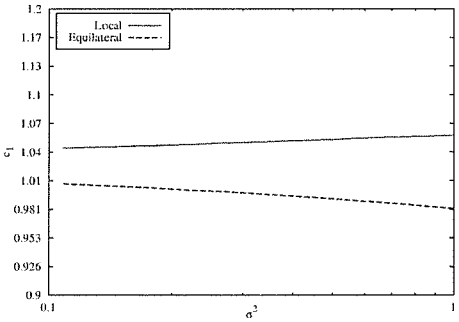
$$\langle \hat{\delta}_j\hat{\delta}_k\hat{\delta}_l \rangle_c = \varepsilon_1(s_j s_k s_l)^{1/2} \;\; ; \;\; \langle \hat{\delta}_j\hat{\delta}_k\hat{\delta}_l\hat{\delta}_m \rangle_c = \varepsilon_2(s_j s_k s_l s_m)^{1/2} \,. \tag{4.28}$$

For clarity, we will display details of the calculation only for this model. In the more realistic case of slowly-varying ε_n, we choose to parametrize the coefficients $\mathcal{G}_3$ and $\mathcal{G}_4$ in a convenient way as follows:

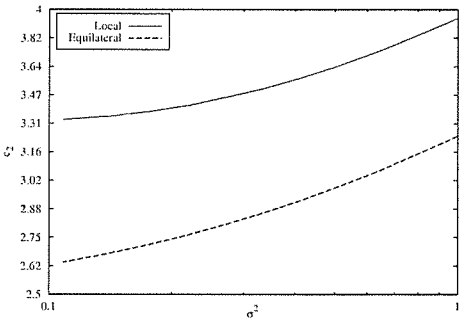
$$\begin{aligned} \mathcal{G}_3^{(1,0,0)} &= \frac{1}{2}\varepsilon_1(s)c_1(s)s^{1/2} \;\; ; \;\; \mathcal{G}_3^{(2,0,0)} = -\frac{1}{4}\varepsilon_1(s)c_2(s)s^{-1/2} \,, \\ \mathcal{G}_3^{(1,1,0)} &= \frac{1}{4}\varepsilon_1(s)c_3(s)s^{-1/2} \;\; ; \;\; \mathcal{G}_4^{(1,0,0,0)} = \frac{1}{2}\varepsilon_2(s)c_4(s)s \,, \end{aligned} \tag{4.29}$$

where the coefficients $c_n(t)$ are smoothly varying functions and depend on the NG model. They are defined in such a way that they all reduce to unity in the toy model defined by Eqn. (4.28). Fig. 4.3 shows the behaviour of c_1, c_2 and c_3 with σ^2, for the local and equilateral models. The ε_n and c_n are independent of redshift by construction, since the linear growth rate $D(z)$ always drops out in their definitions. Also the c_n do not depend on the values of f_{NL} and $g_{\rm NL}$. The calculation of the mass function for this general case proceeds completely analogously to that for the toy model, apart from a few subtleties which we will discuss later, and our final result will be an expression for f in the general case.

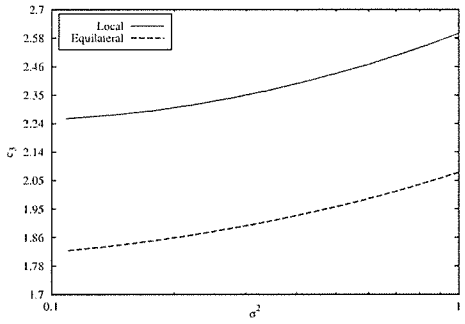
[1]Throughout the chapter we will consider at most 4-point correlators. This truncation is justified given our assumptions, as we will see later.



(a)



(b)



(c)

Figure 4.3: The derivative coefficients c_1 (panel (a)), c_2 (panel (b)) and c_3 (panel (c)), as a function of σ^2, for local and equilateral NG models. These quantities are independent of redshift and the NG amplitudes f_{NL} and $g_{\rm NL}$. The axes are logscale.

Using the first few terms of the unequal time expansions, in our toy model one can write

$$\sum_{j,k,l=1}^{n} \langle \hat{\delta}_j \hat{\delta}_k \hat{\delta}_l \rangle_c \, \partial_j \partial_k \partial_l = \varepsilon_1 s^{3/2} \Bigg(\sum_{j,k,l=1}^{n} \partial_j \partial_k \partial_l - \frac{3}{2} \sum_{j=1}^{n} (1 - \frac{s_j}{s}) \partial_j \sum_{k,l=1}^{n} \partial_k \partial_l$$
$$- \frac{3}{8} \sum_{j=1}^{n} (1 - \frac{s_j}{s})^2 \partial_j \sum_{k,l=1}^{n} \partial_k \partial_l$$
$$+ \frac{3}{4} \sum_{j,k=1}^{n} (1 - \frac{s_j}{s})(1 - \frac{s_k}{s}) \partial_j \partial_k \sum_{l=1}^{n} \partial_l + \dots \Bigg), \quad (4.30)$$

$$\sum_{j,k,l,m=1}^{n} \langle \hat{\delta}_j \hat{\delta}_k \hat{\delta}_l \hat{\delta}_m \rangle_c \, \partial_j \partial_k \partial_l \partial_m = \varepsilon_2 s^2 \Big(\sum_{j,k,l,m=1}^{n} \partial_j \partial_k \partial_l \partial_m - 2 \sum_{j=1}^{n} (1 - \frac{s_j}{s}) \partial_j \sum_{k,l,m=1}^{n} \partial_k \partial_l \partial_m + \dots \Big) \, . \quad (4.31)$$

These derivative operators are exponentiated in the path integral, and act on $W^{\rm gm}$. One simplification that occurs in our toy model, is that the path integral in Eqn. (4.22) becomes a function only of ν (although this is not obvious at this stage), and hence eventually only the $\nu\partial_\nu$ part of the overall derivative contributes. However, the structure of the exponentiated derivatives is still rather formidable. Moreover, the truncation of the series at this stage is based more on the intuition that higher order terms should somehow be smaller, rather than on a strict identification of the small parameters. In fact, we will see in detail in section 4.4 that the issue of truncation involves several subtleties.

To make progress, it helps to analyze the effect on $W^{\rm gm}$ of each of the terms in the above series, *before* exponentiation. The leading term in Eqn. (4.22) involves the multiple integral of $W^{\rm gm}$, which is just the quantity $\mathcal{P}_{\rm gauss}$ encountered in Eqn. (4.20). The operator $\nu\partial_\nu$ acts on the error function to give the Gaussian rate of Eqn. (4.21). Next, notice that the action of the operator $\sum_{j=1}^{n} \partial_j$ on *any* function $g(\delta_1, \dots \delta_n)$ under the multiple integral, is simply

$$\int_{-\infty}^{\delta_c} d\delta_1 \dots d\delta_n \sum_{j=1}^{n} \partial_j g = \frac{\partial}{\partial \delta_c} \int_{-\infty}^{\delta_c} d\delta_1 \dots d\delta_n g \, , \quad (4.32)$$

Using this, and the fact that $s^{1/2}(\partial/\partial\delta_c)|_s = \partial_\nu|_s$, we see that the leading term in Eqn. (4.30) (i.e. the term with no powers of $(1 - s_j/s)$), leads to a term involving

$$\varepsilon_1 \nu \partial_\nu (\partial_\nu)^3 {\rm erf}\left(\nu/\sqrt{2}\right) \sim f_{\rm PS} \varepsilon_1 \nu^3 (1 + \mathcal{O}(\nu^{-2})) \, ,$$

The problem with this term is that the quantity $\varepsilon_1\nu^3$ can be of order unity, and hence cannot be treated perturbatively. To be consistent, we should keep all terms involving powers of $\varepsilon_1\nu^3$. Luckily, this can be done in a straightforward way due to the result in Eqn. (4.32). We see that the entire exponential operator $\exp[-(\varepsilon_1 t^{3/2}/3!) \sum_{j,k,l=1}^{n} \partial_j\partial_k\partial_l]$ in Eqn. (4.22) can be pulled across the multiple integral and converted to $\exp[-(\varepsilon_1/3!)\partial_\nu^3]$ acting on the remaining integral. Similarly, the operator $\exp[(\varepsilon_2 s^2/4!) \sum_{j,k,l,m=1}^{n} \partial_j\partial_k\partial_l\partial_m]$ can be pulled out and converted to $\exp[(\varepsilon_2/4!)\partial_\nu^4]$, and the same applies for all such equal time operators. We will see later that the action of these operators can be easily accounted for, using a saddle-point approximation. To summarize, the function f at

this stage is given by

$$f = \nu\, e^{-(\varepsilon_1/3!)\partial_\nu^3+(\varepsilon_2/4!)\partial_\nu^4+\ldots}\partial_\nu \int_{-\infty}^{\delta_c} d\delta_1 \ldots d\delta_n \exp\Bigg[\frac{1}{3!}\varepsilon_1 s^{3/2}\Big(\frac{3}{2}\sum_{j=1}^{n}(1-\frac{s_j}{s})\partial_j \sum_{k,l=1}^{n}\partial_k\partial_l + \frac{3}{8}\sum_{j=1}^{n}(1-\frac{s_j}{s})^2\partial_j\sum_{k,l=1}^{n}\partial_k\partial_l - \frac{3}{4}\sum_{j,k=1}^{n}(1-\frac{s_j}{s})(1-\frac{s_k}{s})\partial_j\partial_k\sum_{l=1}^{n}\partial_l + \ldots\Big) - \frac{1}{4!}\varepsilon_2 s^2\Big(2\sum_{j=1}^{n}(1-\frac{s_j}{s})\partial_j \sum_{k,l,m=1}^{n}\partial_k\partial_l\partial_m + \ldots\Big)\Bigg] W^{\rm gm}\,. \tag{4.33}$$

Now consider the action of the individual terms in the remaining exponential under the integrals, but without exponentiation. From MR [109], we have the following results[1],

$$\sum_{j=1}^{n}(1-\frac{s_j}{s})\sum_{k,l=1}^{n}\int_{-\infty}^{\delta_c} d\delta_1\ldots d\delta_n \partial_j\partial_k\partial_l W^{\rm gm} = \left(\frac{2}{\pi}\right)^{1/2}\frac{1}{s^{3/2}}e^{-\nu^2/2}\,, \tag{4.34a}$$

$$\sum_{j=1}^{n}(1-\frac{s_j}{s})^2\sum_{k,l=1}^{n}\int_{-\infty}^{\delta_c} d\delta_1\ldots d\delta_n \partial_j\partial_k\partial_l W^{\rm gm} = \left(\frac{2}{\pi}\right)^{1/2}\frac{3}{s^{3/2}}h(\nu)\,, \tag{4.34b}$$

$$\sum_{j,k=1}^{n}(1-\frac{s_j}{s})(1-\frac{s_k}{s})\sum_{l=1}^{n}\int_{-\infty}^{\delta_c} d\delta_1\ldots d\delta_n \partial_j\partial_k\partial_l W^{\rm gm} = \left(\frac{2}{\pi}\right)^{1/2}\frac{4}{s^{3/2}}h(\nu)\,, \tag{4.34c}$$

where we have defined

$$h(\nu) \equiv e^{-\nu^2/2} - \left(\frac{\pi}{2}\right)^{1/2}\nu\,\mathrm{erfc}\left(\frac{\nu}{\sqrt{2}}\right) = \frac{\nu}{2^{3/2}}\Gamma\left(-\frac{1}{2},\frac{\nu^2}{2}\right)\,, \tag{4.35}$$

where $\Gamma(-1/2,\nu^2/2)$ is an incomplete gamma function. Let us focus on the term in Eqn. (4.34a). If we linearize in ε_1 in Eqn. (4.33), then this term appears with $\varepsilon_1 s^{3/2}\partial_\nu$ acting on it, leading to $\sim f_{\rm PS}\varepsilon_1\nu \ll f_{\rm PS}$. This term can therefore be treated perturbatively. Similarly, one can check that the terms given by Eqns. (4.34b) and (4.34c) also lead to perturbatively small quantities, which are in fact further suppressed compared to $\varepsilon_1\nu$ by powers of ν^{-2}. Specifically, one obtains terms involving $\varepsilon_1\mathrm{erfc}\left(\nu/\sqrt{2}\right)$ which, for large ν, reduces to $\sim f_{\rm PS}\cdot\varepsilon_1\nu\cdot\nu^{-2}(1+\mathcal{O}(\nu^{-2}))$.

A few comments are in order at this stage. First, this ordering in powers of ν^{-2} is a generic feature of integrals involving an increasing number of powers of $(1-s_j/s)$ being summed. This can be understood in a simple way from the asymptotic properties of the incomplete gamma function, as we show in Appendix 4.A. We are therefore justified in truncating the Taylor expansion of the unequal time correlators, even though superficially (on dimensional grounds) each term in the series appears to be equally important. Secondly, we have not yet accounted for the effect of the exponential derivatives. In fact we will see in the next section that when $\epsilon\nu^3 \sim \mathcal{O}(1)$, it is these terms that impose

[1]The terms in Eqns. (4.34a), (4.34b) and (4.34c) are, upto prefactors, the integrals of what MR denote as $\Pi^{(3,\mathrm{NL})}$, $\Pi^{(3,\mathrm{NNLa})}$ and $\Pi^{(3,\mathrm{NNLb})}$ respectively in Ref. [109].

stricter conditions on the series truncations. For now, however, we have no guidance other than the fact that if we account for one term of order $\sim \epsilon^n \nu^n$, then we should account for *all* terms at this order. Given this, note that for $\epsilon\nu^3 \sim \mathcal{O}(1)$ we have $\nu^{-2} \sim \epsilon\nu$, and hence the terms arising from Eqns. (4.34b) and (4.34c) are of order $\sim \epsilon^2\nu^2$. To consistently retain them, we must therefore also retain the term linear in ε_2 and the one quadratic in ε_1, when expanding the exponential. These involve the following quantities:

$$\sum_{j=1}^{n}(1-\frac{s_j}{s})\sum_{k,l,m=1}^{n}\int_{-\infty}^{\delta_c} d\delta_1 \dots d\delta_n \partial_j\partial_k\partial_l\partial_m W^{\rm gm} = -\left(\frac{2}{\pi}\right)^{1/2}\frac{1}{s^2}\nu\, e^{-\nu^2/2}\,, \quad (4.36a)$$

$$\sum_{j,k=1}^{n}(1-\frac{s_j}{s})(1-\frac{s_k}{s})\sum_{l,l_1,l_2,l_3=1}^{n}\int_{-\infty}^{\delta_c} d\delta_1 \dots d\delta_n \partial_j\partial_k\partial_l\partial_{l_1}\partial_{l_2}\partial_{l_3} W^{\rm gm} = -\left(\frac{2}{\pi}\right)^{1/2}\frac{4}{s^3}\nu\, e^{-\nu^2/2}\,, \quad (4.36b)$$

where we have used the result (4.32), and in Eqn. (4.36b) also the identity

$$\partial_\nu^3 h(\nu) = -\nu\, e^{-\nu^2/2}\,. \quad (4.37)$$

We now see that the result of the path integral depends only on ν. Putting things together and computing the overall ν derivative, we find

$$f = \left(\frac{2}{\pi}\right)^{1/2}\nu\, e^{-(\varepsilon_1/3!)\partial_\nu^3+(\varepsilon_2/4!)\partial_\nu^4+\dots}\left[e^{-\nu^2/2} - \frac{1}{4}\varepsilon_1\nu\, e^{-\nu^2/2} + \frac{5}{16}\varepsilon_1\left(\frac{\pi}{2}\right)^{1/2}\mathrm{erfc}\left(\frac{\nu}{\sqrt{2}}\right) + \frac{1}{8}\left(\varepsilon_1^2 - \frac{2}{3}\varepsilon_2\right)e^{-\nu^2/2}\left(\nu^2-1\right) + \mathcal{O}(\epsilon^3\nu^3)\right], \quad (4.38)$$

where we ignore terms like $\varepsilon_1\nu\mathcal{O}(\nu^{-4})$. To be consistent, we should also expand the complementary error function up to terms of order $\epsilon^2\nu^2$. However, we will see that it is more convenient to leave this term as is for now, and truncate its effects *after* we perform the saddle point integral below. Also we will find that the term $\sim \epsilon^2$ which we have retained above will cancel in the saddle point calculation.

To compute the action of the exponentiated derivative operators, we start by writing the expression in square brackets in Eqn. (4.38) in terms of its Fourier transform, using

the relations[1]

$$\begin{aligned} e^{-\nu^2/2} &= \int_{-\infty}^{\infty} \frac{d\lambda}{\sqrt{2\pi}} e^{i\lambda\nu} e^{-\lambda^2/2}\,, \\ -\nu\, e^{-\nu^2/2} &= \int_{-\infty}^{\infty} \frac{d\lambda}{\sqrt{2\pi}} (i\lambda) e^{i\lambda\nu} e^{-\lambda^2/2}\,, \\ \nu^2 e^{-\nu^2/2} &= -\int_{-\infty}^{\infty} \frac{d\lambda}{\sqrt{2\pi}} (\lambda^2 - 1) e^{i\lambda\nu} e^{-\lambda^2/2}\,, \\ \left(\frac{\pi}{2}\right)^{1/2} \mathrm{erfc}\left(\frac{\nu}{\sqrt{2}}\right) &= \int_{-\infty}^{\infty} \frac{d\lambda}{\sqrt{2\pi}} \frac{i}{\lambda} e^{i\lambda\nu} e^{-\lambda^2/2}\,. \end{aligned} \tag{4.39}$$

Together with the identity $e^{A(-d/d\nu)^n} e^{i\lambda\nu} = e^{A(-i\lambda)^n} e^{i\lambda\nu}$, for constant A and B, this gives

$$f(\nu) = \left(\frac{2}{\pi}\right)^{1/2} \nu \int_{-\infty}^{\infty} \frac{d\lambda}{\sqrt{2\pi}} e^{i\lambda\nu} e^{-\lambda^2/2 + (-i\lambda)^3 \varepsilon_1/6 + (-i\lambda)^4 \varepsilon_2/24 + \ldots} \mathcal{P}(\lambda) \tag{4.40}$$

where $\mathcal{P}(\lambda)$ is the truncated series given by

$$\mathcal{P}(\lambda) = 1 + \frac{1}{4} i\varepsilon_1 \lambda + \frac{5}{16} \frac{i\varepsilon_1}{\lambda} - \frac{1}{4}\lambda^2 \left(\frac{\varepsilon_1^2}{2} - \frac{\varepsilon_2}{3}\right) + \ldots \tag{4.41}$$

The integral in eq. (4.40) can be performed using the saddle point approximation. We write it as

$$f(\nu) = \left(\frac{2}{\pi}\right)^{1/2} \nu \int_{-\infty}^{\infty} \frac{d\lambda}{\sqrt{2\pi}} e^{\phi(\lambda)}\,, \tag{4.42}$$

where

$$\phi(\lambda) \equiv i\lambda\nu - \frac{1}{2}\lambda^2 + \frac{i\varepsilon_1}{6}\lambda^3 + \frac{\varepsilon_2}{24}\lambda^4 + \ln \mathcal{P}(\lambda) + \ldots \tag{4.43}$$

The location of the saddle point, $\lambda = \lambda_*$, is the solution of $\phi'(\lambda_*) = 0$, and the saddle point approximation then tells us that

$$\int_{-\infty}^{\infty} \frac{d\lambda}{\sqrt{2\pi}} e^{\phi(\lambda)} = e^{\phi(\lambda_*)} (|\phi''(\lambda_*)|)^{-1/2}\,, \tag{4.44}$$

(see Appendix 4.B for a discussion of the errors introduced by this approximation). It turns out that in order to obtain $f(\nu)$ correctly up to order $\sim \epsilon^2\nu^2$, we only need λ_* correct up to order $\sim \epsilon\nu$. The expression for ϕ' at the relevant order is,

$$\phi'(\lambda) = i\nu - \lambda + \frac{i\varepsilon_1}{2}\lambda^2 + \ldots\,, \tag{4.45}$$

and solving for λ_* perturbatively up to order $\epsilon\nu$, we find

$$\lambda_* = i\nu \left[1 - \frac{1}{2}\varepsilon_1 \nu + \mathcal{O}(\epsilon^2\nu^2)\right]\,. \tag{4.46}$$

[1] We are using a regulator which shifts the pole at $\lambda = 0$ in the last expression in Eqn. (4.39), to $\lambda = -i\alpha$ where α is real, positive and small.

The expression for the mass function $f(\nu)$ then works out to

$$f(\nu) = \left(\frac{2}{\pi}\right)^{1/2} \nu \exp\left[-\frac{1}{2}\nu^2\left(1 - \frac{\varepsilon_1}{3}\nu + \frac{1}{4}\left(\varepsilon_1^2 - \frac{\varepsilon_2}{3}\right)\nu^2 + \mathcal{O}(\epsilon^3\nu^3)\right)\right] \times \left(1 - \frac{1}{4}\varepsilon_1\nu\left(3 - \frac{5}{4\nu^2}\right) + \left(\varepsilon_1^2 - \frac{\varepsilon_2}{3}\right)\nu^2 + \mathcal{O}(\epsilon^3\nu^3)\right), \tag{4.47}$$

which superficially at least, is comprised of *two* series expansions, one in the exponential and one as a polynomial, both based on the small parameter $\epsilon\nu$ (see however the next section).

This derivation assumed that ε_1 and ε_2 are constant, and that the c_n are unity. If we relax these assumptions and allow a scale dependence in these parameters, Eqn. (4.33) is replaced with

$$\begin{aligned} f &= (\nu\partial_\nu - 2s\partial_s)e^{-(\varepsilon_1(s)/3!)\partial_\nu^3 + (\varepsilon_2(s)/4!)\partial_\nu^4 + \cdots} g(\nu, t) \\ &= \left[\nu + \frac{1}{3}\dot{\varepsilon}_1\varepsilon_1\partial_\nu^2 - \frac{1}{12}\dot{\varepsilon}_2\varepsilon_2\partial_\nu^3\right] e^{-(\varepsilon_1(s)/3!)\partial_\nu^3 + (\varepsilon_2(s)/4!)\partial_\nu^4 + \cdots} \partial_\nu g(\nu, t) \\ &\qquad - 2te^{-(\varepsilon_1(s)/3!)\partial_\nu^3 + (\varepsilon_2(s)/4!)\partial_\nu^4 + \cdots} \partial_s g(\nu, s), \end{aligned} \tag{4.48}$$

where, for any function $v(s)$, the dot is defined as

$$\dot{v}(t) \equiv \frac{d\ln v}{d\ln s}, \tag{4.49}$$

(recall $s = \sigma^2$) and the function $g(\nu, s)$ can be shown to be

$$g(\nu, s) = \left(\frac{2}{\pi}\right)^{1/2}\left[\left(\frac{\pi}{2}\right)^{1/2}\mathrm{erf}\left(\frac{\nu}{\sqrt{2}}\right) + \frac{1}{4}\varepsilon_1 c_1 e^{-\nu^2/2} + \frac{\varepsilon_1}{4}\left(\frac{3}{4}c_2 - 2c_3\right)h(\nu) - \frac{1}{8}\varepsilon_1^2 c_1^2 \nu e^{-\nu^2/2} + \frac{1}{12}\varepsilon_2 c_4 \nu e^{-\nu^2/2} + \ldots\right], \tag{4.50}$$

The expression in Eqn. (4.48) can be evaluated analogously to Eqn. (4.38), since the additional derivatives pose no conceptual difficulty. The result of the saddle point calculation, correct up to quadratic order assuming $\epsilon\nu^3 \sim \mathcal{O}(1)$, is

$$\begin{aligned} f(\nu, s) = \left(\frac{2}{\pi}\right)^{1/2} \nu \exp\left[-\frac{1}{2}\nu^2\left(1 - \frac{\varepsilon_1}{3}\nu + \frac{1}{4}\left(\varepsilon_1^2 - \frac{\varepsilon_2}{3}\right)\nu^2 + \mathcal{O}(\epsilon^3\nu^3)\right)\right] \\ \times\left\{1 - \frac{1}{4}\varepsilon_1\nu\left(\left(c_1 + 2 - \frac{4}{3}\dot{\varepsilon}_1\right) + \frac{1}{\nu^2}\left(\frac{3}{4}c_2 - 2c_3 + \frac{4}{3}\dot{\varepsilon}_1 + 2c_1(\dot{\varepsilon}_1 + \dot{c}_1)\right)\right)\right. \\ + \frac{1}{8}\nu^2\left(\varepsilon_1^2\left(c_1^2 + 2c_1 + 5 - \frac{2}{3}\dot{\varepsilon}_1(c_1 + 6)\right) - 2\varepsilon_2\left(1 + \frac{1}{3}c_4 - \frac{1}{3}\dot{\varepsilon}_2\right)\right) \\ \left. + \mathcal{O}(\epsilon^3\nu^3)\right\}, \end{aligned} \tag{4.51}$$

which reduces to Eqn. (4.47) if we take ε_1, ε_2 to be constant and set the c_n to unity.

One issue which we have ignored so far, is that the definition of ν involves the variance $s = \sigma^2$ of the *non-Gaussian* field. Computationally it is more convenient to work with the variance $\sigma_{\rm g}^2$ of the *Gaussian* field in terms of which cosmological NG are typically defined. We should then ask whether this difference will require changes in our expressions for f. We start by noting that this difference in variances is of order $\sim \epsilon^2$. For example, in the local model one has $\sigma^2(R) = Ad_1(R) + A(Af_{NL}^2)d_2(R)$ where $A \sim 10^{-9}$ is an overall normalization constant, d_1 and d_2 are scale dependent functions of comparable magnitude on all relevant scales, and ϵ is estimated as $\epsilon \sim f_{NL}A^{1/2}$. We therefore have $\nu = \delta_c/\sigma = (\delta_c/\sigma_{\rm g})(1 + \mathcal{O}(\epsilon^2))$. However, with our assumption that $\epsilon\nu^3 \sim \mathcal{O}(1)$, we see that this correction is actually of order $\sim (\epsilon^2\nu^2)\nu^{-2} \sim \epsilon^3\nu^3$, which we have been consistently ignoring. We will see that even when we relax the assumption $\epsilon\nu^3 \sim \mathcal{O}(1)$ and probe smaller scales where $\epsilon\nu^3 \ll 1$, this correction can still be consistently ignored. Hence we can safely set $\nu = \delta_c/\sigma_{\rm g}$ in all of our expressions.

4.4 Consistency of the truncation

4.4.1 Comparative sizes of terms in the mass function

Now that all the derivative operators which we consider important have been accounted for, we can check whether our final result is consistently truncated, i.e. whether we have retained *all* terms at any given order in the expansion. Symbolically, our current result for the mass function can be written as

$$f \sim e^{-\frac{1}{2}\nu^2\left(1+\epsilon\nu+\epsilon^2\nu^2+\mathcal{O}(\epsilon^2,\epsilon^3\nu^3)\right)}\left[1+\epsilon\nu+\frac{\epsilon}{\nu}+\epsilon^2\nu^2+\mathcal{O}(\epsilon\nu^{-3},\epsilon^2,\epsilon^3\nu^3)\right], \qquad (4.52)$$

with the understanding that coefficients are computed (but not displayed) for all terms except those indicated by the $\mathcal{O}()$ symbols. Also, ϵ^2 refers to both ε_1^2 and ε_2.

Since the expansions involve two parameters, $\epsilon\nu$ and ν^{-2}, they make sense only if we additionally prescribe a relation between these parameters. So far we assumed that ϵ is fixed and ν is such that $\epsilon\nu^3 \simeq 1$, which was based on the observation that the term $\epsilon\nu^3$ naturally appears in the exponent and is not restricted in principle to small values. In this case, in the polynomial in (4.52) we retain the terms $\epsilon\nu \simeq \nu^{-2}$, $(\epsilon\nu^{-1}, \epsilon^2\nu^2) \simeq \nu^{-4}$, and we discard $(\epsilon\nu^{-3}, \epsilon^2, \epsilon^3\nu^3) \simeq \nu^{-6}$. It would seem that our expression is then correct upto order $\sim \nu^{-4}$. However, the terms discarded in the exponential have the form $\exp(\mathcal{O}(\epsilon^3\nu^5)) \sim \exp(\mathcal{O}(\nu^{-4})) \sim 1 + \mathcal{O}(\nu^{-4})$. The error we are making is thus of the same order as the smallest terms we are retaining, and it therefore makes sense to *also* ignore all the terms of order $\sim \nu^{-4}$ which we computed in the polynomial. The consistent expression when $\epsilon\nu^3 \simeq 1$ is then given by

$$f \sim e^{-\frac{1}{2}\nu^2\left(1+\epsilon\nu+\epsilon^2\nu^2\right)}\left[1+\epsilon\nu+\mathcal{O}\left(\nu^{-4}\right)\right]. \qquad (4.53)$$

Clearly, similar arguments can be applied at smaller scales where, e.g. one might have $\epsilon\nu^3 \simeq \nu^{-1}, \nu^{-2}$, etc. It is then important to ask which mass scales correspond

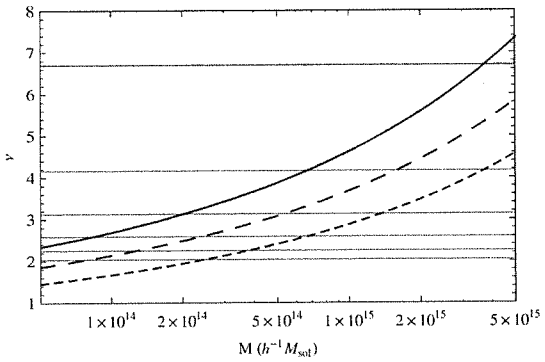


Figure 4.4: $\nu \equiv \delta_c(z)/\sigma(M)$ in the range $5 \cdot 10^{13} < (M/h^{-1}M_{\rm sol}) < 5 \cdot 10^{15}$ for three different redshifts, with $\epsilon = 1/300$. The solid, long dashed and short dashed curves correspond to redshifts $z = 1$, 0.5 and 0 respectively. The horizontal lines mark the transition points where $\epsilon\nu^3$ becomes equal to (from top to bottom) 1, ν^{-1}, ν^{-2}, ν^{-3}, ν^{-4} and ν^{-5}.

to these "transition points". In Fig. 4.4 we plot $\nu(M, z)$ given by Eqn. (4.25) in an observationally interesting mass range, for three different redshifts. The horizontal lines mark the transition points where $\epsilon\nu^3$ becomes equal to (from top to bottom) 1, ν^{-1}, ν^{-2}, ν^{-3}, ν^{-4} and ν^{-5}. We fix $\epsilon = 1/300$ which follows from the fact that in the local model with $f_{NL} = 100$ we have $\varepsilon_1 \simeq 0.02$ (see Fig. 4.1), and the expression for $f(\nu, M)$ contains the quantity $\varepsilon_1/6$ in the exponential. From the intersections of the horizontal lines with the curves, we see that different transition points are relevant at different redshifts, and their locations also obviously depend on the value of ϵ. For example, we find that the transition point where $\epsilon\nu^3 \simeq \nu^{-2}$, remains accessible even when ϵ is an order of magnitude smaller (with $\epsilon \simeq 1/3000$, this transition occurs at $\nu \simeq 4.96$). The transitions at $\epsilon\nu^3 \simeq 1, \nu^{-1}$ on the other hand, are not accessible for this level of NG. The transition at $\epsilon\nu^3 \simeq \nu^{-2}$ is therefore observationally very interesting.

We will now discuss in some detail the truncation of our expression for f, at various transition points. The goal is to try and settle on a *single* expression which is valid over a wide range of scales (i.e. across several transition points). This can then be applied without worrying about truncation inconsistencies. Of course, the order of the discarded terms will then depend on the particular transition point being considered, leading to a scale dependent theoretical error. At this point, the reader may skip to the end of the present subsection, where we present such a single consistent expression.

4.4.1.1 $\epsilon\nu^3 \simeq \nu^{-1}$

At this transition point, the terms we retain in the exponential are

$$\epsilon\nu^3 \simeq \nu^{-1} \quad ; \quad \epsilon^2\nu^4 \simeq \nu^{-4} \,,$$

while discarding $\mathcal{O}(\epsilon^3\nu^5) = \mathcal{O}(\nu^{-7})$. In the polynomial meanwhile, we retain

$$\epsilon\nu \simeq \nu^{-3} \;\; ; \;\; \epsilon\nu^{-1} \simeq \nu^{-5} \;\; ; \;\; \epsilon^2\nu^2 \simeq \nu^{-6} ,$$

while discarding

$$\mathcal{O}(\epsilon\nu^{-3}) = \mathcal{O}(\nu^{-7}) \;\; ; \;\; \mathcal{O}(\epsilon^2) = \mathcal{O}(\nu^{-8}) \;\; ; \;\; \mathcal{O}(\epsilon^3\nu^3) = \mathcal{O}(\nu^{-9}) .$$

Our expression (4.52) therefore retains all terms correctly up to order $\sim \nu^{-6}$, and is consistent. With some foresight however, it turns out to be more convenient to degrade this expression somewhat by also discarding the polynomial quadratic term $\epsilon^2\nu^2 \simeq \nu^{-6}$. The remaining expression,

$$f \sim e^{-\frac{1}{2}\nu^2\left(1+\epsilon\nu+\epsilon^2\nu^2\right)} \left[1 + \epsilon\nu + \frac{\epsilon}{\nu} + \mathcal{O}\left(\nu^{-6}\right)\right] , \tag{4.54}$$

is also consistent at this transition point, and has a form which is identical to the ones we will see next.

4.4.1.2 $\epsilon\nu^3 \simeq \nu^{-2}$

As we mentioned earlier, this transition point is observationally quite interesting. The terms we retain in the exponential are

$$\epsilon\nu^3 \simeq \nu^{-2} \;\; ; \;\; \epsilon^2\nu^4 \simeq \nu^{-6} ,$$

while discarding $\mathcal{O}(\epsilon^3\nu^5) = \mathcal{O}(\nu^{-10})$, and in the polynomial we retain

$$\epsilon\nu \simeq \nu^{-4} \;\; ; \;\; \epsilon\nu^{-1} \simeq \nu^{-6} \;\; ; \;\; \epsilon^2\nu^2 \simeq \nu^{-8} ,$$

while discarding

$$\mathcal{O}(\epsilon\nu^{-3}) = \mathcal{O}(\nu^{-8}) \;\; ; \;\; \mathcal{O}(\epsilon^2) = \mathcal{O}(\nu^{-10}) \;\; ; \;\; \mathcal{O}(\epsilon^3\nu^3) = \mathcal{O}(\nu^{-12}) .$$

This time we see that the term $\epsilon\nu^{-3}$ has become as important as the quadratic term $\epsilon^2\nu^2$ in the polynomial, and to be consistent we should discard the quadratic term. The expansion should read

$$f \sim e^{-\frac{1}{2}\nu^2\left(1+\epsilon\nu+\epsilon^2\nu^2\right)} \left[1 + \epsilon\nu + \frac{\epsilon}{\nu} + \mathcal{O}\left(\nu^{-8}\right)\right] . \tag{4.55}$$

4.4.1.3 $\epsilon\nu^3 \simeq \nu^{-3}$

A similar analysis as above shows that at this stage $\epsilon\nu^{-3} \simeq \nu^{-9} > \epsilon^2\nu^2$, and a consistent expression again requires dropping the quadratic term in the polynomial, leaving

$$f \sim e^{-\frac{1}{2}\nu^2\left(1+\epsilon\nu+\epsilon^2\nu^2\right)} \left[1 + \epsilon\nu + \frac{\epsilon}{\nu} + \mathcal{O}\left(\nu^{-9}\right)\right] . \tag{4.56}$$

4.4.1.4 $\epsilon\nu^3 \simeq \nu^{-4}$ and smaller

Beyond this point, the term $\epsilon\nu^{-3}$ which we discard in the polynomial, becomes comparable or larger than the quadratic term of the exponential as well, and a consistent expression becomes

$$f \sim e^{-\frac{1}{2}\nu^2(1+\epsilon\nu)}\left[1+\epsilon\nu+\frac{\epsilon}{\nu}+\ldots\right] \tag{4.57}$$

The parametric order of the terms now discarded, depends on the exact relation between $\epsilon\nu^3$ and ν^{-1}.

Finally, note that the error introduced by setting $\nu \to \nu_{\rm g}$ where $\nu_{\rm g}$ is defined using the variance of a *Gaussian* field, was estimated in section 4.3 as $\mathcal{O}(\epsilon^2)$. When $\epsilon\nu^3 \simeq 1$, this error is of order $\mathcal{O}(\epsilon^3\nu^3)$ and can therefore be consistently ignored. It is not hard to see that at *all* lower transition points, this error continues to be comparable to or smaller than the largest terms being discarded, and can hence be consistently ignored. In summary, we can state that for observationally accessible mass scales *larger* than the transition point where $\epsilon\nu^3 \simeq \nu^{-3}$, the single expression

$$f \sim e^{-\frac{1}{2}\nu^2\left(1+\epsilon\nu+\epsilon^2\nu^2\right)}\left[1+\epsilon\nu+\frac{\epsilon}{\nu}+\mathcal{O}(\epsilon^3\nu^5,\epsilon^2\nu^2,\epsilon\nu^{-3})\right], \tag{4.58}$$

is parametrically consistent as it stands – the terms ignored are smaller than the smallest terms retained – and in fact it remains a very good approximation even when $\epsilon\nu^3 \simeq 1$, since the only "inconsistent" term then is $\epsilon\nu^{-1}$, whose effect *reduces* as ν increases. On scales where $\epsilon\nu^3 \simeq \nu^{-4}$ and lower, the theoretical error becomes comparable to or larger than the quadratic term in the exponential. Plugging back all the coefficients, we have the following result for the mass function (excluding filter effects, see section 4.5),

$$\begin{aligned} f(\nu,t) = & f_{\rm PS}(\nu)\,\exp\left(\frac{1}{6}\varepsilon_1\nu^3-\frac{1}{8}\left(\varepsilon_1^2-\frac{\varepsilon_2}{3}\right)\nu^4\right) \\ & \times\left\{1-\frac{1}{4}\varepsilon_1\nu\left(\left(c_1+2-\frac{4}{3}\dot{\varepsilon}_1\right)+\frac{1}{\nu^2}\left(\frac{3}{4}c_2-2c_3+\frac{4}{3}\dot{\varepsilon}_1+2c_1(\dot{\varepsilon}_1+\dot{c}_1)\right)\right)\right. \\ & \left. +\ \mathcal{O}(\epsilon^3\nu^5,\epsilon^2\nu^2,\epsilon\nu^{-3})\right\}. \end{aligned} \tag{4.59}$$

4.4.2 Comparing with previous work

In this subsection we compare our results with previous work on the non-Gaussian mass function. As mentioned in the Introduction, this quantity has been computed by several authors in different ways [104; 105; 109]. If one considers the range of validity of the perturbative expansion, the strongest result so far has been due to MVJ [104], who explicitly retain the exponential dependence on ε_1. Their expression for f can be

written as[1]

$$f_{\rm MVJ} = f_{\rm PS}(\nu)\frac{e^{\varepsilon_1\nu^3/6}}{(1-\varepsilon_1\nu/3)^{1/2}}\left(1-\frac{1}{2}\varepsilon_1\nu\left(1-\frac{2}{3}\dot{\varepsilon}_1\right)\right) . \tag{4.60}$$

The major shortcoming of their result is that it is based on a Press-Schechter like prescription, and must therefore be normalized by an appropriate Gaussian mass function, typically taken to be the one due to Sheth & Tormen [106]. Additionally, it always misses the contributions due to the unequal time correlators, which contribute to the terms $\sim \epsilon\nu$, $\epsilon\nu^{-1}$, etc. in Eqn. (4.59). When one considers formal correctness on the other hand, MR have presented a result based on explicit path integrals, which accounts for the unequal time contributions, and which also does not need any *ad hoc* normalizations (in this context see also Lam & Sheth [113]). Indeed, our calculations in section 4.3 were based on techniques discussed by MR in Refs. [107; 109]. As we discuss below however, the fact that MR do not explicitly retain the exponential dependence of $\varepsilon_1\nu^3$, means that their result is subject to significant constraints on the range of its validity. Their expression for f, ignoring filter effects, is[2]

$$f_{\rm MR} = f_{\rm PS}(\nu)\Bigg(1+\frac{1}{6}\varepsilon_1\nu^3\Bigg\{1-\frac{3}{2\nu^2}\left(c_1+2-\frac{4}{3}\dot{\varepsilon}_1\right) - \frac{3}{2\nu^4}\left(\frac{3}{4}c_2-2c_3+4\dot{\varepsilon}_1+2c_1(\dot{\varepsilon}_1+\dot{c}_1)\right)\Bigg\}\Bigg) . \tag{4.61}$$

This expression is precisely what one obtains by linearizing our expression (4.59) in ε_1, which serves as a check on our calculations. LMSV [105] present a result based on an Edgeworth expansion of the type encountered when studying NG generated by nonlinear gravitational effects [124]. The result most often quoted in the literature is their expression linear in ε_1 (and hence in $\varepsilon_1\nu^3$), which is

$$f_{\rm LMSV,lin} = f_{\rm PS}(\nu)\left(1+\frac{1}{6}\varepsilon_1\nu^3\left\{1-\frac{1}{\nu^2}(3-2\dot{\varepsilon}_1)-\frac{2}{\nu^4}\dot{\varepsilon}_1\right\}\right) . \tag{4.62}$$

In Appendix B.3 of Ref. [105], LMSV also give an expression involving ε_1^2 and ε_2, which can be written as

$$f_{\rm LMSV,quad} = f_{\rm PS}(\nu)\Bigg[1+\frac{1}{6}\varepsilon_1\left(H_3(\nu)+\frac{2}{\nu}\dot{\varepsilon}_1 H_2(\nu)\right) + \frac{1}{72}\varepsilon_1^2\left(H_6(\nu)+\frac{4}{\nu}\dot{\varepsilon}_1 H_5(\nu)\right)+\frac{1}{24}\varepsilon_2\left(H_4(\nu)+\frac{2}{\nu}\dot{\varepsilon}_2 H_3(\nu)\right)\Bigg] , \tag{4.63}$$

[1]The analysis presented by MVJ in fact allows one to retain terms like $\sim \epsilon^2\nu^4$ in the exponential as well, and we have seen that when $\epsilon\nu^3 \simeq 1$, these terms are as important as the polynomial $\epsilon\nu$ term retained by MVJ. However, since the MVJ expression misses unequal time effects of order $\sim \epsilon\nu$ anyway, it is reasonable to compare our results with the expression (4.60), which is also the one used by most other authors (see e.g. Refs. [122; 123]).

[2]We have corrected a typo in MR's result [109] : the object they define as $\mathcal{V}_3$ should appear with an overall positive coefficient in their Eqns. (85), (87) and (92).

where the $H_n(\nu)$ are the Hermite polynomials of order n. This expression was used by LMSV only as a check on the validity of their *linear* expression. By comparing with our expression which is non-perturbative in $\varepsilon_1\nu^3$, we will see below that these quadratic terms in fact significantly improve LMSV's prediction.

Sticking to the linearized results, we see that the expressions of both MR and LMSV have the symbolic form

$$f \sim e^{-\nu^2/2}\left[1 + \epsilon\nu^3 + \epsilon\nu + \frac{\epsilon}{\nu} + \ldots\right], \tag{4.64}$$

where the ellipsis denotes all terms of the type $\epsilon\nu^{-3}$, $\epsilon\nu^{-5}$, etc., as well as all terms containing ϵ^2. As we have seen, deciding where to truncate the expression for f is not trivial, and using our more detailed expression we can ask whether the expression (4.64) is consistent at all the relevant length scales. Immediately, we see that this expression cannot be correct once $\epsilon\nu^3$ becomes close to unity. However, this case is on the border of the observed mass window (for galaxy cluster observations), even at high redshifts.

Let us therefore directly look at the case $\epsilon\nu^3 \simeq \nu^{-2}$ which, as we saw, is accessible over a wide range of redshifts for $\epsilon \sim 10^{-2}$, and at high redshifts also for $\epsilon \sim 10^{-3}$. In this case the terms MR and LMSV retain have magnitudes

$$\epsilon\nu^3 \simeq \nu^{-2}\ ;\ \ \epsilon\nu \simeq \nu^{-4}\ ;\ \ \epsilon\nu^{-1} \simeq \nu^{-6}\,,$$

and terms like $\epsilon\nu^{-3} \simeq \nu^{-8}$ are discarded. We know from our expression however, that $\epsilon\nu^3$ appears in the exponential, and therefore leads to terms like $(\epsilon\nu^3)^2 \simeq \nu^{-4}$ and $(\epsilon\nu^3)^3 \simeq \nu^{-6}$ when the exponential is expanded, which are of the same order as the terms retained in (4.64). The exponential also contributes a term $\epsilon^2\nu^4 \simeq \nu^{-6}$, which in fact involves the *tri*spectrum of NG, again at the order retained by MR and LMSV. The error in the expression (4.64) when $\epsilon\nu^3 \simeq \nu^{-2}$, is therefore $\mathcal{O}(\epsilon\nu)$. (A similar analysis shows that the error at transition point where $\epsilon\nu^3 \simeq \nu^{-1}$, is $\mathcal{O}(\nu^{-2}) > \mathcal{O}(\epsilon\nu)$.)

From a purely parametric point of view, the situation for MR and LMSV improves as ν is decreased further, and the expression (4.64) as it stands, becomes exactly consistent (in the sense discussed in the previous subsection, see below Eqn. (4.58)) when $\epsilon\nu^3 \simeq \nu^{-5}$, because at this stage $\epsilon\nu^{-1} \simeq \nu^{-9}$ while $(\epsilon\nu^3)^2 \simeq \nu^{-10}$ and $\epsilon^2\nu^4 \simeq \nu^{-12}$, and hence the exponential only contributes a single linear term $\epsilon\nu^3$. More importantly, LMSV's expression also has errors due to the absence of the unequal time terms discussed earlier, which are of order $\sim \epsilon\nu$ and can be dominant over the others. For the intermediate transitions, the analysis shows that when $\epsilon\nu^3 \simeq \nu^{-3}$, the error in (4.64) is $\mathcal{O}(\nu^{-6}) > \mathcal{O}(\epsilon\nu^{-1})$, and when $\epsilon\nu^3 \simeq \nu^{-4}$, the error is $\mathcal{O}(\epsilon\nu^{-1})$. This should be compared with our result (4.59), in which the error (at least on large scales) is always parametrically *smaller* than the smallest terms we retain.

4.5 Effects of the diffusing barrier and the filter

In Ref. [108], MR showed that the agreement between a Gaussian mass function calculated using the statistics of random walks, and mass functions observed in numerical

simulations with Gaussian initial conditions, can be dramatically improved by allowing the barrier itself to perform a random walk. This approach is motivated by the fact that the ignorance of the details of the collapse introduces a scatter in the value of the collapse threshold for different virialized objects. The width of this scatter was found by Robertson *et al.* [125] to be a growing function of $\sigma(M)$, which is consistent with the physical expectation that deviations from spherical collapse become relevant at small scales. The barrier can thus be treated (at least on a first approximation) as a stochastic variable whose probability density function obeys a Fokker-Planck equation with a diffusion coefficient D_B, which can be estimated numerically in a given N-body simulation. In particular, MR found $D_B \simeq 0.25$ using the simulations of Ref. [125].

Conceptually, the variation of the value of the barrier is due to two types of effects, one intrinsically physical and one more inherent to the way in which one interprets the results of simulations. From a physical point of view, the dispersion accounts for deviations from the simple model of spherical collapse, for instance the effects of ellipsoidal collapse, baryonic physics, etc. On the other hand, the details of the distribution of the barrier (and therefore the precise value of D_B) will depend on the halo finder algorithm used to identify halos in a particular simulation, since different halo finders identify collapsed objects with different properties. MR concluded that the final effect of this barrier diffusion on large scales can be accounted for in a simple way, by changing $\delta_{c0} \to \sqrt{a}\delta_{c0}$ where $a = (1 + D_B)^{-1}$. In practice this change is identical to the one proposed by Sheth *et al.* [112][1]. As MR argue in section 3.4 of Ref. [109], this barrier diffusion effect can also be accounted for in the *non*-Gaussian case, again by the simple replacement of $\delta_{c0} \to \sqrt{a}\delta_{c0}$. It is easy to see that their arguments go through for all our calculations as well, and we have implemented this change in our definition of ν in Eqn. (4.25), setting $\sqrt{a} = 0.89$.

In Ref. [107], MR also accounted for the non-Markovian effects of the real space top-hat filter, as opposed to the sharp-k filter for which the results of section (4.3) apply. This is done by writing the 2-point function $\langle\, \hat{\delta}(R_1)\, \hat{\delta}(R_2)\,\rangle$ calculated using the real space top-hat filter, as the Markovian value plus a correction, $\langle\, \hat{\delta}(R_j)\, \hat{\delta}(R_k)\,\rangle = \min(s_j, s_k) + \Delta_{jk}$, and noting that the correction Δ_{jk} remains small over the interesting range of length scales. In fact, MR show that a very good analytical approximation for the symmetric object Δ_{jk}, is

$$\Delta_{jk} \simeq \kappa \min(s_j, s_k) \left(1 - \frac{\min(s_j, s_k)}{\max(s_j, s_k)}\right) , \tag{4.65}$$

[1] A potential issue in this argument lies in the assumption of a linear Langevin equation for the stochastic barrier B, resulting in a simple Fokker-Planck equation with a constant D_B like the one in MR, while the distribution of B was found to be approximately log-normal (and therefore non-Gaussian) in Ref. [125]. One can see that a Langevin equation of the type $\dot{B} = B\xi$ (which would produce a log-normal distribution) can be approximated by $\dot{B} = \langle B\rangle\xi$, whenever the fluctuations around $\langle B\rangle$ are small, and gives a constant diffusion coefficient as long as $\langle B\rangle$ is constant. Although both approximations are reasonable on the scales of interest, non-Gaussian and scale dependent corrections to the barrier diffusion should be studied, since in principle they could be of the same order as the other corrections retained here. This investigation is left for future work.

where in our case we find $\kappa(R) \simeq 0.464 + 0.002R$, with R measured in h^{-1}Mpc. The mass function is then obtained by perturbatively expanding in Δ_{ij}, with the leading effect being due to the integral

$$\int_{-\infty}^{\delta_c} d\delta_1 \dots d\delta_n \frac{1}{2} \sum_{j,k=1}^{n} \Delta_{jk} \partial_j \partial_k W^{\rm gm} ,$$

which on evaluation leads to

$$f_{\rm g,sharp-x}(\nu, t) = \left(\frac{2}{\pi}\right)^{1/2} \nu \left[(1-\kappa) e^{-\nu^2/2} + \frac{\kappa}{2} \Gamma\left(0, \frac{\nu^2}{2}\right) + \mathcal{O}(\kappa^2)\right] , \qquad (4.66)$$

where the subscript stands for Gaussian noise with the top-hat filter in real space, and κ introduces a weak explicit $s(=\sigma^2)$ dependence. In Ref. [109] MR proposed an extension of this result to the non-Gaussian case, by assuming that *all* the non-Gaussian terms that they computed with the sharp-k filter, would simply get rescaled by the factor $(1-\kappa)$ at the lowest order, but otherwise retain their coefficients. Symbolically, their result (Eqn. 88 of Ref. [109]) is

$$f_{\rm ng,sharp-x}(\nu, s) \sim \nu \left[(1-\kappa) e^{-\nu^2/2} \left(1 + \epsilon\nu^3 + \epsilon\nu + \epsilon\nu^{-1}\right) + \frac{\kappa}{2} \Gamma\left(0, \frac{\nu^2}{2}\right)\right] , \qquad (4.67)$$

with the specific coefficients of the $\epsilon\nu^3$, $\epsilon\nu$ and $\epsilon\nu^{-1}$ terms being identical to those in Eqn. (4.61). However, the coefficient of e.g. the $\kappa\epsilon\nu$ term arises from the action of an operator $\sim \sum_{j,k} \Delta_{jk} \partial_j \partial_k$ combining with the first unequal time operator $\sim \varepsilon_1 t^{1/2} \sum_j (s - s_j) \partial_j \sum_{k,l} \partial_k \partial_l$, and there is no simple way of predicting its exact value beforehand. Since MR explicitly neglect such "mixed" terms, their formula is not strictly inconsistent, as long as one keeps in mind that the theoretical error in their expression is of the same order as the terms $\sim \kappa\epsilon\nu$ that they include. However, if one wants to consistently retain such terms, a detailed calculation is needed[1]. Our calculations (not displayed) indicate that the coefficient of the $\kappa\epsilon\nu$ term depends on certain details of the continuum limit of the path integral near the barrier, which require a more careful study. We are currently investigating methods of computing these effects. At present however, we conclude that the mixed terms involving both filter effects and NG, must be truncated at order $\sim \kappa\epsilon\nu$.

Finally, the filter-corrected mass function is also subject to effects of barrier diffusion. Here we make the same assumptions as MR do in Ref. [108], namely that the barrier location satisfies a Langevin equation with white noise and diffusion constant D_B, which can be accounted for by replacing $\kappa \to \tilde{\kappa} = \kappa/(1+D_B) = a\kappa$. However, it is difficult to theoretically predict the unequal time behaviour of the barrier correlations, and these simple assumptions must also be tested, perhaps by suitably comparing with the detailed results of Robertson *et al.* [125]. We leave this for future work. Our final

[1]Notice that this issue is completely decoupled from the subtleties in truncation discussed in section 4.4 – this problem remains even at scales where MR's expression is formally consistent.

expression for the mass function, corrected for effects of the diffusing barrier and the top-hat real space filter, is

$$f(\nu,s) = f_{\rm PS}(\nu)\left(1-\tilde{\kappa}+\mathcal{O}(\tilde{\kappa}^2)\right)\exp\left[\frac{1}{6}\varepsilon_1\nu^3-\frac{1}{8}\left(\varepsilon_1^2-\frac{\varepsilon_2}{3}\right)\nu^4\right] \times\left\{1+\frac{\left(1-2\dot{\tilde{\kappa}}\right)}{1-\tilde{\kappa}}\tilde{\kappa}\nu^{-2}\left(1-2\nu^{-2}\right)-\frac{1}{4}\varepsilon_1\nu\left(\frac{c_1}{1-\tilde{\kappa}}+2-\frac{4}{3}\dot{\varepsilon}_1\right) -\frac{1}{4}\varepsilon_1\nu^{-1}\left(\frac{3}{4}c_2-2c_3+\frac{4}{3}\dot{\varepsilon}_1+2c_1(\dot{\varepsilon}_1+\dot{c}_1)\right) +\mathcal{O}(\tilde{\kappa}^2\nu^{-2},\tilde{\kappa}\epsilon\nu,\tilde{\kappa}\nu^{-6})+\mathcal{O}(\epsilon^2\nu^2,\epsilon^3\nu^5,\epsilon\nu^{-3})\right\}, \quad (4.68)$$

where we have chosen to account for the scale independent $\mathcal{O}(\tilde{\kappa}^2)$ error arising from filter effects, as an overall normalization uncertainty, and have explicitly displayed the orders of the various terms we ignore. Here $f_{\rm PS}(\nu)$ is given by Eqn. (4.21) with $\nu(M,z)$ defined in Eqn. (4.25).

To summarize, Eqn. (4.68) gives an analytical expression for the non-Gaussian mass function. This expression is based on approximations that are valid over a larger range of length scales than the ones presented by MR and LMSV, and incorporates effects which are ignored in the expression presented by MVJ and LMSV. Like all these other mass functions, it suffers from the errors introduced by filter effects. However, the largest of these can be accounted for as an overall normalization constant, which can be fixed using, for instance, results of a Gaussian simulation. In Table 4.1 we provide analytical fits for ε_1, ε_2, c_1, c_2 and c_3, for the local and equilateral case as a function of σ^2. As mentioned earlier, all these quantities are independent of redshift, although they depend on the choice of cosmological parameters in a complicated way in general due to the presence of the transfer function in their definitions. However, the dependence on σ_8 is simple, and one can check that we have $\varepsilon_1 \propto \sigma_8$, $\varepsilon_2 \propto \sigma_8^2$ and that the c_n are independent of σ_8. Recall that the c_n are also independent of f_{NL} and $g_{\rm NL}$. Also, we have the following relations for $\dot{\varepsilon}_1$ and $\dot{c}_1$, which can be proved using the definitions of ε_1 and the c_n,

$$\dot{\varepsilon}_1 = \frac{3}{2}(c_1-1) \;;\; \dot{c}_1 = 1-\frac{3}{2}c_1+\frac{1}{c_1}\left(c_3-\frac{1}{2}c_2\right). \quad (4.69)$$

For completeness, in Table 4.1 we also give fits for the filter parameters $\tilde{\kappa}$ and $\dot{\tilde{\kappa}}$ which appear in the mass function, as functions of σ^2.

4.6 Results and Discussion

In this section we conclude with our final results for the non-Gaussian halo mass function, comparing our approach with previous work. In principle, we should compare the full expressions for the mass functions of various authors with ours. However,

Parameter	Fitting form $b + c\, s^n$		
Local NG	b	c	n
ε_1	0.0096	0.015	0.18
c_1	0.98	0.073	0.094
c_2	3.15	0.79	0.69
c_3	2.15	0.45	0.65
$\varepsilon_2(f_{NL}^2)$	-0.0049	0.0059	0.011
$\varepsilon_2(g_{\rm NL})$	$7.9 \cdot 10^{-4}$	0.0022	0.25

Parameter	Fitting form $b + c\, s^n$		
Equilateral NG	b	c	n
ε_1	0.01	$-4 \cdot 10^{-4}$	1.25
c_1	1.03	-0.052	0.30
c_2	2.32	0.93	0.49
c_3	1.72	0.36	0.54
Filter			
$\tilde{\kappa}$	0.36	0.015	-0.47
$\dot{\tilde{\kappa}}$	0.046	-0.064	-0.17

Table 4.1: Analytical fits for the various NG parameters, in the local and equilateral cases, as a function of $s = \sigma^2$, in the range $2 \cdot 10^{13} < M/(h^{-1} M_{\rm sol}) < 5 \cdot 10^{15}$, for $f_{NL} = 100$ and $g_{\rm NL} = 10^4$. We have $\varepsilon_1 \propto f_{NL}$ in both cases, and for ε_2 in the local case we give separate fits for the terms proportional to f_{NL}^2 and $g_{\rm NL}$. We do not consider ε_2 in the equilateral case, since the trispectrum in this case is highly model dependent. We also give fits for the filter parameters $\tilde{\kappa}$ and $\dot{\tilde{\kappa}}$ as functions of t, in the same mass range. The errors on all the fits are less than 1%, except for $\varepsilon_2(f_{NL}^2)$ where the error is $\sim 6\%$. This was due to numerical difficulties in calculating this object. These fits of course depend on our choice of cosmological parameters.

recall that for MVJ and LMSV one has to multiply an analytically predicted ratio $R_{\rm ng} = f(\nu, M, f_{NL})/f(\nu, M, f_{NL} = 0)$ with a suitable *Gaussian* mass function based on fits to simulations. It is not clear how to compute theoretical error bars on the latter. On the other hand, the object $R_{\rm ng}$ itself is an unambiguous theoretical prediction of every approach, that is MVJ, LMSV, MR and our work, and we can compute theoretical errors on it. In this work, we will restrict ourselves to comparing the different expressions for $R_{\rm ng}$. In future work, we hope to compare both $R_{\rm ng}$ and the full mass function with the results of N-body simulations.

In Fig. 4.5 and Fig. 4.6 we plot the ratio $R_{\rm ng}$, respectively without and with the filter effects, at redshift $z = 1$. In this way we can explicitly disentangle the errors due to an approximate treatment of non-Gaussian effects, from those due to the filter effects.

We compare our expression (4.68) with the expressions of MR (4.61), LMSV (4.62) and (4.63), and of MVJ (4.60). Notice that, when considering the filter effects, the Gaussian function that enters in the ratio $R_{\rm ng}$ is defined to be the function with $f_{NL} = 0$ (i.e. without NG but with filter effects when present). We use the local model, setting $f_{NL} = 100$ and $g_{\rm NL} = 0$, and use the reference ΛCDM cosmology described in section 4.2. We do not explicitly show the final results for the equilateral model, but they are qualitatively similar. As is commonly done in the literature, we modify the LMSV and MVJ curves by applying the Sheth *et al.* correction of $\delta_{c0} \to \sqrt{a}\delta_{c0}$. An identical correction is already present in the expressions (4.68) and (4.61) due to the barrier diffusion. We set $\sqrt{a} = 0.89$, which is the value inferred by MR in Ref. [108] using the simulations of Robertson *et al.* [125].

To make the comparison meaningful, we introduce theoretical error bars on the curves. These error bars have no intrinsic statistical meaning – they simply keep track of the absolute magnitude of the terms that are ignored in any given prescription for computing the mass function. As we have discussed at length in section 4.4, these theoretical errors are scale dependent. The estimated error magnitude for each point is the maximum among the terms ignored in the expression. More explicitly, the errors for the linearized LMSV expression (4.62) are estimated as the maximum of $(\epsilon\nu^3)^2$ which comes from the expansion of the exponential, $\epsilon\nu$ which is the order of the largest unequal time terms missing, and $\tilde{\kappa}\nu^{-2}$ which comes from the filter effects. The errors for the LMSV expression (4.63) are similarly estimated as the maximum of $(\epsilon\nu^3)^3$, $\epsilon\nu$ and $\tilde{\kappa}\nu^{-2}$. The largest error for the MVJ expression (4.60) is the maximum of $\epsilon\nu$ (unequal time terms) and $\tilde{\kappa}\nu^{-2}$ (filter effects). Finally, the error for the MR expression (4.61) is the maximum of $(\epsilon\nu^3)^2$ from the expansion of the exponential, $\epsilon\nu^{-3}$ from the largest unequal time terms ignored, and $\tilde{\kappa}^2\nu^{-2}$ and $\tilde{\kappa}\epsilon\nu$ from the filter effects. We include the filter effects and the associated errors only in Fig. 4.6.

From these figures, we can draw some interesting conclusions. First of all, we see that it is important to retain terms which are quadratic in the NG, either with a saddle point method like in MVJ and in our formula, or by expanding the exponential up to second order, like in LMSV. Actually, we argue that it is correct to keep the exponential, otherwise the expansion breaks down when $\epsilon\nu^3$ is of order unity. We notice in passing that the term proportional to ε_2 which comes from the trispectrum, partially cancels with the ε_1^2 term. Secondly, comparing our expression with MVJ's, we can observe that keeping the unequal time terms allows us to sensibly reduce the theoretical errors due to the approximate treatment of NG. In fact, if these terms are missing, they provide the largest theoretical error on large scales. Instead, the largest theoretical error on small scales comes from the approximations involved in dealing with a real space top-hat filter, as is apparent from Fig. 4.6.

To conclude, in this work we have calculated the non-Gaussian halo mass function in the excursion set framework, improving over previous calculations. We started from a path integral formulation of the random walk of the smoothed density field, following Ref. [107]. This allows us to take into account effects due to multi-scale correlations of the smoothed density field ("unequal time" correlations), and due to the real space

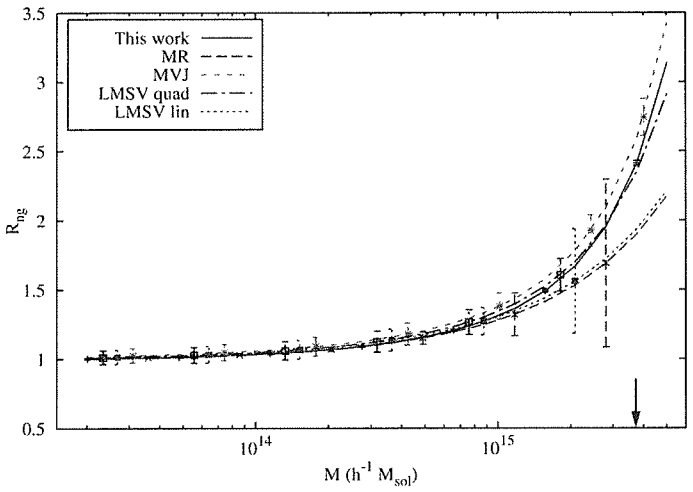


Figure 4.5: Theoretical comparison of the different mass functions at $z = 1$, without the filter effects, i.e. setting $\tilde{\kappa} = 0$. We plot the ratio $R_{\rm ng}$ of the non-Gaussian and Gaussian mass functions, in the local model with $f_{NL} = 100$ and $g_{\rm NL} = 0$. See main text for a discussion of the error bars. The arrow indicates the mass scale where $\varepsilon_1\nu^3/6 = 1$, i.e. where the expansions of LMSV (both linearized and quadratic) and MR break down.

top-hat filter, which generates non-Markovianities in the random walk. We recognize two small parameters in which we perturb: ϵ, defined below Eqn. (4.24), which measures the magnitude of the primordial NG; and $\nu^{-1} = \sigma_R/\delta_c$, which is small on very large scales. In order to do a consistent expansion and to estimate the theoretical errors, one must study the (scale dependent) relation between these two parameters, which we have discussed in Sec. 4.4. We then used saddle point techniques which allowed us to non-perturbatively retain the dependence on $\epsilon\nu^3$, which naturally appears in the calculation and whose magnitude becomes of order unity at high masses and high redshift. Finally, we included effects due to the choice of filter function and due to deviations from spherical collapse, as explained in Sec. 4.5. Our final result is presented in Eqn. (4.68), which we reproduce here:

$$f(\nu,s) = f_{\rm PS}(\nu)\Big(1-\tilde{\kappa}+\mathcal{O}(\tilde{\kappa}^2)\Big)\exp\left[\frac{1}{6}\varepsilon_1\nu^3-\frac{1}{8}\left(\varepsilon_1^2-\frac{\varepsilon_2}{3}\right)\nu^4\right]$$
$$\times\Bigg\{1+\frac{\left(1-2\dot{\tilde{\kappa}}\right)}{1-\tilde{\kappa}}\tilde{\kappa}\nu^{-2}\left(1-2\nu^{-2}\right)-\frac{1}{4}\varepsilon_1\nu\left(\frac{c_1}{1-\tilde{\kappa}}+2-\frac{4}{3}\dot{\varepsilon}_1\right)$$
$$-\frac{1}{4}\varepsilon_1\nu^{-1}\left(\frac{3}{4}c_2-2c_3+\frac{4}{3}\dot{\varepsilon}_1+2c_1(\dot{\varepsilon}_1+\dot{c}_1)\right)$$
$$+\mathcal{O}(\tilde{\kappa}^2\nu^{-2},\tilde{\kappa}\epsilon\nu,\tilde{\kappa}\nu^{-6})+\mathcal{O}(\epsilon^2\nu^2,\epsilon^3\nu^5,\epsilon\nu^{-3})\Bigg\}. \quad (4.70)$$

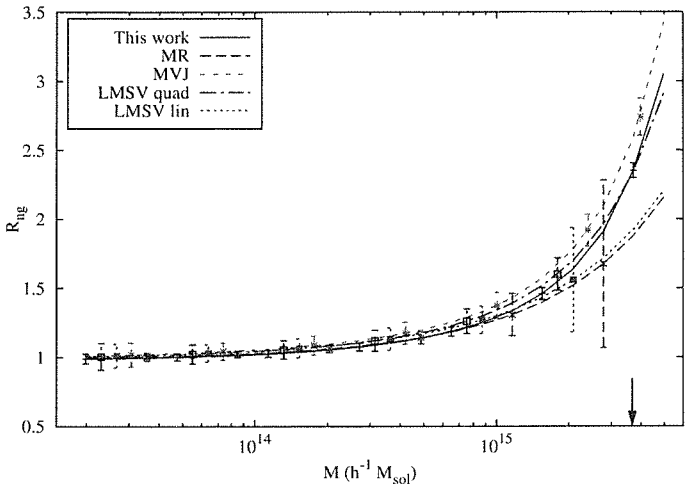


Figure 4.6: Same as Fig. 4.5, but including filter effects. These affect only the error bars for MVJ and LMSV, and they affect both the curve and the error bars for MR and our result. For MR and our result, the Gaussian mass function used to construct the ratio $R_{\rm ng}$, is taken as the non-Gaussian result at $f_{NL} = 0$, and hence includes filter effects.

In Table 4.1 we provide analytical fits for the various parameters that appear in this expression. We also considered other expressions for the mass function found in the literature, which use different expansion methods but do not estimate the theoretical errors. We estimated the theoretical errors for each formula, and we show comparative plots in Fig. 4.5 and Fig. 4.6. In our work we have improved over the calculations of MVJ [104] and LMSV [105] (who ignore unequal time correlations) and of MR [109] (who do not retain the exponential dependence on $\epsilon\nu^3$). We have also demonstrated that the (linearized) result of LMSV can be significantly improved by retaining the quadratic terms of their calculation which are usually ignored in the literature. We find that at large scales and high redshifts, the biggest theoretical errors are introduced by ignoring the exponential dependence on $\epsilon\nu^3$, followed by the neglect of unequal time correlations. The errors on our expression (4.70) are therefore significantly smaller than those of the others. The strength of our approach lies in the combination of path integral methods as laid out by MR [109], and the saddle point approximation as used by MVJ [104].

Very recently an approach like the one of this chapter was used for the study of the non-gaussian mass function with an ellipsoidal collapse barrier (see reference [126]). They partially apply our saddle point techniques, and our expressions can be seen to match in the appropriate limits. They ignore in their final result the contributions coming from unequal time correlators which we saw are of the same order as the leading contributions, these are however trivial to include.

Our work can be continued in several directions. First, a thorough calculation of the effects due to the choice of the filter should be performed, since they lead to significant uncertainties in our final expression. This would include a study of the details of the continuum limit of the path integral near the barrier, and also a study of the statistics of the barrier diffusion process in the presence of filter effects. Second, a comparison with N-body simulations should be performed, in order to quantitatively assess the possibility of constraining NG using our work. Finally, an application to the void statistics along the same lines should be feasible. The problem here is made more interesting by the presence of *two* barriers, as discussed by Sheth & van de Weygaert [127], and since voids probe larger length/mass scales than halos, they constitute a promising future probe of primordial NG [128].

4.A Appendix: Hierarchy of terms in Eqn. (4.33)

Here we argue why the hierarchy of terms ordered by powers of ν^{-2} emerges on expanding the exponentiated derivative operators in Eqn. (4.33). Focusing on terms involving the 3-point correlator, one sees that a generic term in the expansion contains some powers of $(\varepsilon_1 s^{3/2})$, multiplying an n-dimensional integral containing some summations $\sim \sum_{j_1,j_2,\ldots=1}^{n}(1-s_{j_1}/s)^{p_1}(1-s_{j_2}/s)^{p_2}\ldots\partial_{j_1}\partial_{j_2}\ldots$, and also some summations over "free" derivatives $\sim \sum_{k_1,k_2\ldots=1}^{n}\partial_{k_1}\partial_{k_2}\ldots$, all of this acting on $W^{\rm gm}$. More precisely, the structure of the terms is

$$\sim (\varepsilon_1 s^{3/2})^m \sum_{j_1,\ldots,j_{3m}} \int_{-\infty}^{\delta_c} d\delta_1\ldots d\delta_n \left[(1-s_{j_1}/s)\ldots(1-s_{j_m}/s)\right]^p \times \left[(1-s_{j_{m+1}}/s)\ldots(1-s_{j_{2m}}/s)\right]^q \left[(1-s_{j_{2m+1}}/s)\ldots(1-s_{j_{3m}}/s)\right]^r \times \partial_{j_1}\ldots\partial_{j_{3m}} W^{\rm gm}, \quad (4.71)$$

for $m \geq 1$ and non-negative p, q, r such that not all three are zero. The terms we have considered in the text are $(m,p,q,r) = (1,1,0,0)$, $(1,1,1,0)$, $(1,2,0,0)$ and $(2,1,0,0)$. We have already discussed how the "free" derivatives can be pulled out of the integral and converted to ∂_ν. For the "non-free" derivatives, we see that what is important is the *total* number of $(1-s_j/s)$ factors accompanying these derivatives. For example, the $(1,1,1,0)$ term in Eqn. (4.34c) has the same structure as the $(1,2,0,0)$ term in Eqn. (4.34b) – the effect of $\sum_{j,k}(1-s_j/s)(1-s_k/s)\partial_j\partial_k$, up to numerical factors, is identical to that of $\sum_{j,k}(1-s_j/s)^2\partial_j\partial_k$. This is expected to be true also with higher numbers of non-free derivatives.

It is then possible to understand the hierarchy of terms by only considering terms containing $\sum_j(1-s_j/s)^p\partial_j$, and no other non-free derivatives. The basic object to study now becomes

$$\sum_j (1-s_j/s)\int d\delta_1\ldots d\delta_n \partial_j W^{\rm gm},$$

which in the continuum limit can be shown to reduce to the integral

$$g_{(0)}\left(\frac{\nu^2}{2}\right) \equiv \int_0^1 \frac{dy}{y^{3/2}}(1-y)^{1/2}e^{-\nu^2/2y} = \frac{\sqrt{\pi}}{2}\Gamma\left(-\frac{1}{2},\frac{\nu^2}{2}\right) . \tag{4.72}$$

Notice the similarity with the function $h(\nu)$ in Eqn. (4.37), which of course is not accidental given the definitions of these objects. It is now easy to check that increasing the powers of $(1-s_j/s)$ in the summation amounts to increasing the powers of $(1-y)$ in $g_{(0)}$. We are then comparing (with $A=\nu^2/2$) $g_{(0)}(A)$ with $g_{(p)}(A)$ where

$$g_{(p)}(A) \equiv \int_0^1 \frac{dy}{y^{3/2}}(1-y)^{1/2+p}e^{-A/y} . \tag{4.73}$$

Starting with $p=1$ and manipulating the integrals, it is straightforward to establish the recurrence

$$g_{(p+1)}(A) = g_{(p)}(A) - \int_A^\infty d\tilde{A}\, g_{(p)}(\tilde{A}) . \tag{4.74}$$

The argument is now almost complete. We know that for large $A=\nu^2/2$, we have $\Gamma(n,A) = e^{-A}A^{n-1}(1+\mathcal{O}(A^{-1}))$. Hence $g_{(0)}(A) = (\sqrt{\pi}/2)A^{-3/2}e^{-A}(1+\mathcal{O}(A^{-1}))$, and its integral from A to ∞ gives a leading term proportional to $\Gamma(-3/2,A) = e^{-A}A^{-5/2}(1+\mathcal{O}(A^{-1}))$. The pattern is now clear: $g_{(p)}(A) \sim A^{-3/2-p}e^{-A}(1+\mathcal{O}(A^{-1}))$, and since $A=\nu^2/2$, this explains the hierarchy of terms in powers of ν^{-2}, in Eqn. (4.33).

4.B Appendix: The saddle point approximation

In this appendix we discuss the saddle point approximation of the integrals of the type appearing in section 4.3.1, and estimate the error it induces. We will argue that the errors introduced by the saddle point approximation are much smaller than those due to truncating the perturbative series in the small parameters ϵ and ν^{-1}. For an introduction to the saddle point approximation see Ref. [129]. Since we only wish to discuss the saddle point method in this appendix, we will ignore here the complications introduced by the unequal time correlators, i.e. in Eqn. (4.40) we set $\mathcal{P}(\lambda)=1$. We will also work here to first order in $\epsilon\nu$. The extension to a more general case is straightforward and the result is given by (4.51) as described in section 4.3.1. We begin with expression (4.40):

$$f(\nu) = \left(\frac{2}{\pi}\right)^{1/2}\nu\int_{-\infty}^{\infty}\frac{\mathrm{d}\lambda}{\sqrt{2\pi}}e^{g(\lambda)} , \tag{4.75}$$

where $g(\lambda) \equiv i\nu\lambda - \lambda^2/2 + (-i\lambda)^3\varepsilon_1/6 + \mathcal{O}(\epsilon^2\lambda^4)$.

We first find the location of a saddle point λ_* of the function $g(\lambda)$, by perturbatively solving $g'(\lambda_*)=0$ using $\epsilon\nu$ as the small parameter and demanding $g''(\lambda_*)<0$. The first-order solution is

$$\lambda_* = i\nu\big(1-\varepsilon_1\nu/2+\mathcal{O}(\epsilon^2\nu^2)\big) , \tag{4.76}$$

$$g(\lambda_*) = -\frac{\nu^2}{2}(1 - \frac{1}{3}\varepsilon_1\nu + \mathcal{O}(\epsilon^2\nu^2)) \,, \tag{4.77}$$

$$g''(\lambda_*) = -1 - \varepsilon_1\nu + \mathcal{O}(\epsilon^2\nu^2) \,. \tag{4.78}$$

The saddle point approximation consists roughly of performing a Taylor expansion of $g(\lambda)$ to second order around λ_* in the integrand of (4.75) and performing the resulting Gaussian integral. We will carry this out explicitly below. The saddle point prescription will give a good approximation to the integral as long as $g(\lambda)$ attains a global maximum at λ_* (along the contour of integration); this is indeed our case since the integrand in Eqn. (4.75) will be nearly a Gaussian centered at λ_* in the complex plane.

Notice that Im $\lambda_* \neq 0$, requiring a deformation of the contour of integration such that it passes through λ_*. The deformation of the path of integration can be performed by taking a closed contour formed by four pieces: The real axis C_1, the line Im $\lambda =$ Im λ_* which we call here $-C_2$, and the closures of this contour at possitive and negative infinity. The integral in this closed contour must be zero, and since the integral on the closures of the contour at infinity can be assumed to vanish, we have $\int_{C_1} = \int_{C_2}$. Therefore C_2 is the desired deformation of the contour which passes through λ_*[1]. We can then make a series of approximations in the integral (4.75), which we discuss below,

$$\begin{aligned}\int_{-\infty}^{\infty} \frac{d\lambda}{\sqrt{2\pi}} e^{g(\lambda)} &\approx \int_{-\infty}^{\infty} \frac{d\lambda}{\sqrt{2\pi}} e^{g(\lambda_*)+g''(\lambda_*)(\lambda-\lambda_*)^2/2} \\ &= e^{-\frac{1}{2}\nu^2\left(1-\varepsilon_1\nu/3+\mathcal{O}(\epsilon^2\nu^2)\right)} \left(1 + \varepsilon_1\nu + \mathcal{O}(\epsilon^2\nu^2)\right)^{-1/2} \,. \end{aligned} \tag{4.79}$$

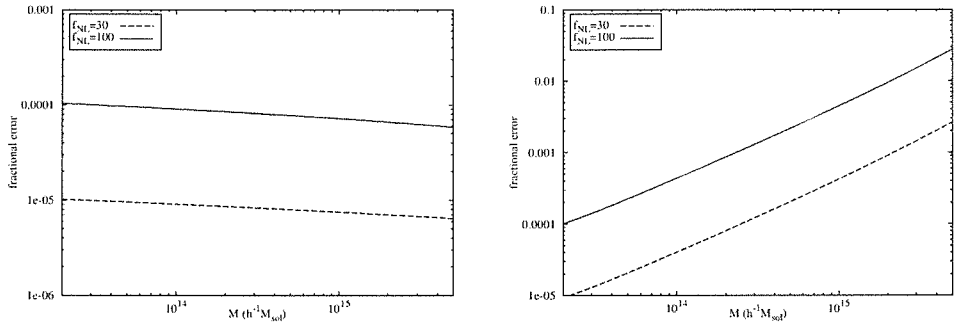
Here the integrations are performed along the deformed contour.

In order to estimate the errors induced by the approximation done in equation (4.79), one can keep higher orders in the Taylor expansion of the function in the exponential:

$$\begin{aligned}\int_{-\infty}^{\infty} \frac{d\lambda}{\sqrt{2\pi}} e^{g(\lambda)} &\approx \int_{-\infty}^{\infty} \frac{d\lambda}{\sqrt{2\pi}} e^{g(\lambda_*)+g''(\lambda_*)(\lambda-\lambda_*)^2/2+g^{(3)}(\lambda_*)(\lambda-\lambda_*)^3/6+g^{(4)}(\lambda_*)(\lambda-\lambda_*)^4/24+\ldots} \\ &\approx e^{g(\lambda_*)} \left(-g''(\lambda_*)\right)^{-1/2} \\ &\quad + \int_{-\infty}^{\infty} \frac{\mathrm{d}z}{\sqrt{2\pi}} \left\{ \frac{1}{6} g^{(3)}(\lambda_*) z^3 + \frac{1}{72} \left[g^{(3)}(\lambda_*)\right]^2 z^6 + \frac{1}{24} g^{(4)}(\lambda_*) z^4 + \ldots \right\} \\ &\qquad \times e^{g(\lambda_*)+g''(\lambda_*)z^2/2} \\ &= e^{g(\lambda_*)} \left(-g''(\lambda_*)\right)^{-1/2} (1 + \mathcal{O}(\epsilon^2)) \,. \end{aligned} \tag{4.80}$$

Here we used the fact that $g^{(3)}(\lambda_*) = \mathcal{O}(\epsilon)$ and $g^{(4)}(\lambda_*) = \mathcal{O}(\epsilon^2)$. The integrals in the second equality of this derivation can be computed analytically, which allows one to go

[1] Technically, one should also require that Im $g(\lambda)$ be nearly constant along the deformed contour for the saddle point approximation to work. In our case one can show that Im g will be suppressed by ϵ. This and all errrors induced by the saddle point are accounted for in equation (4.80).



(a) (b)

Figure 4.7: Panel (a) : Fractional difference between the saddle point approximation on the r.h.s. of Eqn. (4.81) and the numerical integration of the l.h.s. of the same equation. Panel (b) : Total error induced on the result of the toy model (4.81) by both the saddle point approximation and the perturbative expansion to leading order in $\epsilon\nu$. We plot the fractional difference between the numerical integration of the l.h.s. of Eqn. (4.81) and the approximation (4.79). Both panels show the results for a local NG with two values of f_{NL}.

to arbitrary accuracy with the saddle point technique. Notice that the results of these integrations are of higher order than the terms we retain. In the main text, where the integral contains also a polynomial $\mathcal{P}(\lambda)$ one can again compute the errors via similar Taylor expansions. These errors can be shown to be of order $\mathcal{O}(\epsilon^2)$, comparable to other terms which we ignore.

One can also estimate the errors introduced by our opproximations by using the following toy model in which everything is computable: Take the 3-point cumulant ε_1 to be different from zero and all higher order cumulants ε_n for $n \geq 2$ to be zero[1]. For such a model the integral is

$$\int_{-\infty}^{\infty} \mathrm{d}\lambda \; e^{i\nu\lambda - \lambda^2/2 + (-i\lambda)^3\varepsilon_1/6} \approx \left(\frac{2\pi}{\sqrt{1+2\varepsilon_1\nu}}\right)^{1/2} \exp\left(\frac{1-\sqrt{1+2\varepsilon_1\nu}+\varepsilon_1\nu\left(3-2\sqrt{1+2\varepsilon_1\nu}\right)}{3\varepsilon_1^2}\right). \quad (4.81)$$

vIn the r.h.s of this equation we have used the saddle point approximation but have made no expansion in $\epsilon\nu$. By comparing the numerical integration of the l.h.s. with the expression on the r.h.s. (panel (a) of Fig. 4.7), one can see that the errors introduced by the saddle point approximation are indeed of order ϵ^2 as indicated by (4.80). On the

[1]This toy model is inconsistent because if the third cumulant is different from zero, then all higher cumulants must also be different from zero. We use it here only to estimate how good the saddle point prescription is in approximating an integral, and compare it with errors induced by a perturbative expansion in $\epsilon\nu$.

other hand, one can use the numerical integration of the left hand side of this equation and compare it with the approximation (4.79) (panel (b) of Fig. 4.7), to see that the biggest error is of order $\epsilon^2\nu^2$ induced by the fact that we perform a perturbative expansion in $\epsilon\nu$. Notice that here we considered only the leading order in $\epsilon\nu$ and ignored unequal time correlators, while in the main text we present a result which is more precise (to next to leading order in $\epsilon\nu$) and complete (using the excursion set formalism rigorously).

5

Discussion.

We studied nonlinear cosmological perturbations during the post-inflationary evolution, using the equivalence between a perfect barotropic fluid and a derivatively coupled scalar field with Lagrangian $[-(\partial\phi)^2]^{(1+w)/2w}$. Since this Lagrangian is just a special case of k-inflation, this approach is analogous to the one employed in the study of non-Gaussianities from inflation. We used this method to derive the second order metric during matter dominance in the comoving gauge directly as function of the primordial inflationary perturbation ζ. Going to Poisson gauge, we recover the metric previously derived in the literature. We briefly discussed an interpretation of the apparent non-conservation of ζ due to the coupling of subhorizon modes. This calculation is a first step in the estimation of the gravitational contribution to the non-gaussianity observed in the CMB.

We calculated the CMB anisotropy bispectrum on large angular scales in the absence of primordial non-Gaussianities, assuming exact matter dominance and extending at second order the classic Sachs-Wolfe result $\delta T/T = \Phi/3$. The calculation was done in Poisson gauge. Besides intrinsic contributions calculated at last scattering, one must consider integrated effects. These are associated to lensing, and to the time dependence of the potentials (Rees-Sciama) and of the vector and tensor components of the metric generated at second order. The bispectrum was explicitly computed in the flat-sky approximation. It scales as l^{-4} in the scale invariant limit and the shape dependence of its various contributions is represented in 3d plots. Although all the contributions to the bispectrum are parametrically of the same order, the full bispectrum is dominated by lensing. In the squeezed limit it corresponds to $f_{\rm NL}^{\rm local} = -1/6 - \cos(2\theta)$, where θ is the angle between the short and the long modes; the angle dependent contribution comes from lensing. In the equilateral limit it corresponds to $f_{\rm NL}^{\rm equil} \simeq 3.13$.

The abundance of collapsed objects in the universe, or halo mass function, is an important theoretical tool in studying the effects of primordially generated non-Gaussianities on the large scale structure. The non-Gaussian mass function has been calculated by several authors in different ways, typically by exploiting the smallness of certain parameters which naturally appear in the calculation to set up a perturbative

expansion. We improved upon the existing results for the mass function by combining path integral methods and saddle point techniques (which have been separately applied in previous approaches). Additionally, we carefully accounted for the various scale dependent combinations of small parameters which appear. Since the expansion was done on several small parameters which depend on the mass, a careful treatment is not only comfortable for the calculation but also strictly necessary. Some of these combinations in fact become of order unity for large mass scales and at high redshifts, and must therefore be treated non-perturbatively. Our approach allows us to do this, and to also account for multi-scale density correlations which appear in the calculation. We thus derived an accurate expression for the mass function which is based on approximations that are valid over a larger range of mass scales and redshifts than those of other authors. By tracking the terms ignored in the analysis, we estimated theoretical errors for our result and also for the results of others. We also discussed the complications introduced by the choice of smoothing filter function, which we take to be a top-hat in real space, and which leads to the dominant errors in our expression. Finally, we presented a detailed comparison between the various expressions for the mass functions, exploring the accuracy and range of validity of each; this is also one of the main results of our computation.

References

[1] Lotfi Boubekeur, Paolo Creminelli, Jorge Norena, and Filippo Vernizzi. Action approach to cosmological perturbations: the 2nd order metric in matter dominance. *JCAP*, 0808:028, 2008, 0806.1016. 1, 33, 54, 57

[2] Guido D'Amico, Marcello Musso, Jorge Norena, and Aseem Paranjape. An Improved Calculation of the Non-Gaussian Halo Mass Function. 2010, 1005.1203. 1

[3] Lotfi Boubekeur, Paolo Creminelli, Guido D'Amico, Jorge Norena, and Filippo Vernizzi. Sachs-Wolfe at second order: the CMB bispectrum on large angular scales. *JCAP*, 0908:029, 2009, 0906.0980. 1, 29

[4] Juan Martin Maldacena. Non-Gaussian features of primordial fluctuations in single field inflationary models. *JHEP*, 05:013, 2003, astro-ph/0210603. 2, 8, 10, 21, 22, 33, 38, 39, 40, 45, 57, 82, 83, 84

[5] Licia Verde. Non-Gaussianity from Large-Scale Structure Surveys. 2010, 1001.5217. 2, 30, 92

[6] E. Komatsu et al. Non-Gaussianity as a Probe of the Physics of the Primordial Universe and the Astrophysics of the Low Redshift Universe. 2009, 0902.4759. 2, 26, 29

[7] Daniel Baumann. TASI Lectures on Inflation. 2009, 0907.5424. 2, 8, 16

[8] Adam G. Riess et al. Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. *Astron. J.*, 116:1009–1038, 1998, astro-ph/9805201. 3

[9] S. Perlmutter et al. Measurements of Omega and Lambda from 42 High-Redshift Supernovae. *Astrophys. J.*, 517:565–586, 1999, astro-ph/9812133. 3

[10] Steven Weinberg. The cosmological constant problem. *Rev. Mod. Phys.*, 61:1–23, 1989. 3

[11] E. Komatsu et al. Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation. 2010, 1001.4538. 4, 17, 29, 53, 92, 94

[12] B.D Fields and S. Sakar. Review of particle physics: Big Bang Nucleosynthesis. *Phys. Lett.*, B667:202, 2008. 4

[13] Alan H. Guth. The Inflationary Universe: A Possible Solution to the Horizon and Flatness Problems. *Phys. Rev.*, D23:347–356, 1981. 6

[14] Bruce A. Bassett, Shinji Tsujikawa, and David Wands. Inflation dynamics and reheating. *Rev. Mod. Phys.*, 78:537–589, 2006, astro-ph/0507632. 7

[15] James M. Bardeen. Gauge Invariant Cosmological Perturbations. *Phys. Rev.*, D22:1882–1905, 1980. 9

[16] Hideo Kodama and Misao Sasaki. Cosmological Perturbation Theory. *Prog. Theor. Phys. Suppl.*, 78:1–166, 1984. 9

[17] Richard L. Arnowitt, Stanley Deser, and Charles W. Misner. The dynamics of general relativity. 1962, gr-qc/0405109. 10, 38

[18] David H. Lyth, Karim A. Malik, and Misao Sasaki. A general proof of the conservation of the curvature perturbation. *JCAP*, 0505:004, 2005, astro-ph/0411220. 12

[19] Viatcheslav Mukhanov and Sergei Winitzki. *Introduction to quantum effects in gravity.* Cambridge Univ. Pr., Cambridge, UK, 2007. 13

[20] D. Larson et al. Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Power Spectra and WMAP-Derived Parameters. 2010, 1001.4635. 18, 19

[21] Antony Lewis, Anthony Challinor, and Anthony Lasenby. Efficient Computation of CMB anisotropies in closed FRW models. *Astrophys. J.*, 538:473–476, 2000, astro-ph/9911177. 19, 95

[22] Uros Seljak and Matias Zaldarriaga. A Line of Sight Approach to Cosmic Microwave Background Anisotropies. *Astrophys. J.*, 469:437–444, 1996, astro-ph/9603033. 19, 95

[23] Scott Dodelson. *Modern Cosmology.* Elsevier, Academic Press, Oxfort, UK, 2003. 19, 20

[24] Chung-Pei Ma and Edmund Bertschinger. Cosmological perturbation theory in the synchronous and conformal Newtonian gauges. *Astrophys. J.*, 455:7–25, 1995, astro-ph/9506072. 19

[25] Uros Seljak. A Two fluid approximation for calculating the cosmic microwave background anisotropies. *Astrophys. J.*, 435:L87–L90, 1994, astro-ph/9406050. 19

[26] David N. Spergel and Matias Zaldarriaga. CMB polarization as a direct test of inflation. *Phys. Rev. Lett.*, 79:2180–2183, 1997, astro-ph/9705182. 20

[27] Viviana Acquaviva, Nicola Bartolo, Sabino Matarrese, and Antonio Riotto. Second-order cosmological perturbations from inflation. *Nucl. Phys.*, B667:119–148, 2003, astro-ph/0209156. 21, 57

[28] Paolo Creminelli and Matias Zaldarriaga. Single field consistency relation for the 3-point function. *JCAP*, 0410:006, 2004, astro-ph/0407059. 22, 82, 83, 84

[29] Clifford Cheung, A. Liam Fitzpatrick, Jared Kaplan, and Leonardo Senatore. On the consistency relation of the 3-point function in single field inflation. *JCAP*, 0802:021, 2008, 0709.0295. 22, 82, 83, 84

[30] Paolo Creminelli. On non-gaussianities in single-field inflation. *JCAP*, 0310:003, 2003, astro-ph/0306122. 23, 24, 69, 96

[31] N. Bartolo, E. Komatsu, Sabino Matarrese, and A. Riotto. Non-Gaussianity from inflation: Theory and observations. *Phys. Rept.*, 402:103–266, 2004, astro-ph/0406398. 23, 26

[32] David H. Lyth, Carlo Ungarelli, and David Wands. The primordial density perturbation in the curvaton scenario. *Phys. Rev.*, D67:023503, 2003, astro-ph/0208055. 24, 96

[33] Eiichiro Komatsu. The Pursuit of Non-Gaussian Fluctuations in the Cosmic Microwave Background. 2002, astro-ph/0206039. 26

[34] Daniel Babich, Paolo Creminelli, and Matias Zaldarriaga. The shape of non-Gaussianities. *JCAP*, 0408:009, 2004, astro-ph/0405356. 28, 55, 63, 64, 65, 69, 79, 96

[35] Clifford Cheung, Paolo Creminelli, A. Liam Fitzpatrick, Jared Kaplan, and Leonardo Senatore. The Effective Field Theory of Inflation. *JHEP*, 03:014, 2008, 0709.0293. 29

[36] Leonardo Senatore, Kendrick M. Smith, and Matias Zaldarriaga. Non-Gaussianities in Single Field Inflation and their Optimal Limits from the WMAP 5-year Data. *JCAP*, 1001:028, 2010, 0905.3746. 29, 80

[37] Duncan Hanson, Kendrick M. Smith, Anthony Challinor, and Michele Liguori. CMB lensing and primordial non-Gaussianity. *Phys. Rev.*, D80:083004, 2009, 0905.4732. 29, 85

[38] Leonardo Senatore, Svetlin Tassev, and Matias Zaldarriaga. Cosmological Perturbations at Second Order and Recombination Perturbed. *JCAP*, 0908:031, 2009, 0812.3652. 29, 53

[39] Leonardo Senatore, Svetlin Tassev, and Matias Zaldarriaga. Non-Gaussianities from Perturbing Recombination. *JCAP*, 0909:038, 2009, 0812.3658. 29, 53

[40] Cyril Pitrou, Jean-Philippe Uzan, and Francis Bernardeau. Cosmic microwave background bispectrum on small angular scales. *Phys. Rev.*, D78:063526, 2008, 0807.0341. 29, 53

[41] Cyril Pitrou, Jean-Philippe Uzan, and Francis Bernardeau. The cosmic microwave background bispectrum from the non- linear evolution of the cosmological perturbations. *JCAP*, 1007:003, 2010, 1003.0481. 29, 53, 81

[42] Daisuke Nitta, Eiichiro Komatsu, Nicola Bartolo, Sabino Matarrese, and Antonio Riotto. CMB anisotropies at second order III: bispectrum from products of the first-order perturbations. *JCAP*, 0905:014, 2009, 0903.0894. 29, 53, 81

[43] Vincent Desjacques and Uros Seljak. Primordial non-Gaussianity in the large scale structure of the Universe. 2010, 1006.4763. 30

[44] Anze Slosar, Christopher Hirata, Uros Seljak, Shirley Ho, and Nikhil Padmanabhan. Constraints on local primordial non-Gaussianity from large scale structure. *JCAP*, 0808:031, 2008, 0805.3580. 30, 31, 32, 53, 91

[45] Emiliano Sefusatti and Eiichiro Komatsu. The bispectrum of galaxies from high-redshift galaxy surveys: Primordial non-Gaussianity and non-linear galaxy bias. *Phys. Rev.*, D76:083004, 2007, 0705.0343. 30, 31

[46] Emiliano Sefusatti, Martin Crocce, and Vincent Desjacques. The Matter Bispectrum in N-body Simulations with non- Gaussian Initial Conditions. 2010, 1003.0007. 30

[47] Carmelita Carbone, Licia Verde, and Sabino Matarrese. Non-Gaussian halo bias and future galaxy surveys. *Astrophys. J.*, 684:L1–L4, 2008, 0806.1950. 32

[48] R. K. Sachs and A. M. Wolfe. Perturbations of a cosmological model and angular variations of the microwave background. *Astrophys. J.*, 147:73–90, 1967. 33, 54, 56

[49] Sabino Matarrese, Silvia Mollerach, and Marco Bruni. Second-order perturbations of the Einstein-de Sitter universe. *Phys. Rev.*, D58:043504, 1998, astro-ph/9707278. 34, 46, 47, 48, 54, 56, 57

[50] Nicola Bartolo, Sabino Matarrese, and Antonio Riotto. The Full Second-Order Radiation Transfer Function for Large-Scale CMB Anisotropies. *JCAP*, 0605:010, 2006, astro-ph/0512481. 34, 46, 48, 54, 55, 57, 85

[51] David Seery and James E. Lidsey. Primordial non-gaussianities in single field inflation. *JCAP*, 0506:003, 2005, astro-ph/0503692. 34, 37, 38, 41

[52] Xingang Chen, Min-xin Huang, Shamit Kachru, and Gary Shiu. Observational signatures and non-Gaussianities of general single field inflation. *JCAP*, 0701:002, 2007, hep-th/0605045. 34, 37, 38, 41, 42

[53] D. S. Salopek and J. R. Bond. Nonlinear evolution of long wavelength metric fluctuations in inflationary models. *Phys. Rev.*, D42:3936–3962, 1990. 34, 44

[54] A. H. Taub. General Relativistic Variational Principle for Perfect Fluids. *Phys. Rev.*, 94:1468–1470, 1954. 35, 36

[55] Bernard F. Schutz. Perfect Fluids in General Relativity: Velocity Potentials and a Variational Principle. *Phys. Rev.*, D2:2762–2773, 1970. 35, 36

[56] S. Dubovsky, T. Gregoire, A. Nicolis, and R. Rattazzi. Null energy condition and superluminal propagation. *JHEP*, 03:025, 2006, hep-th/0512260. 35, 36

[57] Steven Weinberg. *Gravitation and Cosmology*. John Wiley and Sons, 1972. 36

[58] S. Matarrese. On the classical and quantum irrotation motions of a relativistic perfect fluid. 1. Classical theory. *Proc. Roy. Soc. Lond.*, A401:53–66, 1985. 36

[59] Jaume Garriga and Viatcheslav F. Mukhanov. Perturbations in k-inflation. *Phys. Lett.*, B458:219–225, 1999, hep-th/9904176. 36, 37

[60] C. Armendariz-Picon, T. Damour, and Viatcheslav F. Mukhanov. k-Inflation. *Phys. Lett.*, B458:209–218, 1999, hep-th/9904075. 37

[61] C. Armendariz-Picon, Viatcheslav F. Mukhanov, and Paul J. Steinhardt. Essentials of k-essence. *Phys. Rev.*, D63:103510, 2001, astro-ph/0006373. 37

[62] Frederico Arroja and Kazuya Koyama. Non-gaussianity from the trispectrum in general single field inflation. *Phys. Rev.*, D77:083517, 2008, 0802.1167. 37

[63] Daniel Baumann, Alberto Nicolis, Leonardo Senatore, and Matias Zaldarriaga. Cosmological Non-Linearities as an Effective Fluid. 2010, 1004.2488. 44, 49, 50, 51

[64] Kishore N. Ananda, Chris Clarkson, and David Wands. The cosmological gravitational wave background from primordial density perturbations. *Phys. Rev.*, D75:123518, 2007, gr-qc/0612013. 52

[65] Daniel Baumann, Paul J. Steinhardt, Keitaro Takahashi, and Kiyotomo Ichiki. Gravitational Wave Spectrum Induced by Primordial Scalar Perturbations. *Phys. Rev.*, D76:084019, 2007, hep-th/0703290. 52

[66] Kendrick M. Smith, Leonardo Senatore, and Matias Zaldarriaga. Optimal limits on f_{NL}^{local} from WMAP 5-year data. *JCAP*, 0909:006, 2009, 0901.2572. 53

[67] Uros Seljak and Matias Zaldarriaga. Direct Signature of Evolving Gravitational Potential from Cosmic Microwave Background. *Phys. Rev.*, D60:043504, 1999, astro-ph/9811123. 53, 55, 76

[68] David M. Goldberg and David N. Spergel. Microwave background bispectrum. 2. A probe of the low redshift universe. *Phys. Rev.*, D59:103002, 1999, astro-ph/9811251. 53, 55, 76

[69] Rishi Khatri and Benjamin D. Wandelt. Crinkles in the last scattering surface: Non-Gaussianity from inhomogeneous recombination. *Phys. Rev.*, D79:023501, 2009, 0810.4370. 53

[70] Nicola Bartolo and Antonio Riotto. On the non-Gaussianity from Recombination. *JCAP*, 0903:017, 2009, 0811.4584. 53

[71] Nicola Bartolo, Sabino Matarrese, and Antonio Riotto. CMB Anisotropies at Second-Order II: Analytical Approach. *JCAP*, 0701:019, 2007, astro-ph/0610110. 53

[72] Cyril Pitrou. The radiative transfer at second order: a full treatment of the Boltzmann equation with polarization. *Class. Quant. Grav.*, 26:065006, 2009, 0809.3036. 53, 81

[73] Nicola Bartolo, Sabino Matarrese, and Antonio Riotto. Gauge-invariant temperature anisotropies and primordial non-Gaussianity. *Phys. Rev. Lett.*, 93:231301, 2004, astro-ph/0407505. 54, 61, 67

[74] Nicola Bartolo, Sabino Matarrese, and Antonio Riotto. Non-Gaussianity of Large-Scale Cosmic Microwave Background Anisotropies beyond Perturbation Theory. *JCAP*, 0508:010, 2005, astro-ph/0506410. 54, 59, 62

[75] Silvia Mollerach, Alejandro Gangui, Francesco Lucchin, and Sabino Matarrese. Contribution to the three point function of the cosmic microwave background from the Rees-Sciama effect. *Astrophys. J.*, 453:1, 1995, astro-ph/9503115. 55, 69

[76] Dipak Munshi, Tarun Souradeep, and Alexei A. Starobinsky. Skewness of cosmic microwave background temperature fluctuations due to nonlinear gravitational instability. *Astrophys. J.*, 454:552, 1995, astro-ph/9501100. 55, 69

[77] Paolo Creminelli and Matias Zaldarriaga. CMB 3-point functions generated by non-linearities at recombination. *Phys. Rev.*, D70:083532, 2004, astro-ph/0405428. 55, 62, 85

[78] Ted Pyne and Sean M. Carroll. Higher-Order Gravitational Perturbations of the Cosmic Microwave Background. *Phys. Rev.*, D53:2920–2929, 1996, astro-ph/9510041. 56, 61

[79] Silvia Mollerach and Sabino Matarrese. Cosmic microwave background anisotropies from second order gravitational perturbations. *Phys. Rev.*, D56:4494–4502, 1997, astro-ph/9702234. 56, 61

[80] Charles W. Misner, Kip S. Thorne, and John A. Wheeler. *Gravitation*. W. H. Freeman, 1973. 56

[81] David Langlois and Filippo Vernizzi. Nonlinear perturbations of cosmological scalar fields. *JCAP*, 0702:017, 2007, astro-ph/0610064. 59

[82] Wayne Hu and Asantha Cooray. Gravitational time delay effects on cosmic microwave background anisotropies. *Phys. Rev.*, D63:023504, 2001, astro-ph/0008001. 61

[83] Uros Seljak. Gravitational lensing effect on cosmic microwave background anisotropies: A Power spectrum approach. *Astrophys. J.*, 463:1, 1996, astro-ph/9505109. 61, 85

[84] Wayne Hu. Weak lensing of the CMB: A harmonic approach. *Phys. Rev.*, D62:043007, 2000, astro-ph/0001303. 63, 78, 85, 86

[85] E. Komatsu et al. First Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Tests of Gaussianity. *Astrophys. J. Suppl.*, 148:119–134, 2003, astro-ph/0302223. 64

[86] M. Liguori, F. K. Hansen, E. Komatsu, S. Matarrese, and A. Riotto. Testing Primordial Non-Gaussianity in CMB Anisotropies. *Phys. Rev.*, D73:043505, 2006, astro-ph/0509098. 67

[87] Phillip James Edwin Peebles. *The Large-Scale Structure of the Universe*. Princeton University Press, 1980. 69

[88] M. J. Rees and D. W. Sciama. Large scale Density Inhomogeneities in the Universe. *Nature*, 217:511–516, 1968. 69

[89] David N. Spergel and David M. Goldberg. Microwave background bispectrum. 1. Basic formalism. *Phys. Rev.*, D59:103001, 1999, astro-ph/9811252. 69

[90] Antony Lewis and Anthony Challinor. Weak Gravitational Lensing of the CMB. *Phys. Rept.*, 429:1–65, 2006, astro-ph/0601594. 75, 77, 78

[91] Matias Zaldarriaga. Lensing of the CMB: Non-Gaussian aspects. *Phys. Rev.*, D62:063510, 2000, astro-ph/9910498. 78

[92] Pieter Daniel Meerburg, Jan Pieter van der Schaar, and Pier Stefano Corasaniti. Signatures of Initial State Modifications on Bispectrum Statistics. *JCAP*, 0905:018, 2009, 0901.4044. 80

[93] David Seery, Martin S. Sloth, and Filippo Vernizzi. Inflationary trispectrum from graviton exchange. *JCAP*, 0903:018, 2009, 0811.3934. 84

[94] Neal Dalal, Olivier Dore, Dragan Huterer, and Alexander Shirokov. The imprints of primordial non-gaussianities on large- scale structure: scale dependent bias and abundance of virialized objects. *Phys. Rev.*, D77:123514, 2008, 0710.4560. 91

[95] Sabino Matarrese and Licia Verde. The effect of primordial non-Gaussianity on halo bias. *Astrophys. J.*, 677:L77, 2008, 0801.4826. 91

[96] B. Sartoris et al. The potential of X-ray cluster surveys to constrain primordial non-Gaussianity. 2010, 1003.0841. 92

[97] Carlos Cunha, Dragan Huterer, and Olivier Dore. Primordial non-Gaussianity from the covariance of galaxy cluster counts. *Phys. Rev.*, D82:023004, 2010, 1003.2416. 92

[98] Emiliano Sefusatti. 1-loop Perturbative Corrections to the Matter and Galaxy Bispectrum with non-Gaussian Initial Conditions. *Phys. Rev.*, D80:123002, 2009, 0905.0717. 92

[99] Raul Jimenez and Licia Verde. Implications for Primordial Non-Gaussianity (f_{NL}) from weak lensing masses of high-z galaxy clusters. *Phys. Rev.*, D80:127302, 2009, 0909.0403. 92

[100] Vincent Desjacques and Uros Seljak. Primordial non-Gaussianity from the large scale structure. *Class. Quant. Grav.*, 27:124011, 2010, 1003.5020. 92

[101] James E. Gunn and J. Richard Gott, III. On the infall of matter into cluster of galaxies and some effects on their evolution. *Astrophys. J.*, 176:1–19, 1972. 92

[102] William H. Press and Paul Schechter. Formation of galaxies and clusters of galaxies by selfsimilar gravitational condensation. *Astrophys. J.*, 187:425–438, 1974. 92

[103] J. R. Bond, S. Cole, G. Efstathiou, and Nick Kaiser. Excursion set mass functions for hierarchical Gaussian fluctuations. *Astrophys. J.*, 379:440, 1991. 92, 97

[104] Sabino Matarrese, Licia Verde, and Raul Jimenez. The abundance of high-redshift objects as a probe of non- Gaussian initial conditions. *Astrophys. J.*, 541:10, 2000, astro-ph/0001366. 92, 93, 111, 120

[105] Marilena LoVerde, Amber Miller, Sarah Shandera, and Licia Verde. Effects of Scale-Dependent Non-Gaussianity on Cosmological Structures. *JCAP*, 0804:014, 2008, 0711.4126. 92, 93, 111, 112, 120

[106] Ravi K. Sheth and Giuseppe Tormen. Large scale bias and the peak background split. *Mon. Not. Roy. Astron. Soc.*, 308:119, 1999, astro-ph/9901122. 93, 112

[107] Michele Maggiore and Antonio Riotto. The Halo Mass Function from the Excursion Set Method. I. First principle derivation for the non-markovian case of gaussian fluctuations and generic filter. *Astrophys. J.*, 711:907–927, 2010, 0903.1249. 93, 94, 97, 98, 112, 114, 118

[108] Michele Maggiore and Antonio Riotto. The halo mass function from the excursion set method. II. The diffusing barrier. *Astrophys. J.*, 717:515–525, 2010, 0903.1250. 93, 94, 100, 113, 115, 118

[109] Michele Maggiore and Antonio Riotto. The halo mass function from the excursion set method. III. First principle derivation for non-Gaussian theories. *Astrophys. J.*, 717:526–541, 2010, 0903.1251. 93, 97, 104, 111, 112, 114, 115, 120

[110] P. Valageas. Mass function and bias of dark matter halos for non- Gaussian initial conditions. 2009, 0906.1042. 93

[111] James M. Bardeen, J. R. Bond, Nick Kaiser, and A. S. Szalay. The Statistics of Peaks of Gaussian Random Fields. *Astrophys. J.*, 304:15–61, 1986. 93, 95

[112] Ravi K. Sheth, H. J. Mo, and Giuseppe Tormen. Ellipsoidal collapse and an improved model for the number and spatial distribution of dark matter haloes. *Mon. Not. Roy. Astron. Soc.*, 323:1, 2001, astro-ph/9907024. 93, 114

[113] Tsz Yan Lam and Ravi K. Sheth. Halo abundances in the f_{nl} model. 2009, 0905.1702. 93, 112

[114] Naoshi Sugiyama. Cosmic background anistropies in CDM cosmology. *Astrophys. J. Suppl.*, 100:281, 1995, astro-ph/9412025. 95

[115] N. Bartolo, S. Matarrese, and A. Riotto. On non-Gaussianity in the curvaton scenario. *Phys. Rev.*, D69:043503, 2004, hep-ph/0309033. 96

[116] Gia Dvali, Andrei Gruzinov, and Matias Zaldarriaga. A new mechanism for generating density perturbations from inflation. *Phys. Rev.*, D69:023505, 2004, astro-ph/0303591. 96

[117] Mohsen Alishahiha, Eva Silverstein, and David Tong. DBI in the sky. *Phys. Rev.*, D70:123505, 2004, hep-th/0404084. 96

[118] Nima Arkani-Hamed, Paolo Creminelli, Shinji Mukohyama, and Matias Zaldarriaga. Ghost Inflation. *JCAP*, 0404:001, 2004, hep-th/0312100. 96

[119] Paolo Creminelli, Alberto Nicolis, Leonardo Senatore, Max Tegmark, and Matias Zaldarriaga. Limits on non-Gaussianities from WMAP data. *JCAP*, 0605:004, 2006, astro-ph/0509029. 96

[120] Jean Zinn-Justin. Quantum field theory and critical phenomena. *Int. Ser. Monogr. Phys.*, 113:1–1054, 2002. 98

[121] Subrahmanyan Chandrasekhar. Stochastic problems in physics and astronomy. *Rev. Mod. Phys.*, 15:1–89, 1943. 98

[122] Tommaso Giannantonio and Cristiano Porciani. Structure formation from non-Gaussian initial conditions: multivariate biasing, statistics, and comparison with N- body simulations. *Phys. Rev.*, D81:063530, 2010, 0911.0017. 112

[123] C. Carbone et al. The properties of the dark matter halo distribution in non-Gaussian scenarios. *Nucl. Phys. Proc. Suppl.*, 194:22–27, 2009. 112

[124] F. Bernardeau, S. Colombi, E. Gaztanaga, and R. Scoccimarro. Large-scale structure of the universe and cosmological perturbation theory. *Phys. Rept.*, 367:1–248, 2002, astro-ph/0112551. 112

[125] Brant Robertson, Andrey Kravtsov, Jeremy Tinker, and Andrew Zentner. Collapse Barriers and Halo Abundance: Testing the Excursion Set Ansatz. *Astrophys. J.*, 696:636–652, 2009, 0812.3148. 114, 115, 118

[126] Andrea De Simone, Michele Maggiore, and Antonio Riotto. Excursion Set Theory for generic moving barriers and non- Gaussian initial conditions. 2010, 1007.1903. 120

[127] Ravi K. Sheth and Rien van de Weygaert. A hierarchy of voids: Much ado about nothing. *Mon. Not. Roy. Astron. Soc.*, 350:517, 2004, astro-ph/0311260. 121

[128] Marc Kamionkowski, Licia Verde, and Raul Jimenez. The Void Abundance with Non-Gaussian Primordial Perturbations. *JCAP*, 0901:010, 2009, 0809.0506. 121

[129] A. Erdélyi. *Asymptotic Expansions.* Dover, New York, NY, 1956. 122