Dark Energy and Non-Gaussianity Through the Large Scale Structure

Guido D'Amico

Astroparticle Physics Curriculum

Advisor: Prof. Paolo Creminelli

A thesis submitted in partial satisfaction of the requirements for the degree of

Philosophiae Doctor (PhD)

2010

Contents

CONTENTS

Chapter 1

Introduction

Can we actually know the Universe?
My God, it's hard enough finding your way
around in Chinatown.

Woody Allen

In the last twenty years, cosmology has become a precision science. Since the COBE era, a wealth of data, along with advances in theoretical interpretation and numerical analysis, provided us with a convincing model of the content and evolution of our Universe. It is now generally accepted that an early period of inflation produced a nearly homogeneous flat Universe, with small inhomogeneities (of quantum origin!) which can be seen as temperature anisotropies in the cosmic microwave background (CMB), and which grow by gravitational instability to form the large scale structure (LSS) of the Universe.

Measurements of the CMB anisotropies, which provide a spectacular 2-dimensional snapshot of the Universe at the time of recombination, by the WMAP satellite (for the latest release see [1]) and other missions ([2, 3]) have been the dominant source of information in the past decade. From the theoretical side, we have a thorough understanding of the physics of CMB [4, 5, 6, 7, 8, 9] and fast numerical codes for anisotropy spectrum calculation [10, 11], since perturbations at decoupling are very small and a perturbative treatment works well. Precision measurements of the CMB will go on in the next years, with the Planck mission ([12]) and with CMBPOL [13].

These missions will probably end the epoch of CMB data mining and analysis. However, a lot of experiments aiming at LSS observations have been planned for the near (and not so near) future. In the last few years, in fact, large scale structure has been considered more and more as

an important and complementary probe to CMB measurements, and it will likely become the main subject of observational cosmology. For instance, measurements of primordial non-Gaussianity using the scale-dependence of bias [14, 15, 16] are already competitive with the latest WMAP results [1]. In addition, large scale structures are a unique probe of the late acceleration of the Universe, and therefore the only window on the nature of dark energy and on theories of modified gravity.

Unfortunately, LSS observations are not ideal both experimentally, since what is observed is often just the luminous baryon fraction which constitute only one sixth of the total matter (exceptions are lensing and velocity measurements), and theoretically, since structures have undergone non-linear gravitational and complex hydrodynamical evolution. However, if one focuses on very large structures, such as galaxy clusters, some theoretical predictions can be done. In fact, clusters are mainly dominated by gravitational physics, and simplified theoretical models and N-body simulations can be applied with success.

In this Thesis, after discussing an effective theory of perturbations of single-field dark energy, we will explore the modifications on the mass function of dark matter haloes due to single-field dark energy models and to non-Gaussianity in the initial conditions. We first present an effective theory description of single field, minimally coupled dark energy models. A generic prediction of this study is that, whenever the equation of state w becomes less than -1, the speed of sound c_s of perturbations should be set to zero because of stability requirements. We will then show, using the spherical collapse and the excursion set theory formalism, how the mass function of haloes would be modified in models with non-clustering ($c_s = 1$) and clustering ($c_s = 0$) dark energy, and we speculate about possible signals in galaxy cluster observations. In the last chapter, we use a path integral approach to the excursion set theory to derive the mass function of haloes for a generic (small) non-Gaussianity in the initial conditions, which is a unique window into the physics of inflation.

Chapter 2

The accelerating Universe

2.1 Observational evidences

On large scales, the Universe is observed to be homogeneous and isotropic, and it can be described by the Friedmann-Lemaître-Robertson-Walker (FLRW) metric:

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2(t)\left[\frac{\mathrm{d}r^2}{1-kr^2} + r^2\mathrm{d}\Omega^2\right] , \tag{2.1}$$

where r, ϑ, ϕ are comoving polar coordinates, k quantifies the spatial curvature of constant-t slices, and the function $a(t)$ is the scale factor, the only degree of freedom allowed by symmetry.

The evolution equations are:

$$H^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} , \tag{2.2}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p) , \tag{2.3}$$

where $H \equiv \dot{a}/a$ is the Hubble parameter, and ρ and p are respectively the total energy density and pressure of the species present in the Universe at a given time, including the cosmological constant term. The first equation can be rewritten as

$$\sum_i \Omega_i(t) - 1 = \frac{k}{a^2(t)} , \tag{2.4}$$

where we have introduced the relative abundance $\Omega_i \equiv 8\pi G\rho_i/(3H^2)$ for any species i. Thus, the total energy density in the Universe determines its curvature: for a flat universe, $\Omega_{\mathrm{tot}} = 1$.

Measurements of the distances and of the expansion history are one of the main concerns of observational cosmology, since they determine the geometry and the energy content of the Universe.

2.1.1 Distance indicators

The physical distance to an object of (comoving) coordinate r depends on time, and it is $d(t) = a(t)r$. Cosmologists use the redshift, i.e. the fractional amount by which the wavelength of a photon has been stretched between the time the photon is emitted and the time it is received, as a measure of the scale factor: $z \equiv 1/a - 1$, having normalized $a_0 = 1$.

If an object of luminosity L is at a redshift z, a proxy for its distance to us is the *luminosity distance*

$$d_L \equiv \sqrt{\frac{L}{4\pi F}} \,, \tag{2.5}$$

where F is the energy flux received from that object. This relation generalizes the usual one in Minkowski spacetime. It is easy to compute that, in a flat Universe,

$$d_L = c(1+z) \int_0^z \frac{\mathrm{d}z'}{H(z')} \,. \tag{2.6}$$

Therefore, if there are sources of known intrinsic luminosity, known as "standard candles", a measure of their apparent brightness (i.e. the energy flux received on Earth) at different redshifts translates into a measure of the expansion history.

Type Ia supernovae are excellent standard candles. These are stellar explosions that occur when a white dwarf, onto which gas is accreting from some companion star, reach the Chandrasekhar mass limit (about 1.4 solar masses), the maximum possible mass that can be supported by electron Fermi pressure. When this happens, the gravitational collapse that follows results in an extremely bright explosion, with a luminosity that can reach a significant fraction of that of the host galaxy. Because of the prompt explosion after reaching the Chandrasekhar mass, the variations in absolute luminosity are small. Such variations can depend on the environment or the chemical composition of the star, and they appear to be well correlated with the rise and decline time of the ligth curve. The relationship between peak luminosities and light curve shape has been well-studied, and today type Ia supernovae are reliably used as standard candles for measurements of cosmological distances.

In 1998, two independent groups, the Supernova Cosmology Project [17] and the High-z Supernovae Search Team [18] measured luminosity distances of many SNIa as a function of redshifts up to $z \lesssim 1$. The results show that the most distant supernovae appear dimmer than expected in an Einstein-de Sitter Universe, which means that the expansion is accelerating. In subsequent years, many more supernovae at high redshifts have been detected and studied. The trend towards fainter supernovae at redshift $z \gtrsim 1$ has been shown to reverse, which means that the acceleration took

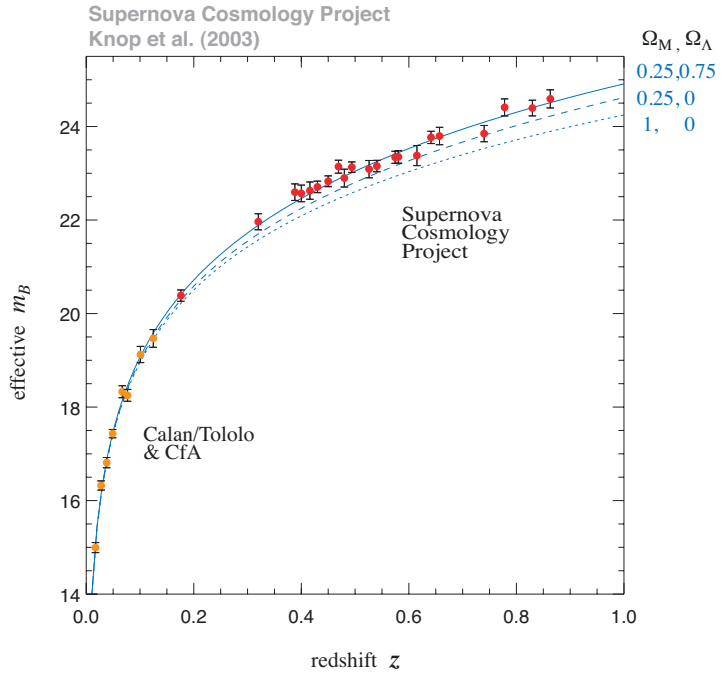


Figure 2.1: The Hubble diagram of SNIa, showing the apparent magnitude versus redshift. The solid line is the best fit ΛCDM cosmology. The dashed line is the prediction for an open Universe, while the dotted one corresponds to an EdS cosmology. Figure taken from [19].

place around $z \simeq 1$; such measurements play also a key role in disregarding dust as an explanation of the dimming. We show the Hubble diagram taken from [19] in fig. 2.1.

Another definition of distance is the *angular diameter distance*. If an object has a proper size l_p orthogonal to the line of sight, and it is observed to subtend an angle θ on the sky, its angular diameter distance is

$$d_A \equiv \frac{l_p}{\theta} \ . \tag{2.7}$$

The relation to the luminosity distance is $d_A(z) = d_L(z)/(1+z)^2$. An object of which we know the proper size is called a "standard ruler". In our Universe a standard ruler is the size of the sound horizon at recombination H_r^{-1}, which is imprinted on the CMB temperature spectrum as the maximum amplitude scale. This is because perturbations on larger length scales cannot be affected by causal physics and they just bear the imprint of the primordial perturbations, while on smaller scales there is the propagation of damped sound waves, generated by the competing forces of gravitational infall and photon pressure. When photons decouple from the plasma, we expect to see a peak in the CMB temperature power spectrum corresponding to the size of the sound horizon

at recombination. Since we know the physical size of H_r^{-1}, a precise measurement of the angular distance of the CMB is given by the position of the first peak in the temperature power spectrum. The observed value, which is $l \simeq 220$, corresponds to the geometry of a flat Universe, and it is almost independent of the relative abundances Ω_m and Ω_Λ.

2.1.2 CMB and LSS

The analysis of the CMB anisotropies is an invaluable tool for extracting the cosmological parameters. The large-scale plateau of the power spectrum can be used as a probe of acceleration, through the integrated Sachs-Wolfe (ISW) effect. In a flat, matter-dominated Universe, the gravitational potentials are constant and therefore do not affect the energy of a photon traversing them. On the contrary, photons traversing a shallowing (deepening) potential will be blueshifted (redshifted), as the energy they lose when they climb out of the well is less (more) than the energy they gain when they enter. The cleanest way to detect the ISW effect is to cross-correlate the CMB power spectrum with the galaxy one, as the signal is expected to be correlated with the tracers of the gravitational potentials [20]. Analyses show a 4σ detection [21, 22].

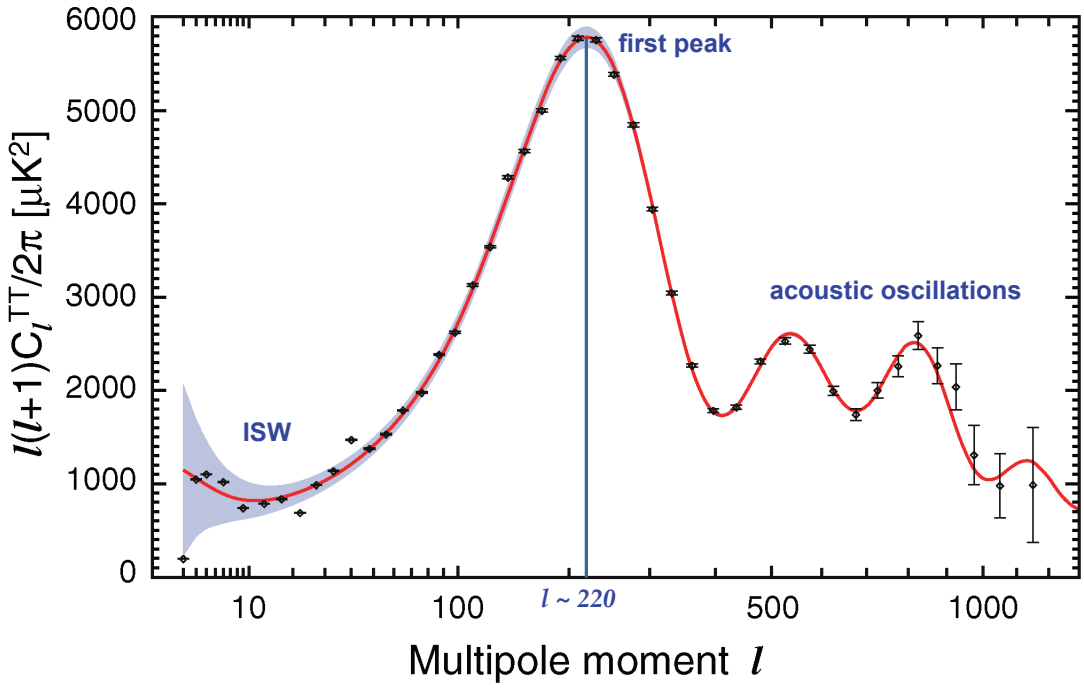


Figure 2.2: The CMB temperature power spectrum, as measured by the WMAP collaboration [1]. The solid line is the best fit ΛCDM model. We have shown the ISW on large scales, the position of the first acoustic peak and the subsequent acoustic oscillations.

2.1 Observational evidences

The acoustic oscillations in the plasma before recombination leave also a characteristic imprint on the distribution of matter. There is a bump (called acoustic peak) in the correlation function of the density fluctuations at a comoving separation of $\sim 100h^{-1}$Mpc; equivalently, we observe oscillations in the matter power spectrum, known as baryon acoustic oscillations (BAOs). These have been detected in galaxy surveys [23]. The physical wavenumber at which they occur is well understood from linear perturbation theory, which allows us to use BAOs as a cosmological probe. In a galaxy survey, the acoustic scale can be measured both across and along the line of sight, to place constraints both on the angular diameter distance and on the expansion history.

Another powerful probe of acceleration is the growth rate of density inhomogeneities. At early times, or on large scales, the fractional density perturbation is very small and it can be studied by perturbation theory. In the linear regime, it satisfies the equation

$$\ddot{\delta}_m + 2H\dot{\delta}_m - \frac{3}{2}\Omega_m H^2 \delta_m = 0 \,, \tag{2.8}$$

whose growing mode solution is $\delta_m(z) \propto D(z)$, where $D(z)$ is called the growth function and depends on the background expansion history. The growth function can be measured by observations of clusters and gravitational lensing.

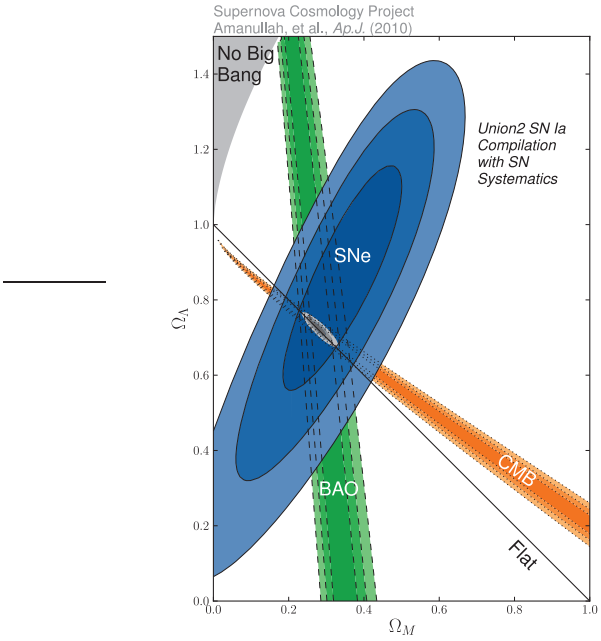


Figure 2.3: Constraints in the $\Omega_m - \Omega_\Lambda$ parameter space from SNeIA, BAOs and CMB data. Different datasets are degenerate along different lines. Figure taken from [24].

2.2 Theoretical explanations

The acceleration of the Universe has been now firmly established by several independent observations, and the literature on its theoretical explanation has grown enormously in the last few years.

A possible view is that the acceleration is due to the local environment of our own Galaxy. In particular, if we happen to live near the center of a large void, measurements of background can mimic an accelerating Universe.

Instead, if we stick to the FLRW cosmology, from the equation

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p) \tag{2.9}$$

we infer that acceleration requires a matter source which violates the strong energy condition, $p < -\rho/3$, named dark energy.

Such a source of gravity is, actually, already present in GR. In fact, to the Einstein-Hilbert Lagrangian it is always possible to add the cosmological constant term:

$$\int \mathrm{d}^4x\sqrt{-g}\left[\frac{M_{\rm Pl}^2}{2}R - 2\Lambda\right] , \tag{2.10}$$

where Λ is a costant. The form of such a term is the same as the energy-momentum tensor of a perfect fluid with $p = -\rho$, and it can be interpreted as the energy of the vacuum. The zero-point energy in quantum mechanics is usually subtracted, since only energy differences are measured; in GR, however, it gravitates as all other forms of energy. If one tries to compute the value of Λ from standard quantum field theory, the result is proportional to the fourth power of the cutoff, usually assumed to be the Planck mass. Observations show instead that $\Lambda \simeq 10^{-120} M_{\rm Pl}^4$, which is a huge fine tuning. So far, there is no convincing theoretical explanation for this value. A possible reason is the anthropic one, pioneered by Weinberg [25], and now accepted by many theorists in the context of the the landscape of string theory [26].

The dark energy need not be a constant, and is usually described by parametrizing the equation of state as $p = w\rho$, where w is in general time-dependent; constraints on w show that it is close to -1 with a 10% accuracy. A form of matter that gives rise to this equation of state is a scalar field which slowly rolls down its potential. In this case, perturbations in the scalar field can give interesting observational signatures. The simplest models in this class are minimally coupled models, in which the scalar field couples to gravity through the minimal coupling prescription, and does not directly interact with matter. We study such models in chapter 3, using an effective theory approach.

Another interesting class of models are scalar-tensor theories and $f(R)$ theories, which are a subset of the former. In these models, the scalar field is directly coupled to the Ricci scalar (in Jordan frame) or, performing a conformal transformation of the metric, to the matter fields (in the Einstein frame). It is possible to construct viable models which evade the stringent solar system tests of GR by giving the scalar a large effective mass in high density regions, and whose background evolution $H(z)$ is the same as in ΛCDM cosmology. The difference here arises at the perturbation level. A generic prediction of such models is that the two linear gravitational potentials ϕ and ψ differ, and the growth function of density perturbations becomes scale-dependent. Therefore, only measurements of the growth of structures are able to distinguish these theories from ΛCDM.

A more radical solution is represented by theories in which the gravity theory is different from general relativity, like massive gravity [27], ghost condensation [28], the DGP model [29], the galileon [30]. In this case, there is a non-trivial mixing of the scalar degree of freedom with the graviton and there are deviations from Newton's law already around a Minkowski background, which are constrained to be tiny. Solar system tests are usually evaded by non-linear interactions of the scalar field. In general, modified gravity theories can be falsified or constrained by measurements of the growth function.

2.3 Future prospects

There has recently been a systematic study of the future prospects in dark energy observations, summarized in the Dark Energy Task Force paper [31]. It is generally recommended that the best approach is to use multiple techniques at every experimental stage, at least one of which is a probe sensitive to the growth function of perturbations. The main areas of observations which are useful for dark energy studies are BAOs, galaxy clusters, supernovae and weak lensing.

Baryon acoustic oscillations, as explained above, are a well-understood standard ruler, and this technique allows to measure both the proper distance (from oscillations viewed transversally) and the expansion history (from oscillations viewed radially). This probe is the least affected by systematic uncertainties, but its potential to distinguish evolution of the equation of state parameter is somewhat limited by the fact that the most precise measurements are made at $z > 1$. From the theoretical side, an issue with the use of BAOs is the uncertainty in the effect of non-linear effects on the galaxy power spectrum. Deep galaxy surveys and 21 cm observations will greatly improve BAOs observations and the determination of background quantities.

Cluster abundances are a powerful probe of the growth function, because of the exponential dependence of the mass function on the linear field. The measured number density of clusters is also sensitive to the proper volume in the Universe, which is proportional to a combination of proper distance and expansion history. The strength of cluster counts is that clusters can be detected in several different ways: X-ray emission, Sunyaev-Zeldovich emission, lensing shear measurements and optical observations. The main uncertainty is given by the relation between the mass and the observables, which depends on the complexity of baryonic physics. To overcome the uncertainty in the mass-observable relation, one can apply "self-calibration" methods, which consist in modelling unknown physics using nuisance parameters, which are to be determined along with the cosmological parameters. With future missions, the number of detected cluster will increase substantially, and observations with diverse techniques can allow to reduce systematic uncertainties.

Supernovae are the most established method to constrain the acceleration of the Universe. They are only sensitive to the background expansion history, but they are a useful tool in the interesting redshift range $0 < z < 2$. The main uncertainties are in the standardization of light curves and the possible foreground extinction. Future surveys should provide a huge number of SNeIa, with well determined photometric redshifts.

Weak gravitational lensing is an emerging technique, but it is the method with the greatest potential in constraining dark energy, and it is sensitive to both the background and the growth function. A lot of statistics can be computed, allowing internal tests and corrections of systematic errors. It is difficult to predict how well this method can perform, but it is likely to be dominated by systematic errors in redshift determinations and in galaxy shape measurements. These problems can be bypassed by using different kinds of observations (terrestrial, space and 21 cm), and by further simulations and testing of shape measurements.

Chapter 3

The Effective Theory of Quintessence

The origin of the present acceleration of the Universe is likely to be the most important theoretical problem in physics today. Given the general reluctance in accepting as explanation an incredibly small cosmological constant and the absence of compelling alternatives, it seems that one should keep an open-minded approach, concentrating on very general theoretical constraints and on observables more than on specific models.

In this chapter we study in generality, and focusing on perturbations, dark energy scenarios where the dark sector is described by a single scalar degree of freedom, without direct coupling to matter (in the Einstein frame). We will often call this general model *quintessence*, although in the literature this name is usually reserved to a scalar field with a canonical kinetic term.

Following [32, 33], we will rewrite the scalar field Lagrangian in order to make explicit what is the most general theory of quintessence perturbations around a given background solution characterized by its pressure p_Q and energy density ρ_Q. In this way, the freedom that we have after the unperturbed history is fixed is made clear. This separation is particularly important given that a host of new experiments is going to test dark energy clustering properties [31]. In our formalism, the general theoretical constraints on single field models are also made clear.

In particular, we will study whether a single field model that is safe from ghost and gradient instabilities can have an equation of state $w_Q < -1$, where $w_Q = p_Q/\rho_Q$. In this regime, the stability of the model can be guaranteed by the presence of higher derivative operators, a conclusion already reached in [32], where single field models were studied focusing on the constraints enforced by stability. Here, after reviewing and extending the results of [32], we will concentrate on the behaviour of cosmological perturbations, which are relevant for observations. On cosmological scales we find that these higher derivative terms are irrelevant for the phenomenology, so that

a model with $w_Q < -1$ simply behaves as a k-essence fluid with virtually zero speed of sound. Higher derivative terms are relevant for cosmology only when the equation of state gets very close (and experimentally indistinguishable from) a cosmological constant. In this limit our general Lagrangian reduces to the Ghost Condensate theory [28], smoothly connecting quintessence to this theory of modification of gravity. Notice that, as detailed in AppendixA, we are interested in a regime where higher derivative terms do not introduce additional degrees of freedom (contrary to what happens, for example, in [34]). We find it convenient to summarize our results in the plane $(1 + w_Q)\Omega_Q$ vs. c_s^2, where Ω_Q is the quintessence contribution to the critical density. We dub this plane of parameters the *quintessential plane*.

We also study the issue of whether it is possible to cross the so-called phantom divide $w_Q = -1$ [35, 36]. We find that the speed of sound vanishes exactly at the divide [37, 38] and since quintessence may remain stable for $w_Q < -1$ there is no general pathology associated with the crossing. We show this explicitly with an example. The phantom divide can be crossed with a single scalar degree of freedom, without introducing ghost-like fields.

This chapter is organized as follows. In section 3.1 we study the most general theory of single field quintessence, taking into account higher derivative operators and focusing on the stability constraints following [32]. An alternative way of deriving the action for perturbations is presented in 3.2, following the approach of [32, 33]. In section 3.3 we study the phenomenology in various limits, considering also the gravitational effect of dark matter on quintessence. In section 3.4 we consider the issue of crossing the phantom divide $w_Q = -1$ and we show explicit examples of the crossing without pathologies. In section 3.5 we concentrate on another kind of higher derivative operators [32, 33], different from the ones studied for the Ghost Condensate. Although the phenomenology on cosmological scales does not change, the modification of gravity at short distances is quite different.

Several issues concerning our effective theory approach are left to the appendices. Appendix A is devoted to reviewing how higher derivative operators must be treated in the effective field theory approach. In appendix B we discuss the modification of gravity induced by the kind of higher derivative operators that were not studied in [28].

3.1 Effective theory of quintessence

Our aim is to study the most general theory of quintessence perturbations. We will do it step by step, first by considering a model with an action containing at most a single derivative acting on

the field. This is known as k-essence [39, 40] and it will be possible to write the action for the perturbations in such a way as to make explicit the dependence on the background energy density and pressure ρ_Q and p_Q. Then we will add higher derivative operators to the k-essence action in such a way as to leave the background invariant. In this section we will consider the kind of operators introduced in the context of ghost condensation. Other higher derivative operators will be discussed later in section 3.5.

An alternative derivation of the most general action for quintessence perturbations is given in the next section 3.2, following the approach of refs. [32, 33], that consists in writing down all the terms preserving the symmetries of the system in a 'unitary' gauge, where the quintessence perturbation is set to zero and appears as a scalar metric degree of freedom. This approach is elegant and straightforward but less pedagogical than the one adopted in this section. Obviously, both approaches lead to the same physical results.

3.1.1 The limit of k-essence

Let us start with a k-essence action

$$S = \int \mathrm{d}^4x \sqrt{-g}\, P(\phi, X) \,, \qquad X = -g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi \,. \tag{3.1}$$

We assume a flat Friedmann-Lemaître-Robertson-Walker (FLRW) Universe with metric $\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2(t)\mathrm{d}\vec{x}^2$. Initially, we will treat this as a fixed background and neglect the perturbations of the metric.

To describe perturbations around a given background solution $\phi_0(t)$, it is useful to write the scalar field as

$$\phi(t, \vec{x}) = \phi_0(t + \pi(t, \vec{x})) \,, \tag{3.2}$$

and expand the action (3.1) in terms of π. In the following, we are going to assume that the function $\phi_0(t)$ is strictly monotonic, $\dot{\phi}_0(t) \neq 0$, to avoid singularities in the relation between ϕ and π.

Using the expansions

$$\phi(t, \vec{x}) = \phi_0 + \dot{\phi}_0\pi + \frac{1}{2}\ddot{\phi}_0\pi^2 + \dots \,, \tag{3.3}$$

$$X(t, \vec{x}) = X_0 + \dot{X}_0\pi + \frac{1}{2}\ddot{X}_0\pi^2 + 2X_0\dot{\pi} + 2\dot{X}_0\pi\dot{\pi} + X_0\left(\dot{\pi}^2 - \frac{(\nabla\pi)^2}{a^2}\right) + \dots \,, \tag{3.4}$$

where $X_0 = \dot{\phi}_0^2$, we have, up to second order in π,

$$S = \int \mathrm{d}^4x\, a^3 \Big[P_0 + \dot{P}_0\pi + \frac{1}{2}\ddot{P}_0\pi^2 + 2P_X X_0 \dot{\pi} + 2\left(P_X X_0\right)^{\cdot}\pi\dot{\pi} + P_X X_0 \left(\dot{\pi}^2 - \frac{(\nabla\pi)^2}{a^2}\right) + 2P_{XX}X_0^2\dot{\pi}^2 \Big] \,, \quad (3.5)$$

where $P_X = \partial P/\partial X|_0$ and $P_{XX} = \partial^2 P/\partial X^2|_0$. The term P_0 does not affect perturbations as it is independent of π, while one can verify that the linear terms cancel using the background equation of motion. Indeed, by integrating by parts the term $\pi\dot{\pi}$ and making use of the background equation of motion, after some manipulations we are left with

$$S = \int \mathrm{d}^4x\, a^3 \left[\left(P_X X_0 + 2P_{XX}X_0^2\right)\dot{\pi}^2 - P_X X_0 \frac{(\nabla\pi)^2}{a^2} + 3\dot{H}P_X X_0\, \pi^2 \right] . \quad (3.6)$$

We can now rewrite the coefficients of this expansion in terms of the stress-energy tensor of the background solution. From the definition of the stress-energy tensor,

$$T_{\mu\nu} = -\frac{2}{\sqrt{-g}}\frac{\delta S}{\delta g^{\mu\nu}} \,, \quad (3.7)$$

one obtains the background energy density and pressure:

$$\rho_Q = 2X_0P_X - P_0 \,, \qquad p_Q = P_0 \,. \quad (3.8)$$

Using these expressions, the action above can be cast in the form

$$S = \int \mathrm{d}^4x\, a^3 \left[\frac{1}{2}\left(\rho_Q + p_Q + 4M^4\right)\dot{\pi}^2 - \frac{1}{2}(\rho_Q + p_Q)\frac{(\nabla\pi)^2}{a^2} + \frac{3}{2}\dot{H}(\rho_Q + p_Q)\pi^2 \right] . \quad (3.9)$$

Here we have defined $M^4 \equiv P_{XX}X_0^2$, where M has the dimension of a mass.

At this stage we can straightforwardly introduce the coupling with metric perturbations. This coupling is particularly simple in synchronous gauge, where the metric takes the form

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2(t)(\delta_{ij} + h_{ij})\mathrm{d}x^i\mathrm{d}x^j \,. \quad (3.10)$$

Indeed, at quadratic order in the action the coupling with gravity only comes through the perturbed $\sqrt{-g}$ in the action. Replacing a^3 with $a^3(1 + h/2)$ in eq. (3.5) we have

$$S = \int \mathrm{d}^4x\, a^3 \left(1 + \frac{h}{2}\right)\left[\dot{P}_0\pi + 2P_X X_0\dot{\pi} + \ldots\right] . \quad (3.11)$$

Integrating by parts and using again the background equation of motion one gets the full action for π:

$$S = \int \mathrm{d}^4x\, a^3 \left[\frac{1}{2}\left(\rho_Q + p_Q + 4M^4\right)\dot{\pi}^2 - \frac{1}{2}(\rho_Q + p_Q)\frac{(\nabla\pi)^2}{a^2} + \frac{3}{2}\dot{H}(\rho_Q + p_Q)\pi^2 - \frac{1}{2}(\rho_Q + p_Q)\dot{h}\pi \right] . \tag{3.12}$$

The quadratic Lagrangian is thus specified by the functions $(\rho_Q + p_Q)(t)$ and $M^4(t)$. It is important to stress that these two functions are completely unconstrained. For any choice of these two functions one can in fact construct a Lagrangian $P(\phi, X)$, such that the quadratic Lagrangian around the unperturbed solution has the form (3.12). Let us see explicitly how this works. First of all, given $\rho_Q + p_Q$ (and possibly the other contributions to the total energy density and pressure coming from other components), one can find the two functions $\rho_Q(t)$ and $p_Q(t)$ solving the Friedmann equations[1]. At this point it is easy to check that the Lagrangian

$$P(\phi, X) = \frac{1}{2}(p_Q - \rho_Q)(\phi) + \frac{1}{2}(\rho_Q + p_Q)(\phi)X + \frac{1}{2}M^4(\phi)(X-1)^2 \tag{3.13}$$

has the solution $\phi = t$, gives the requested pressure and energy density as a function of time, and gives eq. (3.12) as the quadratic action for perturbations. Note that the dimension of ϕ is that of an inverse of a mass. The coefficient M^4 is time dependent and for quintessence we expect that it varies with a time scale of order H^{-1}. Somewhere in the following, to simplify the calculations, we take $M^4 = \text{const}$.

One advantage of the action (3.12) is that the coefficients of all the terms are physically measurable quantities – we will see below that M^4 is related to the sound speed of perturbations. This standard form of rewriting the action of quintessence perturbations does not suffer from field redefinition ambiguities. Indeed, there is an infinite number of physically equivalent Lagrangians $P(\phi, X)$ related by field redefinitions $\phi \to \tilde{\phi}(\phi)$, but they all give the same action (3.12). Note that for most purposes the explicit construction of the action in terms of ϕ is irrelevant – this is apparent in the "unitary" gauge approach, which relies on the symmetries of the Lagrangian. Indeed, one is free to choose *any* function $\rho_Q(a)$ to describe the evolution of the quintessence energy density as a function of the scale factor. Then the action (3.12) will describe the perturbations around this background. In particular, one can always make a field redefinition such that $\phi = t$, as we did in eq. (3.13).

[1] One has to solve the continuity equation for quintessence, $\dot{\rho}_Q + 3H(\rho_Q + p_Q) = 0$, where $\rho_Q + p_Q$ is a known function of time, together with the Friedmann equation, $H^2 = (\rho_Q + \rho_{\rm rest})/3M_{\rm Pl}^2$, where $\rho_{\rm rest}$ includes all the additional sources of energy density in the Universe. These equations can be integrated up to a constant in the initial condition. This ambiguity corresponds to a shift in the cosmological constant, which does not enter in the action for perturbations.

Let us see what are the theoretical constraints that we can put on the general form of the action (3.12). A basic requirement that we will impose on our theory is that it is not plagued by ghosts, i.e. that its kinetic energy term is positive:

$$\rho_Q + p_Q + 4M^4 > 0 \,. \tag{3.14}$$

The presence of a sector with the "wrong" sign of the energy implies that the Hamiltonian is unbounded from below. If one studies this sector alone, no pathology arises as the sign of the energy is a matter of convention. However, quintessence is (at least) gravitationally coupled to the rest of the world, so that there is the danger of exchanging energy without bound between a healthy sector and a negative-energy one. Classically, this is not a problem if quintessence perturbations are very small and remain in the linear regime. Therefore, for a quintessence with negligible clustering (with speed of sound $c_s \sim 1$), there is no obvious classical danger. At the quantum level, however, the situation is more pathological. The vacuum is unstable to the spontaneous decay into positive and negative energy states and the decay rate is UV-divergent and strictly infinite in Lorentz invariant theories [41]. Although it has been shown that it is possible to cut-off this instability in a non-Lorentz invariant theory [42], in this work we take a more conservative approach and forbid the existence of ghosts.

If we set $M^4 = 0$ in the general action (3.12), we reduce to the case of a standard quintessence field with canonical kinetic term $(\partial\phi)^2$. In this case, forbidding the ghost implies $\rho_Q + p_Q > 0$. Thus, as it is well known, a scalar field with minimal kinetic term can violate the null energy condition, equivalent in a cosmological setting to $\rho + p < 0$, only if it is a ghost [43]. In this simple case, the speed of sound of scalar fluctuations is $c_s^2 = 1$. When M^4 does not vanish the speed of sound of fluctuations differs from unity [44] and reads

$$c_s^2 = \frac{\rho_Q + p_Q}{\rho_Q + p_Q + 4M^4} \,. \tag{3.15}$$

One can see that, for $\rho_Q + p_Q > 0$, i.e. for positive c_s^2, $M^4 < 0$ implies that scalar perturbations propagate super-luminally [45, 46]. This is problematic in a theory with a Lorentz invariant UV completion [47].

From the action (3.12) we see that in the presence of M^4 there is no generic connection between the violation of the null energy condition and a wrong sign of the time-kinetic term. The coefficient in front of $\dot{\pi}^2$ can be positive also when $\rho_Q + p_Q < 0$. On the other hand, $\rho_Q + p_Q$ fixes the sign of the term in front of the spatial kinetic term $(\nabla\pi)^2$ [48]. Thus, in absence of ghosts, the violation of

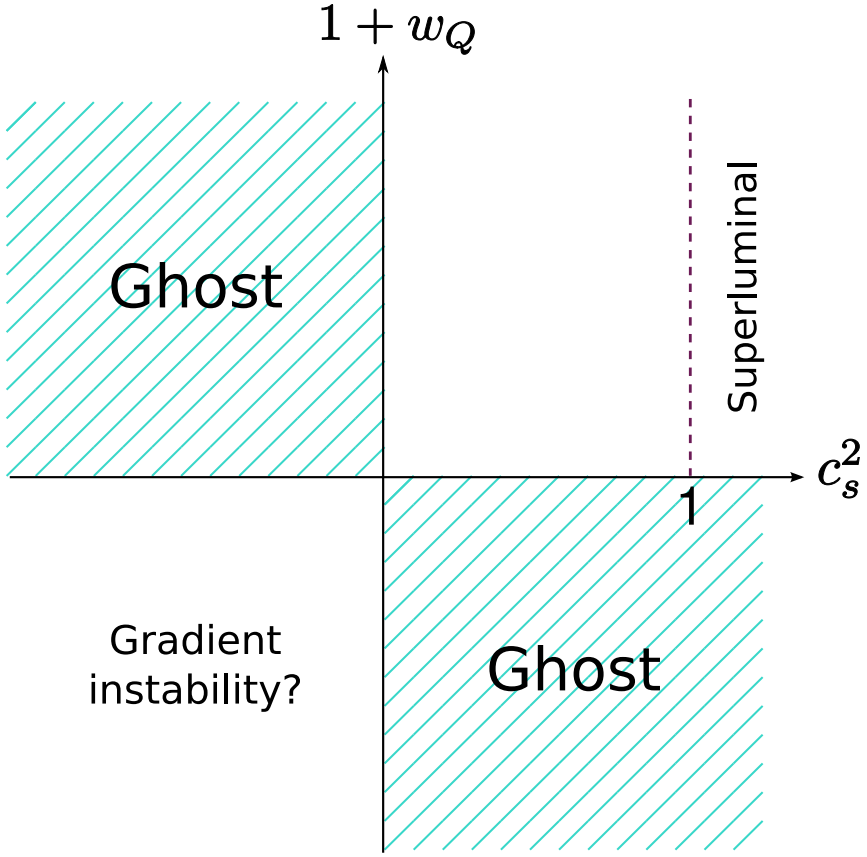


Figure 3.1: The quintessential plane $1 + w_Q$ vs. c_s^2 in the case of k-essence. If we require the absence of ghosts, the sign of the spatial kinetic term is fixed to be the same as $1 + w_Q$, so that one has to worry about gradient instabilities for $1 + w_Q < 0$. For $1 + w_Q > 0$ one has superluminal propagation if $M^4 < 0$.

the null energy condition implies a negative speed of sound squared. The constraints that we have derived are summarized in the *quintessential plane* $1 + w_Q$ vs. c_s^2, represented in figure 3.1.

An imaginary speed of sound – $c_s^2 < 0$ – represents a gradient instability of the system. Taking $M^4 \gg |\rho_Q + p_Q|$ the gradient term is suppressed and the instability is irrelevant for scales of cosmological interest in the sense that the instability time is much longer than the age of the Universe. Still, the instability is relevant for short wavelengths so that it seems difficult to make sense of the $w_Q < -1$ region [37]. However, this conclusion is too hasty. Indeed, when the term $(\nabla\pi)^2$ is suppressed, higher derivative operators may become relevant, as we will now discuss.

3.1.2 Higher derivative terms and stability

We have just shown that, in the limit in which the quintessence speed of sound goes to zero, the standard spatial kinetic term vanishes. Obviously, the action (3.1) is not the end of the story: the full Lagrangian will contain higher derivative operators such as $Q(X)R(\Box\phi)$ and these will give rise to the leading higher derivative spatial kinetic term. A generic higher derivative operator will however change the background solution of (3.1), while here we want to study the effect of higher derivative operators on quintessence perturbations around a given background, as we did in eq. (3.13). To keep the background unchanged, let us add to the Lagrangian (3.13) the operator

$$\mathcal{L}_{\bar{M}} = -\frac{\bar{M}^2}{2}(\Box\phi + 3H(\phi))^2 \ . \tag{3.16}$$

For reasons that will become clear later, we need $\bar{M}^2 > 0$[2]. This term does not alter the background evolution $\phi = t$, $\rho_Q(t)$ and $p_Q(t)$. Indeed, $\Box\phi + 3H(\phi)$ vanishes on the background so that the operator is explicitly quadratic in the perturbations. At quadratic order this operator reads (neglecting for the moment metric perturbations)

$$\mathcal{L}_{\bar{M}} = -\frac{\bar{M}^2}{2}\left(\ddot{\pi} + 3H\dot{\pi} - 3\dot{H}\pi - \frac{\nabla^2\pi}{a^2}\right)^2 \ . \tag{3.17}$$

One may worry about the presence of terms with higher time derivatives, as these would naïvely be associated with additional solutions of the equation of motion. However, if one compares $\bar{M}^2\ddot{\pi}^2$ with the standard time kinetic term $M^4\dot{\pi}^2$ of eq. (3.12) – assuming $\bar{M} \sim M$ – the former is always suppressed with respect to the latter for frequencies below the scale M. In general, we expect that for frequencies $\omega \sim M$ all operators containing higher time derivatives become important, so that the scale $M \sim \bar{M}$ sets the maximum energy scale for which the theory makes sense: it is the energy cutoff. This is the standard situation in an effective field theory: higher derivative terms become important for energies of the order of the cutoff and at lower energies they must be treated perturbatively. In particular, there is no physical meaning in the new solutions that arise from taking higher and higher time derivatives. We postpone a complete discussion about this point to appendix A. Notice that the same argument cannot be used for the operator $-\bar{M}^2(\nabla^2\pi)^2/2$. Indeed, in the limit of small $\rho_Q + p_Q$ there is no spatial kinetic term of the form $(\nabla\pi)^2$ so that $-\bar{M}^2(\nabla^2\pi)^2/2$ is the leading spatial kinetic term. At short scales we have a non-relativistic dispersion relation of the form $\omega \simeq k^2/M$, which implies that energy and momentum behave very

[2]As M^4, also $\bar{M}^2$ can have a time dependence on a time scale of order H^{-1}. For simplicity, in the following we assume that it is constant.

differently (as we will see in appendix A, they have different *scaling dimensions*). In particular, when comparing the first and last terms in brackets in eq. (3.17) we have $\nabla^2\pi \sim M\dot{\pi} \gg \ddot{\pi}$, for energies below the cutoff. This means that we can drop $\ddot{\pi}$ altogether from eq. (3.17).

It is important to stress that there is no fine tuning in the limit $|\rho_Q + p_Q| \ll M^4$ – or equivalently $|c_s^2| \ll 1$ – as this limit is "technically natural", i.e., there is a symmetry that is recovered in the limit $\rho_Q + p_Q = 0$. Indeed, as shown below, in this limit we obtain the Ghost Condensate theory [28], which is invariant under the shift symmetry $\phi \to \phi + \lambda$. In the presence of this symmetry the expansion of the Universe drives the background solution to $\rho_Q + p_Q = 0$. Thus, models with very small speed of sound should be thought of as small deformations of the Ghost Condensate limit [32, 33, 49].

Let us come back to the issue of stability for $\rho_Q + p_Q < 0$ including the new higher gradient term (3.17) and follow the discussion in [32]. As we discussed in the previous section, the only dangerous modes are those on scales much smaller than the Hubble radius, as their instability rate can be arbitrarily large; we thus concentrate on $k/a \gg H$. In this regime the operator (3.17) further simplifies as the second and third terms are negligible with respect to $\nabla^2\pi$. Considering the action (3.12) with the addition of the only remaining operator $-\bar{M}^2(\nabla^2\pi)^2/2$, the dispersion relation of π is thus modified to

$$\left(\rho_Q + p_Q + 4M^4\right)\omega^2 - (\rho_Q + p_Q)\frac{k^2}{a^2} - \bar{M}^2\frac{k^4}{a^4} = 0\,. \tag{3.18}$$

For $\rho_Q + p_Q < 0$, the k^2 gradient term has the unstable sign, but in the presence of the new operator this instability is confined to sufficiently large scales. In particular, the fastest rate of instability is given by

$$\omega^2_{\rm grad} \simeq -\frac{(\rho_Q + p_Q)^2}{\bar{M}^2 M^4}\,, \tag{3.19}$$

where we have taken $M^4 \gg |\rho_Q + p_Q|$. This gradient instability is not dangerous when it is slower than the Hubble rate, i.e., when

$$-\frac{\rho_Q + p_Q}{\bar{M}M^2} \lesssim H\,. \tag{3.20}$$

It is clear that a larger $\bar{M}$ makes the gradient instability slower.

However, a large $\bar{M}$ sources another form of instability, which contrarily to the gradient instability is already present for $\rho_Q + p_Q \geq 0$ and was originally discussed for the Ghost Condensate theory in [28]. Indeed, when the coupling with gravity is taken into account, the system shows a sort of Jeans instability, similarly to a standard fluid. To see this, let us introduce metric perturbations

and consider the limit $\rho_Q + p_Q = 0$. In this case, the complete Lagrangian reads, neglecting the expansion of the Universe,

$$S = \int \mathrm{d}^4x \left[2M^4\dot{\pi}^2 - \frac{\bar{M}^2}{2}\left(\frac{\dot{h}}{2} - \nabla^2\pi\right)^2 \right] . \tag{3.21}$$

We see that a large $\bar{M}$ enhances the mixing of π with gravity, i.e. the Jeans instability.

Thus, the equation of motion for π reads, in this case,

$$\ddot{\pi} + \frac{\bar{M}^2}{4M^4}\nabla^4\pi = \frac{\bar{M}^2}{8M^4}\nabla^2\dot{h} . \tag{3.22}$$

The gravitational perturbation h is sourced by the perturbations of π through Einstein's equations. In synchronous gauge, h satisfies [4]

$$\ddot{h} = -\frac{1}{M_{\rm Pl}^2}(\delta\rho_Q + 3\delta p_Q) , \tag{3.23}$$

where we have neglected the expansion of the Universe, and we have introduced the reduced Planck mass $M_{\rm Pl}^2 \equiv (8\pi G)^{-1}$. The stress-energy tensor can be straightforwardly derived by varying the action and the leading term is $\delta\rho_Q = 4M^4\dot{\pi}$, while the pressure perturbation is negligible[3]. The solution of eq. (3.23) can be plugged back into eq. (3.22). This yields the equation of motion of π taking into account its gravitational back-reaction:

$$\ddot{\pi} + \frac{\bar{M}^2}{4M^4}\nabla^4\pi = -\frac{\bar{M}^2}{2M_{\rm Pl}^2}\nabla^2\pi . \tag{3.26}$$

The mixing with gravity induces an unstable k^2 term in the dispersion relation, similarly to the gradient instability discussed above. We can compute again the fastest instability rate:

$$\omega^2_{\rm Jeans} \simeq -\left(\frac{\bar{M}M^2}{M_{\rm Pl}^2}\right)^2 . \tag{3.27}$$

[3]Indeed, we have

$$\delta\rho_Q = 4M^4\dot{\pi} + \bar{M}^2\left(\frac{\ddot{h}}{2} - \nabla^2\dot{\pi}\right) , \tag{3.24}$$

$$\delta p_Q = \bar{M}^2\left(\frac{\ddot{h}}{2} - \nabla^2\dot{\pi}\right) . \tag{3.25}$$

For $k \ll M$, $\bar{M}^2\nabla^2\dot{\pi} \ll M^4\dot{\pi}$. Moreover, eq. (3.23) shows that also the $\ddot{h}$ terms can be neglected in front of $M^4\dot{\pi}$, so that the operator proportional to $\bar{M}^2$ gives a negligible contribution to the stress-energy tensor.

As expected, in this case the instability gets worse for large $\bar{M}^2$, i.e. when the mixing with gravity is enhanced. By imposing that this instability rate is smaller than the Hubble rate[4], we obtain

$$\frac{\bar{M}M^2}{M_{\rm Pl}^2} \lesssim H\,. \tag{3.28}$$

Requiring that both stability conditions (3.20) and (3.28) are satisfied we get the window [32]

$$-(1+w_Q)\Omega_Q \lesssim \frac{\bar{M}M^2}{HM_{\rm Pl}^2} \lesssim 1\,. \tag{3.29}$$

In conclusion, considering higher derivative terms, a quintessence model with $w_Q \leq -1$ can be completely stable inside the window of parameters (3.29). On the other hand, eq. (3.29) indicates that it is difficult to avoid instabilities when $(1+w_Q)\Omega_Q \ll -1$. These stability constraints were already obtained, for $\Omega_Q = 1$, in [32].

3.2 General action in unitary gauge

In this section, we wish to obtain the general action for the quintessence perturbations by choosing a gauge where the scalar field perturbation is set to zero but it appears as a scalar metric degree of freedom, following [32, 33]. In this gauge, the constant time hypersurfaces are equivalent to the uniform field hypersurfaces. In the fluid language, this implies that the velocity is orthogonal to the constant time surfaces, $T^0_i = 0$, that is why this gauge is called 'velocity orthogonal' [51]. Using a particle physics terminology we can also call it 'unitary gauge', as all the degrees of freedom are in the metric. The unitary gauge corresponds to the choice $\phi(t,\mathbf{x}) = \phi_0(t)$, or $\pi = 0$.

In order to find the effective action, we write down all the terms that preserve the symmetries of the system. Our choice of gauge breaks time diffeomorphism invariance, while preserving invariance under spatial diffeomorphisms. Thus, we include linear and quadratic combinations of generic functions of time t, the time-time component of the inverse metric g^{00} and the extrinsic curvature of the constant time hypersurfaces. The effective action up to second order in perturbations is [32, 33]:

$$\begin{aligned} S = \int \mathrm{d}^3x\,\mathrm{d}t\,\sqrt{-g}\Bigg[&\frac{M_{\rm Pl}^2}{2}R + \mathcal{L}_m + c(t)g^{00} - \Lambda(t) + \frac{M^4(t)}{2}(g^{00}+1)^2 - \frac{\bar{M}^2(t)}{2}\delta K^2 \\ &- \frac{\tilde{M}^2(t)}{2}\delta K^i_{\ j}\delta K^j_{\ i} - \frac{\hat{M}^3(t)}{2}\delta K(g^{00}+1)\Bigg]\,, \end{aligned} \tag{3.30}$$

[4] A more careful analysis [50] indicates that this condition is very conservative and much larger instability rates can be experimentally allowed.

where R is the Ricci scalar and K_{ij} is the extrinsic curvature of constant t hypersurfaces, which at linear order reads

$$K_{ij} = \frac{1}{2}\sqrt{-g^{00}}\left(\partial_0 g_{ij} - \partial_i g_{0j} - \partial_j g_{i0}\right) , \tag{3.31}$$

and we have defined $\delta K_{ij} \equiv K_{ij} - a^2 H \delta_{ij}$, and $\delta K \equiv K^i_{\ i} - 3H$.

There are other operators that are invariant under spatial diffeomorphisms that one would naïvely include in this action. However, these operators are irrelevant at energy scales below the cutoff $M \sim \bar{M} \sim \hat{M}$. For instance, one could include the operator $(\dot{g}^{00})^2$. This would give the term $\ddot{\pi}^2$ in the final action, which indeed also appears as part of the operator (3.16) once expanded in the perturbations, eq. (3.17). However, as explained in section 3.1.2, for frequencies smaller than the cutoff $\ddot{\pi}^2$ is negligible with respect to $(\nabla^2\pi)^2$ so that it can be ignored in the action. For simplicity we will also ignore the operator proportional to $\tilde{M}^2$ as it leads to terms qualitatively similar to those proportional to $\bar{M}^2$.

One can easily fix the coefficients $c(t)$ and $\Lambda(t)$ by computing the background stress-energy tensor. This gives

$$\rho_Q = \Lambda(t) - c(t) , \tag{3.32}$$

$$p_Q = -c(t) - \Lambda(t) . \tag{3.33}$$

Finally, using these relations, we can write the action (3.30) in unitary gauge in terms of the background quantities $\rho_Q(t)$ and $p_Q(t)$:

$$S = \int \mathrm{d}^4x\,\sqrt{-g}\Big[\frac{M_{\rm Pl}^2}{2}R + \mathcal{L}_m + p_Q - \frac{1}{2}(\rho_Q + p_Q)(g^{00}+1) + \frac{M^4(t)}{2}(g^{00}+1)^2 \\ - \frac{\bar{M}^2(t)}{2}\delta K^2 - \frac{\hat{M}(t)^3}{2}\delta K(g^{00}+1)\Big] . \tag{3.34}$$

Now we can rewrite this action in a gauge-invariant form, using the so-called Stueckelberg trick. We perform the following time-coordinate transformation:

$$t \to \tilde{t} = t + \pi(x) \qquad x^i \to \tilde{x}^i = x^i , \tag{3.35}$$

that reintroduces π. The action for π reads, up to second order,

$$\begin{aligned} S = \int \mathrm{d}^4x\,\sqrt{-g}\Big\{ & p_Q + \dot{p}_Q\pi + \frac{1}{2}\ddot{p}_Q\pi^2 \\ & - \frac{1}{2}(\rho_Q + p_Q)\left[(g^{00}+1) - 2\dot{\pi} + 2(g^{00}+1)\dot{\pi} - \dot{\pi}^2 + 2g^{0i}\partial_i\pi + \frac{(\nabla\pi)^2}{a^2}\right] \\ & - \frac{1}{2}(\dot{\rho}_Q + \dot{p}_Q)\pi\left[(g^{00}+1) - 2\dot{\pi}\right] + \frac{M^4(t)}{2}\left[(g^{00}+1) - 2\dot{\pi}\right]^2 \\ & - \frac{\bar{M}^2(t)}{2}\left(\delta K - 3\dot{H}\pi - \frac{\nabla^2\pi}{a^2}\right)^2 - \frac{\hat{M}(t)^3}{2}\left(\delta K - 3\dot{H}\pi - \frac{\nabla^2\pi}{a^2}\right)\left[(g^{00}+1) - 2\dot{\pi}\right]\Big\}\,, \end{aligned} \tag{3.36}$$

while the part of the action containing R and $\mathcal{L}_m$ is invariant under general diffeomorphisms. We now choose to work in the synchronous gauge, which is defined in eq. (3.10). Using the notation of [4], the two scalar degrees of freedom of h_{ij} are its trace $h \equiv \delta^{ij}h_{ij}$ and η which is defined by $\nabla^2 h_{ij} \equiv \partial_i\partial_j h + 6(\partial_i\partial_j - \frac{1}{3}\delta_{ij}\nabla^2)\eta$. Using this metric, after integrating by parts and using the background continuity equation $\dot{\rho}_Q + 3H(\rho_Q + p_Q) = 0$, the action (3.36) takes the form

$$\begin{aligned} S = \int \mathrm{d}^4x\,a^3\Big\{ & p_Q + \frac{1}{2}(\rho_Q + p_Q)\left[\dot{\pi}^2 - \frac{(\nabla\pi)^2}{a^2}\right] + 2M^4\dot{\pi}^2 + \frac{3}{2}\dot{H}(\rho_Q + p_Q)\pi^2 - \frac{1}{2}(\rho_Q + p_Q)\dot{h}\pi \\ & - \frac{\bar{M}^2}{2}\left(\frac{1}{2}\dot{h} - 3\dot{H}\pi - \frac{\nabla^2\pi}{a^2}\right)^2 + \hat{M}^3\dot{\pi}\left(\frac{1}{2}\dot{h} - 3\dot{H}\pi - \frac{\nabla^2\pi}{a^2}\right)\Big\}\,. \end{aligned} \tag{3.37}$$

We can now compute the stress-energy tensor of quintessence using the action (3.36). Expanding in the perturbations, its components read

$$\begin{aligned} T_{00} = {} & \rho_Q + \left(\rho_Q + p_Q + 4M^4\right)\dot{\pi} + \dot{\rho}_Q\pi \\ & - 3H\bar{M}^2\left(\frac{1}{2}\dot{h} - 3\dot{H}\pi - \frac{\nabla^2}{a^2}\pi\right) + \hat{M}^3\left(\frac{1}{2}\dot{h} - 3\dot{H}\pi - \frac{\nabla^2}{a^2}\pi + 3H\dot{\pi}\right)\,, \end{aligned} \tag{3.38}$$

$$T_{0i} = (\rho_Q + p_Q)\partial_i\pi + \bar{M}^2\partial_i\left(\frac{1}{2}\dot{h} - 3\dot{H}\pi - \frac{\nabla^2\pi}{a^2}\right) - \hat{M}^3\partial_i\dot{\pi}\,, \tag{3.39}$$

$$\begin{aligned} T_{ij} = {} & p_Q a^2\delta_{ij} + \left[\dot{p}_Q\pi + (\rho_Q + p_Q)\,\dot{\pi}\right]a^2\delta_{ij} + p_Q a^2 h_{ij} \\ & + 2\bar{M}^2 a^2\delta_{ij}(\partial_0 + 3H)\left(\frac{1}{2}\dot{h} - 3\dot{H}\pi - \frac{\nabla^2\pi}{a^2}\right) - 2\hat{M}^3 a^2\delta_{ij}(\partial_0 + 3H)\dot{\pi}\,. \end{aligned} \tag{3.40}$$

Let us compare the unitary gauge approach with the procedure presented in the previous section. In sec. 3.1, we started from the action of k-essence, eq. (3.6), and we added the two ϕ-dependent higher derivative operators $-\bar{M}^2(\phi)[\Box\phi + 3H(\phi)]^2/2$ and $-\hat{M}^3(\phi)[\Box\phi + 3H(\phi)](X-1)/2$, that do not change the background equations of motion. The action (3.37) is constructed similarly. First of all, note that the first line of eq. (3.37) is the action for k-essence. Indeed, it is equivalent to the

action (3.12), which was found by expanding the k-essence action (3.1) in terms of π. The second line of eq. (3.37) can be constructed by noting that the extrinsic curvature of the hypersurfaces of constant ϕ, defined as

$$K^{\mu}{}_{\nu} \equiv -(g^{\rho}{}_{\nu} + u^{\rho}u_{\nu})\nabla_{\rho}u^{\mu} \,, \tag{3.41}$$

where $u^{\mu} \equiv -\partial^{m}u\phi/\sqrt{X}$ is the unit vector orthogonal to ϕ, can be rewritten as

$$K^{\mu}{}_{\nu} = \frac{1}{\sqrt{-(\partial\phi)^2}}\left[\nabla^{\mu}\partial_{\nu}\phi + \frac{\partial^{\mu}\phi\partial^{\rho}\phi}{-(\partial\phi)^2}\nabla_{\rho}\partial_{\nu}\phi\right] . \tag{3.42}$$

When expanded around the background solution $\phi = t$, then $\delta K \equiv K^{\mu}_{\mu} - 3H(\phi)$ reads[5]

$$\delta K = -3\dot{H}\pi - \frac{\nabla^2}{a^2}\pi + \frac{1}{2}\dot{h} \,. \tag{3.43}$$

Thus the action (3.37) can be constructed by simply adding to the k-essence action the two operators $-\bar{M}^2(\phi)\delta K^2/2$ and $-\hat{M}^3(\phi)\delta K(g^{00}+1)/2$, that do not change the background solution.

3.3 Phenomenology on the quintessential plane

3.3.1 *k*-essence vs. Ghost Condensate

Coming back to the quintessential plane of figure 3.1, in section 3.1.2 we have learned an important lesson: *the gradient instabilities for $w_Q < -1$ can be made harmless by higher derivative operators.* Thus, part of the lower left quadrant of the quintessential plane is allowed.

To discuss the phenomenology of these models (for a related discussion see [52]), let us write the full action for perturbations including the higher derivative operator (3.16):

$$S = \int \mathrm{d}^4x\, a^3\Big[\frac{1}{2}\left(\rho_Q + p_Q + 4M^4\right)\dot{\pi}^2 - \frac{1}{2}(\rho_Q + p_Q)\frac{(\nabla\pi)^2}{a^2} + \frac{3}{2}\dot{H}(\rho_Q + p_Q)\pi^2 - \frac{1}{2}(\rho_Q + p_Q)\dot{h}\pi \\ - \frac{\bar{M}^2}{2}\left(3H\dot{\pi} - 3\dot{H}\pi + \frac{\dot{h}}{2} - \frac{\nabla^2\pi}{a^2}\right)^2\Big] \,. \tag{3.44}$$

First of all, note that it is not possible to switch off quintessence perturbations for $\rho_Q + p_Q \neq 0$; doing it by hand would give gauge dependent unphysical results. On the other hand, the converse is not true: even for $\rho_Q + p_Q = 0$ perturbations may still be present, as in the Ghost Condensate case.

[5] One may wonder why terms of the form $\dot{\pi}\nabla^2\pi$ do not appear in δK^2, while they do appear in $(\Box\phi + 3H)^2$. As seen in eq. (3.42), a time diffeomorphism does not change the extrinsic curvature of constant ϕ hypersurfaces. Thus, δK does not contain $\dot{\pi}$ and $\dot{\pi}\nabla^2\pi$ is not generated by δK^2.

3.3 Phenomenology on the quintessential plane

We saw that the operator in the second line of eq. (3.44) allows the stabilization of the short scale gradient instability; on the other hand, for cosmological purposes we are interested in very large scales. Let us see whether this operator is relevant for scales of the order of the Hubble radius (although our discussion will extend to all scales of cosmological interest). We want to show that, when

$$|\rho_Q + p_Q| \gg \bar{M}^2 H^2 \,, \tag{3.45}$$

the higher derivative operator can be neglected when discussing the cosmological clustering of quintessence. In this case we reduce to a standard k-essence model, with the only difference that there are no short-scale instabilities even for $w_Q < -1$. On the other hand, in the opposite case

$$|\rho_Q + p_Q| \ll \bar{M}^2 H^2 \,, \tag{3.46}$$

all the terms in the action (3.44) proportional to $\rho_Q + p_Q$ can be neglected. In this case the model reduces to the Ghost Condensate theory.

Verifying the existence of these two regimes is quite straightforward. For instance, the dispersion relation at $k/a \sim H$ is dominated either by $(\rho_Q + p_Q)(\nabla\pi)^2$ or by $\bar{M}^2(\nabla^2\pi)^2$, depending on the hierarchy between $|\rho_Q + p_Q|$ and $\bar{M}^2 H^2$. The same applies for the operators involving the metric perturbation, $(\rho_Q + p_Q)\dot{h}\pi$ and $\bar{M}^2\dot{h}\nabla^2\pi$. This check can be done for all the other operators, by taking $\nabla/a \sim H$ and considering that time derivatives are at most of order H. The existence of these two regimes can also be seen by looking at the stress-energy tensor.

Now we can go back and complete our quintessential plane. When w_Q is close to -1,

$$-\frac{\bar{M}^2}{M_{\rm Pl}^2} \lesssim (1 + w_Q)\Omega_Q \lesssim \frac{\bar{M}^2}{M_{\rm Pl}^2} \,, \tag{3.47}$$

the model behaves as the Ghost Condensate. We can estimate the width of this region by imposing the absence of Jeans instability, eq. (3.28). Assuming $M \sim \bar{M}$ one gets a rough upper bound: $\bar{M} \lesssim 10$ MeV [28]. A more accurate analysis shows that this limit is very conservative and it can be relaxed to $\bar{M} \lesssim 100$ GeV [50]. Even in this case the window above is extremely tiny:

$$|1 + w_Q|\Omega_Q \lesssim 10^{-34} \,. \tag{3.48}$$

We can therefore draw an important conclusion: only for values of w_Q which are observationally indistinguishable from the cosmological constant, does quintessence behave as the Ghost Condensate on cosmological scales. This regime corresponds to the strip around the horizontal axis $w_Q = -1$

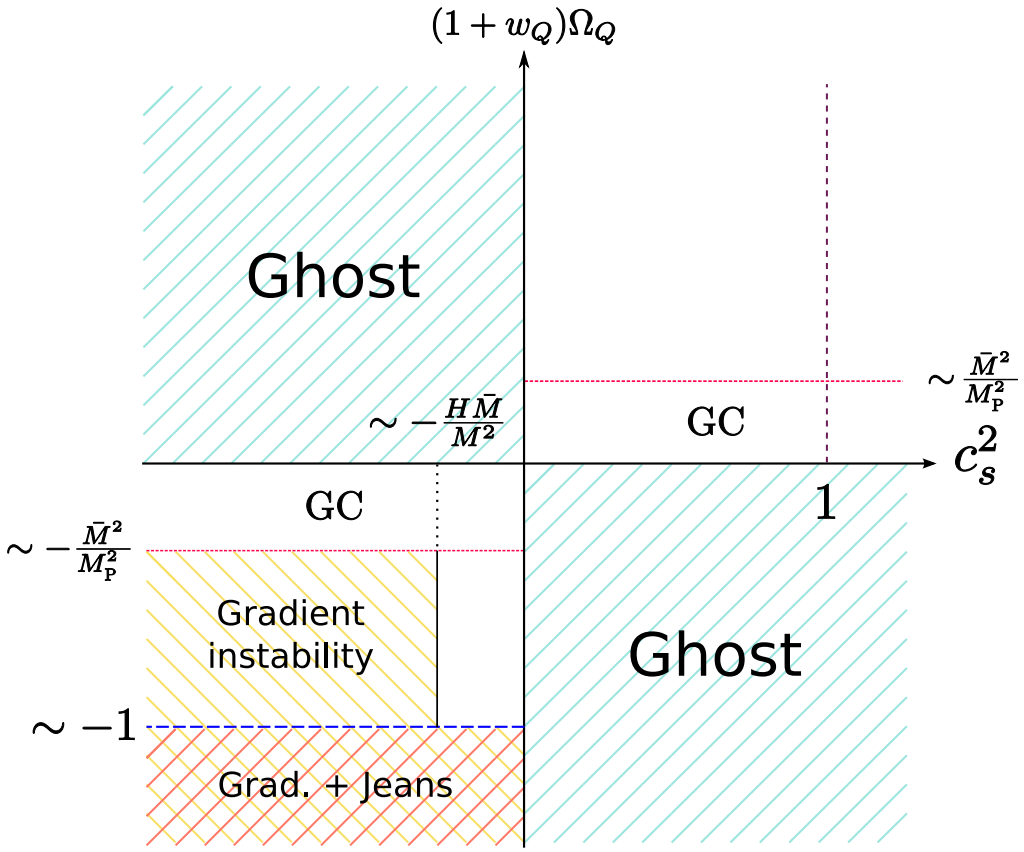


Figure 3.2: On the quintessential plane, we show the theoretical constraints on the equation of state and speed of sound of quintessence, in the presence of the operator $\bar{M}$. Instability regions are dashed. Where $1+w_Q$ and c_s^2 have opposite sign we have a ghost-like instability corresponding to negative kinetic energy. For $w_Q < -1$, the dashed regions in the left-lower panel is unstable by gradient ($c_s^2 \lesssim -H\bar{M}/M^2$) and Jeans ($(1+w_Q)\Omega_Q \lesssim -1$) instabilities, while the strip close to the vertical axis corresponds to the stability window (3.29). Furthermore, the strip around the horizontal axis given in eq. (3.47) corresponds to the Ghost Condensate. Above this region, in the right-upper panel, we find standard k-essence.

in figure 3.2. Notice that in this region the dispersion relation is of the form $\omega \propto k^2$, so that the speed of sound c_s^2 is not well defined, i.e. it becomes scale dependent.

On the other hand, for any value of w_Q which is appreciably different from the one of the cosmological constant, the model reduces to k-essence, as higher derivative terms are cosmologically irrelevant. Their only role is to stabilize the short scale gradient instabilities for $w_Q < -1$. Although in practice not relevant, note however that w_Q cannot be made arbitrarily negative. This is shown by eq. (3.29) and, in the quintessential plane, it excludes the bottom shaded region of the lower-left quadrant.

Let us now constrain the values of the speed of sound c_s^2. For $w_Q > -1$ there are no constraints, besides the possible limit $c_s^2 \leq 1$ already discussed. For $w_Q < -1$ the speed of sound is negative and very small, as it is constrained by the absence of gradient instability, eq. (3.20):

$$-c_s^2 \simeq -\frac{\rho_Q + p_Q}{4M^4} \lesssim \frac{H\bar{M}}{M^2}\,. \tag{3.49}$$

We can numerically constrain the right-hand side of this equation by considering that the scales $M \sim \bar{M}$ represent the cutoff of our effective field theory. By requiring this cutoff to be larger than the minimum scale at which gravity has been probed, i.e., $M \gtrsim 10^{-3}$eV, and using in eq. (3.49) the value of the Hubble parameter today, $H_0 \sim (10^{-3}\text{eV})^2/M_{\rm Pl}$, we obtain

$$-c_s^2 \lesssim \left(\frac{H_0}{M_{\rm Pl}}\right)^{1/2} \sim 10^{-30}\,. \tag{3.50}$$

Thus, for all practical purposes, the speed of sound can be taken to be exactly zero. On the quintessential plane in figure 3.2, in the lower-left quadrant, we can only live in a tiny strip along the vertical axis. Notice however that there is no fine tuning in keeping c_s^2 extremely small. Indeed, as we discussed, in the limit of Ghost Condensate c_s^2 vanishes exactly for symmetry reasons. Thus, the speed of sound remains small for tiny deviations from this limit.

3.3.2 Including dark matter

After the discussion about the stability constraints, we would like to understand the dynamics of quintessence perturbations and their impact on cosmological observations. In order to do this, we will now study quintessence in the presence of cold dark matter, which gravitationally sources quintessence perturbations, and we will focus on the main qualitative features in the various limits.

Let us start from the Ghost Condensate limit (3.46). It is known that the Ghost Condensate affects only short scales, i.e., π perturbations induce a modification of the Newtonian potential at scales parametrically smaller than the Hubble scale [28]. Therefore, we expect to have extremely small effects on cosmological scales. To verify that this is the case, we can study the action (3.44) in the limit of $\rho_Q + p_Q = 0$. This reads

$$S = \int \mathrm{d}^4x\, a^3 \left[2M^4\dot{\pi}^2 - \frac{\bar{M}^2}{2}\left(3H\dot{\pi} - 3\dot{H}\pi + \frac{\dot{h}}{2} - \frac{\nabla^2\pi}{a^2}\right)^2\right]\,. \tag{3.51}$$

For simplicity, let us momentarily disregard the first two terms in parentheses:

$$S = \int \mathrm{d}^4x\, a^3 \left[2M^4\dot{\pi}^2 - \frac{\bar{M}^2}{2}\left(\frac{\dot{h}}{2} - \frac{\nabla^2\pi}{a^2}\right)^2\right]\,. \tag{3.52}$$

Notice that this is the action used in the Ghost Condensate paper [28]. The equation of motion for the π perturbations is given by

$$\ddot{\pi} + 3H\dot{\pi} + \frac{\bar{M}^2}{4M^4}\frac{\nabla^4\pi}{a^4} = \frac{\bar{M}^2}{8M^4}\frac{\nabla^2\dot{h}}{a^2}\,. \tag{3.53}$$

The gradient term on the left hand side can be neglected on cosmological scales. Indeed, the time derivatives will be at least of order $\sim \bar{M}k^2/(aM)^2$, so that the friction term in the previous equation will always dominate the gradient term for $k/a \sim H$. As we want to show that Ghost Condensate perturbations remain small, we assume that the dark matter dominates the perturbed Einstein equations. The validity of this assumption can be checked *a posteriori*.

In a matter dominated Universe, $\dot{h} = -2\dot{\delta}_m$ [4], with $\delta_m \equiv \delta\rho_m/\rho_m$. Using $\dot{\delta}_m = H\delta_m$ and the background Friedmann equation, we get

$$\dot{h} = -\frac{2}{3H}\frac{\delta\rho_m}{M_{\rm Pl}^2}\,. \tag{3.54}$$

We can now replace this as the source of Ghost Condensate perturbations on the right-hand side of eq. (3.53). This yields, neglecting the gradient term:

$$\ddot{\pi} + 3H\dot{\pi} \simeq -\frac{\bar{M}^2}{12M^4M_{\rm Pl}^2}\frac{\nabla^2\delta\rho_m}{a^2H}\,. \tag{3.55}$$

If we now assume that the initial quintessence perturbations are small so that the homogeneous solutions are sub-dominant, similarly to what happens in standard quintessence [53], this equation can be solved to give

$$\dot{\pi} = -\frac{\bar{M}^2}{24M^4M_{\rm Pl}^2}\frac{\nabla^2\delta\rho_m}{a^2H^2}\,. \tag{3.56}$$

As we discussed in the previous section, the energy density and pressure perturbations of the Ghost Condensate are dominated by the M^4 operator so that $\delta\rho_Q \simeq 4M^4\dot{\pi}$ and $\delta p_Q \simeq 0$. Thus, on cosmological scales,

$$\delta\rho_Q \sim \frac{\bar{M}^2}{M_{\rm Pl}^2}\delta\rho_m\,. \tag{3.57}$$

From the simple estimate of $\bar{M}$ below eq. (3.47), we conclude that quintessence perturbations are negligibly small with respect to dark matter perturbations, $\delta\rho_Q \lesssim 10^{-34}\delta\rho_m$. It is straightforward to generalize this analysis including the two terms in parentheses of eq. (3.51) previously neglected and verify that eq. (3.57) remains valid. The conclusion of eq. (3.57) is quantitatively consistent with the (small) modification of the Newton law derived in [28], as one can check for example in their eq. (7.11) for $k/a \sim H$ and $\omega \sim H$.

3.3 Phenomenology on the quintessential plane

Close to the $w_Q = -1$ line, we have just seen that there are no appreciable effects of perturbations on cosmological scales which can help in distinguishing quintessence from a cosmological constant; all the interesting dynamics is limited to modifications of gravity on short scales. As we move away from the $w_Q = -1$ line, we enter in the k-essence regime, as we pointed out in the previous section. The case $w_Q > -1$ is well studied in the literature (see, for instance, [54, 55]). The case $w_Q < -1$ is much less studied: here we have a negative speed of sound squared that is so small – see eq. (3.50) – that can be taken to be zero for all practical purposes. With a small speed of sound we expect quintessence to cluster on scales shorter than the Hubble radius, driven by dark matter gravitational potential wells. To study this, let us repeat the calculation we just did in the Ghost Condensate case for a k-essence with $c_s^2 = 0$.

For simplicity, let us assume for the moment that M is constant. Varying the action (3.12), we get the equation of motion for π:

$$4M^4(\ddot{\pi} + 3H\dot{\pi}) - (\rho_Q + p_Q)\frac{\nabla^2}{a^2}\pi - 3\dot{H}(\rho_Q + p_Q)\pi = -\frac{1}{2}(\rho_Q + p_Q)\dot{h} \ . \tag{3.58}$$

Small $|c_s^2|$ is equivalent to $|\rho_Q + p_Q| \ll M^4$ so that the gradient and mass term in this equation can be neglected. Using again $\dot{h} = -2\dot{\delta}_m$ [4], we thus have

$$4M^4(\ddot{\pi} + 3H\dot{\pi}) = (\rho_Q + p_Q)\dot{\delta}_m \ . \tag{3.59}$$

One can verify that, neglecting decaying modes, the solution of this equation is

$$\delta_Q = \frac{1 + w_Q}{1 - 3w_Q}\delta_m \ . \tag{3.60}$$

It is easy to show that this equation holds, for constant w_Q, for a general time dependent speed of sound which satisfies $|c_s^2(t)| \ll 1$.

Equation (3.60) describes quintessence perturbations both for positive and negative $1 + w_Q$. When $w_Q > -1$, quintessence energy density clusters in the dark matter potential wells, while in the opposite case $w_Q < -1$ it escapes from them [56]. However, clustering of quintessence remains small compared to dark matter as the coupling with gravity is suppressed by $1 + w_Q$. For very small values $|1 + w_Q| \sim \bar{M}^2/M_{\rm Pl}^2$ we smoothly enter in the Ghost Condensate regime. Indeed, it is easy to see that eq. (3.60) smoothly matches eq. (3.57) in the intermediate regime.

In this section we have studied the phenomenology of quintessence in various regimes of $1+w_Q$. Quintessence perturbations smoothly turn off when we approach the cosmological constant limit $w_Q = -1$ from both sides. This suggests that in general there is no pathology in crossing the $w_Q = -1$ line, as we discuss in the next section.

3.4 Crossing the phantom divide

It has been claimed that, during its evolution, single field quintessence cannot cross the $w_Q = -1$ line as perturbations become pathological. For this reason this line has been dubbed "phantom divide" [36]. However, there is no real pathology in crossing this line, besides the fact that, for $w_Q < -1$, short-scale gradient instabilities must be stabilized [37, 38]. If one does not take into account higher derivative terms, a negative c_s^2 leads to catastrophic instabilities at short scales. Once instabilities are cured as we discussed in the previous sections, crossing the phantom divide becomes trivial[6].

Indeed, the Lagrangian (3.13) gives an explicit way to construct a model which crosses the phantom divide. If one assumes for simplicity that quintessence is the only component in the Universe[7], the crossing of the phantom divide corresponds to a change of sign of $\dot{H}$. In particular, considering only quintessence, one can use Friedmann equations to recast the Lagrangian (3.13) in the form (higher derivative operators will be considered later)

$$P(X,\phi) = -3M_{\rm Pl}^2 H^2(\phi) - M_{\rm Pl}^2 \dot{H}(\phi)(X+1) + \frac{1}{2}M^4(\phi)(X-1)^2 \,. \tag{3.61}$$

This is similar to what happens in inflation, where the inflaton is the only relevant component in the Universe [33].

As an example, we consider the case where $\dot{H}$ evolves linearly in time and changes sign from negative to positive:

$$\dot{H}(t) = \frac{\mu^4}{M_{\rm Pl}^2}(mt-1) \,. \tag{3.62}$$

This implies that $H(t)$ will be a parabola of the form

$$H(t) = \frac{\mu^4}{M_{\rm Pl}^2}\left(\frac{m}{2}t^2 - t\right) + H_* \,, \tag{3.63}$$

as shown in figure 3.3 (left panel). Using the general expression (3.61), we deduce that the Lagrangian

$$P(\phi,X) = -3\left[\frac{\mu^4}{M_{\rm Pl}}\left(\frac{m}{2}\phi^2 - \phi\right) + M_{\rm Pl}H_*\right]^2 + \mu^4(m\phi-1)\left[(\partial\phi)^2 - 1\right] + \frac{1}{2}M^4(\phi)\left[(\partial\phi)^2+1\right]^2 \tag{3.64}$$

[6] Models that cross the phantom divide have been found in $f(R)$ theories of gravity. However, for $w_Q < -1$ these models are equivalent to scalar fields with negative kinetic energy, i.e. to ghosts. Thus, according to our stability constraints, they are forbidden.

[7] The conclusions drawn in this section also hold when including dark matter.

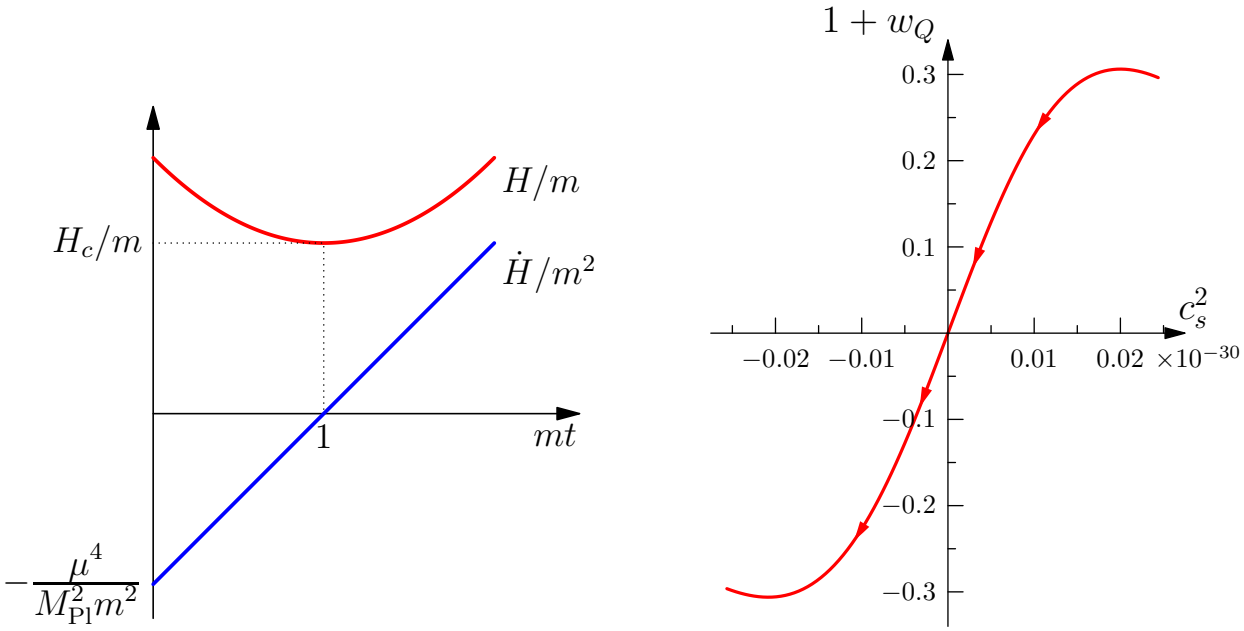


Figure 3.3: Example of phantom divide crossing, as given by eqs. (3.62) and (3.63), where we have defined $H_c \equiv H_* - \mu^4/(2mM_{\rm Pl}^2)$. Left figure: behavior of H and $\dot{H}$; the crossing of $w_Q = -1$ takes place at $t = m^{-1}$ when $\dot{H} = 0$ and $H = H_c$. Right figure: trajectory on the quintessential plane.

admits the background solution $\phi = t$ and the cosmological evolution (3.62) and (3.63). Note that there are no theoretical limitations on the choice of the background evolution $H(t)$. Indeed, we can cross the phantom divide as many times as we want. For example, choosing $\dot{H}(t) \propto \sin(mt)$ the cosmological evolution keeps oscillating up and down around $w_Q = -1$!

We can now study perturbations around a solution crossing $w_Q = -1$ to show that no pathology arises. The evolution equation derived from the action (3.12) reads

$$\left(\rho_Q + p_Q + 4M^4\right)\ddot{\pi} + \frac{1}{a^3}\partial_t\left[a^3(\rho_Q + p_Q + 4M^4)\right]\dot{\pi} - 3\dot{H}(\rho_Q + p_Q)\pi - (\rho_Q + p_Q)\frac{\nabla^2\pi}{a^2} = -\frac{1}{2}(\rho_Q + p_Q)\dot{h}\,. \tag{3.65}$$

At the phantom divide, the last three terms of this equation vanish but the equation is clearly non-singular. In our approach, it is manifest that the speed of sound squared changes sign at $w_Q = -1$. In the example above c_s^2 is given by

$$c_s^2 = \frac{\mu^4(1-mt)}{\mu^4(1-mt) + 2M^4}\,, \tag{3.66}$$

and the trajectory of the crossing on the quintessential plane is shown in figure 3.3 (right panel). The stability in the $w_Q < -1$ region requires that $|c_s^2|$ remains extremely small so that it is mandatory

to have a hierarchy between μ and M, $\mu \ll M$. As we discussed, this hierarchy is naturally realized.

Another approach to study perturbations is using a fluid description, as k-essence is equivalent to a perfect fluid. In synchronous gauge, denoting with a prime the derivative with respect to the conformal time $d\eta \equiv dt/a$ and defining $\mathcal{H} = aH$, the fluid equations read, in Fourier space (see for example [57]):

$$\delta_Q' + 3\mathcal{H}(c_s^2 - w_Q)\delta_Q = -(1+w_Q)\left[k^2 + 9\mathcal{H}^2(c_s^2 - c_a^2)\right]\frac{\vartheta_Q}{k^2} - (1+w_Q)\frac{h'}{2}\,, \tag{3.67}$$

$$\frac{\vartheta_Q'}{k^2} + \mathcal{H}(1-3c_s^2)\frac{\vartheta_Q}{k^2} = \frac{c_s^2}{1+w_Q}\delta_Q\,, \tag{3.68}$$

where θ_Q is the divergence of the velocity field of quintessence, $\theta_Q \equiv ik^i T^0_i/(\rho_Q + p_Q)$ [4], c_s^2 is $\delta p_Q/\delta\rho_Q$ calculated in a velocity orthogonal gauge ($T^0_i = 0$) [51] and it corresponds to the speed of sound squared that can be deduced from the π Lagrangian[8], to be distinguished from the adiabatic speed of sound squared, $c_a^2 \equiv \dot{p}_Q/\dot{\rho}_Q = w_Q - \dot{w}_Q/(3H(1+w_Q))$ [58]. The absence of pathologies at $w_Q \to -1$ can also be shown in this formalism. Indeed, in the continuity equation the divergence of c_a^2 is compensated by the prefactor in front of the squared brackets, while the $1+w_Q$ term at the denominator in the Euler equation is harmless as c_s^2 also vanishes for $w_Q \to -1$. Thus both δ_Q and θ_Q are continuous through the divide. This is not surprising as θ_Q is just the Laplacian of the scalar perturbation π, $\pi = a\,\theta_Q/k^2$.

At this point the reader may be puzzled: in the previous sections we stressed that close to the $w_Q = -1$ line quintessence behaves like the Ghost Condensate on cosmological scales, while eq. (3.65) as well as the fluid eqs. (3.67) and (3.68) do not contain higher derivative terms. Let us see why these additional terms are irrelevant in realistic cases of phantom divide crossing. With these new terms, the equation of motion for π derived from the full action (3.44) is obviously still continuous so that also in this case the line $w_Q = -1$ can be crossed smoothly. The operator proportional to $\bar{M}^2$ dominates in the Ghost Condensate strip around the $w_Q = -1$ line. However, this happens only in the extremely narrow range $|1+w_Q| \lesssim \bar{M}^2/M_{\rm Pl}^2 \lesssim 10^{-34}$. The equation of state parameter w_Q will stay in this range only for a time much smaller than H^{-1}, unless its evolution is tremendously slow. Thus π has no time to evolve in the Ghost Condensate regime, so that for all practical purposes one can totally neglect this strip around the $w_Q = -1$ line on cosmological scales.

[8]The velocity orthogonal condition $T^0_i = 0$ is equivalent to the condition $\pi = 0$. As ϕ is unperturbed in this gauge, the perturbations of pressure and energy density only come from fluctuations of X, i.e. $\delta p_Q/\delta\rho_Q = p_{Q,X}/\rho_{Q,X} = P_X/(2P_{XX}X + P_X)$, which is the speed of sound c_s^2 which appears in the π Lagrangian (3.6).

We have seen that k-essence can be described with the fluid equations (3.67) and (3.68). Even including higher derivative terms, quintessence remains a perfect fluid, but does not satisfy the fluid equations (3.67) and (3.68) as these assume a linear dispersion relation. However, as we discussed, higher derivative terms are phenomenologically irrelevant on cosmological scales, so that one can still use the fluid description above when comparing with observations.

From a practical point of view we conclude that, when comparing with observations a dark energy model which crosses the phantom divide, it is consistent and theoretically motivated to set $c_s^2 = 0$. On the other hand, it is inconsistent to turn off perturbations as sometimes done in the literature.

3.5 Additional higher derivative operators

As we discussed, higher derivative operators become relevant when the speed of sound is very close to zero. This regime is particularly interesting when $w_Q < -1$ so that in the following we will consider mostly the case $\rho_Q + p_Q < 0$.

Theories with very small c_s^2 should be thought of as tiny deformations of the Ghost Condensate theory [32, 33, 49]: in this limit one recovers the shift symmetry $\phi \to \phi + \lambda$, so that a small deviation from the Ghost Condensate is technically natural. In the Ghost Condensate limit there is an additional symmetry that one can impose, i.e. the parity symmetry $\phi \to -\phi$. The background $\phi = t$ in Minkowski space breaks this parity symmetry and the time reversal symmetry to the composition of the two; the theory of perturbations is thus invariant under $\pi \to -\pi$, $t \to -t$ [28]. This symmetry is present only when the background metric is Minkowski: in de Sitter there is a preferred time direction singled out by the expansion. In this case, terms violating the symmetry will be proportional to H, and thus tipically suppressed by H/M. In this paper we have considered small departures from the Ghost Condensate limit, i.e., tiny breakings of the shift symmetry. These also generate terms which are not invariant under parity $\pi \to -\pi$, $t \to -t$, as for instance the mass term in eq. (3.12). These terms will be of the same order of magnitude, i.e. suppressed by H/M, as we are assuming that H^{-1} is the typical time scale of evolution of the operators.

However, one can also consider the case when the parity symmetry $\phi \to -\phi$ is absent in the Ghost Condensate limit [32]. This happens for example if we add to the k-essence Lagrangian (3.13) the operator

$$\mathcal{L}_{\hat{M}} = -\frac{\hat{M}^3}{2}(\Box\phi + 3H)(X - 1) \,, \tag{3.69}$$

which again does not change the background evolution as it starts quadratic in the perturbations. For simplicity we assume that $\bar{M} = 0$ and a constant $\hat{M}$[9]. In synchronous gauge at quadratic order this operator is

$$\mathcal{L}_{\hat{M}} = \hat{M}^3 \dot{\pi} \left(\ddot{\pi} + 3H\dot{\pi} - 3\dot{H}\pi - \frac{\nabla^2 \pi}{a^2} + \frac{\dot{h}}{2} \right) . \tag{3.70}$$

The first two terms in the parentheses contribute (after an integration by parts) to the time kinetic term. Assuming $\hat{M} \sim M$ they can be neglected in comparison with $2M^4\dot{\pi}^2$. The third term gives a mass term that is parametrically smaller than H and can thus be neglected.

To discuss the stability and phenomenology of this model, let us write the full action for perturbations, assuming $|\rho_Q + p_Q| \ll M^4$:

$$S = \int \mathrm{d}^4x\, a^3 \left[2M^4\dot{\pi}^2 - \frac{1}{2}(\rho_Q + p_Q)\frac{(\nabla\pi)^2}{a^2} - \frac{1}{2}(\rho_Q + p_Q)\dot{h}\pi + \hat{M}^3\dot{\pi}\left(\frac{\dot{h}}{2} - \frac{\nabla^2\pi}{a^2} \right) \right] , \tag{3.71}$$

where we have neglected the mass terms. This equation is analogous to eq. (3.44) for the $\bar{M}$ operator (3.16). Analogously to what we have done in sections 3.1.2 and 3.3.2 for the operator proportional to $\bar{M}^2$, we will now study the stability and phenomenology on cosmological scales with the operator $\mathcal{L}_{\hat{M}}$. In appendix B we briefly study the effect of this operator at short distances, i.e. the modification of gravity induced by it.

3.5.1 Stability constraints with $\hat{M}$

Let us first study the stability of the system neglecting other sources of gravity. The equation of motion for π derived varying (3.71) reads

$$\ddot{\pi} + 3H\dot{\pi} - \frac{\rho_Q + p_Q}{4M^4}\frac{\nabla^2\pi}{a^2} - \frac{\hat{M}^3 H}{4M^4}\frac{\nabla^2\pi}{a^2} = -\frac{\rho_Q + p_Q}{8M^4}\dot{h} - \frac{\hat{M}^3}{8M^4}(\ddot{h} + 3H\dot{h}) . \tag{3.72}$$

Notice that the operator $\hat{M}^3\dot{\pi}\nabla^2\pi/a^2$ induces a spatial kinetic term for π proportional to H. Indeed, this operator is a total derivative in Minkowski spacetime. Choosing $\hat{M} > 0$, the spatial kinetic term has the "healthy" sign and can be chosen sufficiently large to cure the gradient instability associated to $\rho_Q + p_Q < 0$, giving a positive and very small c_s^2. This also allows us to neglect the first term on the right hand side in eq. (3.72). To complete the stability analysis one has to take into account the mixing with gravity, i.e., solve for h in terms of the quintessence stress-energy tensor using the Einstein equation (3.23), and plug the result back in the right hand side of eq. (3.72). This, similarly to what happens for the Ghost Condensate, will give rise to a Jeans-like instability.

[9]This assumption can be relaxed by having a time dependence with time scale of order H^{-1}, as in the case of $\bar{M}^2$.

The contribution to the stress-energy tensor of the operator (3.69) is

$$\delta\rho_Q \supset \hat{M}^3 \left(\frac{\dot{h}}{2} - \frac{\nabla^2 \pi}{a^2} \right) , \tag{3.73}$$

$$\delta p_Q \supset -2\hat{M}^3 \left(\ddot{\pi} + 3H\dot{\pi} \right) . \tag{3.74}$$

Given the small speed of sound, time derivatives are much smaller than the spatial ones and the pressure perturbation is negligible, $\delta p_Q \ll \delta\rho_Q$. Concentrating on frequencies much larger than the Hubble rate one can neglect the terms containing $H\dot{\pi}$ and $H\dot{h}$ in eq. (3.72). A further simplification comes from disregarding the standard k-essence contribution to the energy density perturbation $\delta\rho_Q$, i.e. $4M^4\dot{\pi}$, in comparison with $\hat{M}^3\nabla^2\pi/a^2$. Indeed, from eq. (3.72) we have $\hat{M}^3\nabla^2\pi/a^2 \sim M^4\ddot{\pi}/H \gg M^4\dot{\pi}$. Moreover, as we will see, the absence of Jeans instability will impose $\hat{M}^3 \lesssim M_{\rm Pl}^2 H$. This implies that the term with $\dot{h}$ in eq. (3.73) is negligible with respect to $\ddot{h}$ in the Einstein equation (3.23), which becomes

$$\ddot{h} = \frac{\hat{M}^3}{M_{\rm Pl}^2} \frac{\nabla^2\pi}{a^2} . \tag{3.75}$$

Plugging this into the right hand side of eq. (3.72) we finally find

$$\ddot{\pi} - \left(\frac{\rho_Q + p_Q}{4M^4} + \frac{\hat{M}^3 H}{4M^4} - \frac{\hat{M}^6}{8M^4 M_{\rm Pl}^2} \right) \frac{\nabla^2\pi}{a^2} = 0 . \tag{3.76}$$

This same result would have been found using a more rigorous approach, as done in [32]. Again, as in the Ghost Condensate case, mixing with gravity induces a Jeans-like instability, represented by the last term in this equation. Thus, for $\rho_Q + p_Q < 0$ we need to cure both the gradient and the Jeans instabilities. This is possible for

$$-(1+w_Q)\Omega_Q \lesssim \frac{\hat{M}^3}{M_{\rm Pl}^2 H} \lesssim 1 . \tag{3.77}$$

This stability window [32] is analogous to the one discussed in the Ghost Condensate case, eq. (3.29).

We conclude that, with the inclusion of the operator $\mathcal{L}_{\hat{M}}$, we can have a dispersion relation $\omega \propto k$ with positive speed of sound squared; thus, there is no sign of instability even for $\rho_Q + p_Q < 0$.

3.5.2 Including dark matter

Analogously to what done in section 3.3.2, to study the phenomenology induced by the $\hat{M}$ operator we study quintessence perturbations generated by the coupling with dark matter. For simplicity

we assume matter dominance. The action (3.71) gives the equation of motion for π sourced by the dark matter perturbation δ_m:

$$\ddot{\pi} + 3H\dot{\pi} - \frac{H\hat{M}^3}{4M^4}\frac{\nabla^2\pi}{a^2} = \frac{5}{8}\frac{\hat{M}^3}{M^4}H^2\delta_m + \frac{\rho_Q + p_Q}{4M^4}H\delta_m \ , \tag{3.78}$$

where we have used $\dot{h} = -2H\delta_m$ and neglected the gradient term proportional to $\rho_Q + p_Q$ that is subdominant. Since the speed of sound $c_s^2 = H\hat{M}^3/(4M^4)$ is very small and we are interested in cosmological scales, one would naïvely neglect the term with $\nabla^2\pi$. This gives the solution

$$\dot{\pi} = \frac{H\hat{M}^3}{4M^4}\delta_m + \frac{\rho_Q}{4M^4}\frac{1+w_Q}{1-3w_Q}\delta_m \ . \tag{3.79}$$

However, the approximation of neglecting the gradients is not good. Indeed, when we plug this expression into the energy density perturbation

$$\delta\rho_Q = 4M^4\dot{\pi} + \hat{M}^3\left(\frac{\dot{h}}{2} - \frac{\nabla^2\pi}{a^2}\right) \ , \tag{3.80}$$

there is a cancellation of terms proportional to $\hat{M}^3$ up to gradient terms. Thus one is forced to go back to the equation of motion (3.78) and keep the term proportional to the speed of sound.

Once we do that, we obtain

$$\delta_Q = -\frac{\hat{M}^6}{24M^4M_{\rm Pl}^2\Omega_Q}\frac{\nabla^2}{H^2a^2}\delta_m + \frac{1+w_Q}{1-3w_Q}\delta_m \ . \tag{3.81}$$

This equation displays the existence of two different regimes, in strict analogy with what happens in the Ghost Condensate case. For large enough $|1+w_Q|$, the second term on the right hand side dominates and one recovers eq. (3.60), where the system behaves as standard k-essence. Note however that, even for $\rho_Q + p_Q < 0$, there are no stability problems in the stability window (3.77). The dynamics of the system is dominated by the $\hat{M}$ operator only when we are very close to $w_Q = -1$, i.e. for

$$-\frac{\hat{M}^6}{M^4M_{\rm Pl}^2} \lesssim (1+w_Q)\Omega_Q \lesssim \frac{\hat{M}^6}{M^4M_{\rm Pl}^2} \ . \tag{3.82}$$

This region is the analogue of the Ghost Condensate strip around the horizontal axis of figure 3.2. In this range the first term in eq. (3.81) dominates and δ_Q remains extremely small with respect to δ_m. Although the dispersion relation is of the form $\omega \propto c_s k$, quintessence does not follow the simple fluid equations (3.67) and (3.68) because of the presence of higher derivative operators. However, as in the Ghost Condensate case, given the narrowness of the strip (3.82), for all practical purposes we can always use the fluid equations with $c_s^2 = 0$, even when crossing the phantom divide. In

conclusion, the addition of the operator $\hat{M}$ can stabilize k-essence in the phantom case $w_Q < -1$, and the phenomenology of the model is the same as for a k-essence with $c_s^2 = 0$. This general conclusion will hold even when considering both operators $\bar{M}$ and $\hat{M}$ at the same time, and can be extended to all the possible higher derivative operators included in the general action in section 3.2.

3.6 Conclusions

In this chapter we have studied the most generic action describing the perturbations of a single field dark energy – here called quintessence – around a given background. We have constructed the action by adding to the k-essence Lagrangian higher derivative operators that leave the background evolution invariant. Using this action, we have derived the theoretical constraints on the equation of state parameter w_Q as a function of the speed of sound squared c_s^2, by the requirement that perturbations are ghost-free – i.e. that their kinetic energy is positive – and that there are no gradient-like instabilities. These constraints have been conveniently represented on the quintessential plane $(1 + w_Q)\Omega_Q$ vs. c_s^2, in figures 3.1 and 3.2.

In particular, we have considered the case $w_Q < -1$, which is commonly believed to be unstable, and we have shown that for very small c_s^2 both the gradient and the Jeans instabilities can be avoided and perturbations stabilized [32]. Higher derivative operators are crucial for the stabilization. Indeed, it is important to stress that taking an extremely small c_s^2 does not represent a fine tuning, as in the limit $c_s^2 \to 0$ we recover the Ghost Condensate theory which is protected by the shift symmetry $\phi \to \phi + c$. Thus, for $w_Q < -1$, quintessence should be thought of as a small deformation of the Ghost Condensate limit [32, 33, 49]. When the higher order terms containing k^4 dominate over the spatial kinetic term $c_s^2 k^2$, the phenomenology reduces to that of the Ghost Condensate. This always happens on small scales, where the higher order gradients must dominate to stabilize the perturbations, but on cosmological scales this only occurs for values of w_Q extremely close to the one of the cosmological constant, i.e. for $|1 + w_Q|\Omega_Q \lesssim 10^{-34}$. Away from this tiny strip – i.e. for all practical purposes – the behavior on cosmological scales is very different from that of the Ghost Condensate: higher derivative terms are irrelevant so that the phenomenology of the $w_Q < -1$ side of quintessence reduces to that of a k-essence fluid with $c_s^2 = 0$.

Furthermore, we have studied the behavior of quintessence perturbations when crossing the so-called phantom divide $w_Q = -1$. By restricting the analysis to k-essence perturbations around a given background crossing the phantom divide, we have shown that, as the speed of sound vanishes exactly at the divide, perturbations remain finite during the crossing. For $w_Q < -1$,

higher derivative terms, while irrelevant on cosmological scales, are essential to stabilize the short scale gradient instabilities. We conclude that no pathology arises during the crossing: the phantom divide can be crossed *without the addition of new degrees of freedom*. We have illustrated this with an example shown in figure 3.3. An important thing to retain is that a consistent and theoretically motivated way of comparing with data a dark energy evolution which crosses the phantom divide is to set to zero the speed of sound of perturbations.

The present analysis motivates the possibility that quintessence has a virtually vanishing speed of sound, especially when $w_Q < -1$. Such a quintessence can be detected through its effects on structure formation. The speed of sound defines the sound horizon $\ell_Q \equiv a \int c_s dt/a$, which sets the characteristic length scale of smoothness of the perturbations. In the matter dominated era $\ell_Q = 2c_s H_0^{-1}$ for a constant c_s. Hence, for $c_s = 1$ – corresponding to the speed of sound of a scalar field with a canonical kinetic term – quintessence can cluster only on scales larger than the Hubble radius while for $c_s = 0$ it clusters on all scales, thus affecting the gravitational potential and the formation of structures of dark matter and galaxies. The effect of a clustering quintessence can be measured with the cosmic microwave background [55, 56, 57, 59], galaxy redshift surveys [60], large neutral hydrogen surveys [61], or by cross-correlating the integrated Sachs-Wolfe effect in the cosmic microwave background with large scale structures [62, 63]. For instance, in [60, 62] it was found that with future surveys it will be possible to measure a zero speed of sound of dark energy if $w_Q \gtrsim -0.95$. As these analyses were restricted to positive values of $1 + w_Q$ only, it would be interesting to repeat them for negative values. Notice that for a vanishing speed of sound, dark energy will actively participate to the formation of non-linear objects, affecting the halo bias. A study of the non-linear structure formation with clustering quintessence, using the spherical collapse model, will be the subject of chapter 5.

Quintessence is perturbed by the presence of sources and thus modifies gravity as any other kind of matter. When this modification happens on scales much smaller than the Hubble radius one can properly talk about a theory of infrared modification of gravity. This happens when quintessence is close to the Ghost Condensate limit; in this case the modification of gravity is due to the Jeans instability induced by higher derivative operators, and persists even in Minkowski spacetime. Modifications of gravity induced by the Ghost Condensate have been studied in details in [28, 50]. We have considered also the additional operator proportional to $\hat{M}^3$ [32]; it would be interesting to investigate the deviation from General Relativity induced by this operator and its possible observational consequences, extending the preliminary analysis of appendix B, where the treatment has been restricted to linear perturbations in a Minkowski spacetime.

Chapter 4

Excursion Set Approach to Mass Function Calculation

Cluster observations are likely to be an important source of information about cosmology in the upcoming years. One of the main observables is the cluster number density, or mass function, which can be used as a probe of different cosmological scenarios. In this chapter, as an introduction to the following chapters, we discuss the standard calculation of the mass function of virialized objects, explaining the spherical collapse model, the original Press and Schechter approach and the excursion set theory approach. For more details, useful reviews are [64, 65].

4.1 The spherical collapse model

The spherical collapse model is the simplest approximation to the dynamics of formation of a halo. It was pioneered in [66], and it was used by Press and Schechter for their calculation of the halo mass function [67]. We will now describe it in its simplest implementations of an Einstein-de Sitter Universe; an extension to ΛCMD and dark energy models is studied in Chapter 5.

We consider a flat FRW universe filled with dust, with critical energy density $\bar{\rho} = 3H^2/(8\pi G)$. Take a homogeneous spherical overdensity with energy density $\rho_h > \bar{\rho}$ and radius R, and imagine there is an empty thin layer around the ball which compensates the inner overdensity. By Birkhoff's theorem (see [68]), the metric inside the thin layer is Schwarzschild with $M = \frac{4}{3}\pi R^3 \rho_h$, and it is independent on what happens outside the ball. At the same time, since the total density of the ball is the same of the background one, the external spacetime follows the FRLW evolution. This means that, to solve for the evolution of the spherical overdensity, we can imagine the Universe

filled with energy density ρ_h as well. In the limit in which the thin layer goes to zero, the evolution inside the ball is the same of a homogeneous closed FRW Universe, for which we can write the Friedmann equation

$$\frac{\dot{R}^2}{R^2} = \frac{8\pi G}{3}\rho_h - \frac{k}{R^2}\,, \tag{4.1}$$

where $k = +1$ is the curvature constant and $R(t)$ is the scale factor inside the ball. The equation simplifies if we introduce conformal time η such that $\mathrm{d}\eta = \mathrm{d}t/R$. The well-known solution, in parametric form, describes a cycloid:

$$t(\eta) = GM(\eta - \sin\eta) \tag{4.2}$$

$$R(\eta) = GM(1 - \cos\eta)\,. \tag{4.3}$$

Therefore, the ball expands until it reaches a maximum radius $R_M = 2GM$ at $t_M = \pi GM$, then it collapses to $R = 0$ at $t_c = 2t_M$. The non-linear evolution of the density field is given by $\rho = 3\pi M/(4R^3)$, while the average universe density evolves as $\bar{\rho} = 1/(6\pi G t^2)$.

To make contact with linear theory, we expand for small times:

$$R \simeq \frac{GM}{2}\left(\frac{6t}{GM}\right)^{\frac{2}{3}}\left[1 - \frac{1}{20}\left(\frac{6t}{GM}\right)^{\frac{2}{3}}\right]\,. \tag{4.4}$$

One important prediction of the model is the value of the overdensity linearly extrapolated at collapse time:

$$\delta_c = \frac{3M}{4\pi}\frac{R^{-3}}{\bar{\rho}_c} - 1 = \frac{3M}{4\pi}\frac{8}{G^3M^3}6\pi G t_c^2\left(\frac{G^2M^2}{36t_c^2}\right)\frac{3}{20}\left(\frac{6t_c}{GM}\right)^{\frac{2}{3}} = \frac{3}{20}(12\pi)^{\frac{2}{3}} = 1.686\,. \tag{4.5}$$

Since the geometry of this model is special, particles never cross and the collapse will not end until a singularity is formed at the center of the ball. In the real world, shell-crossing happens before that, and the collapse will stop at the radius where virial equilibrium is reached. The virial condition is

$$\frac{1}{2}\dot{R}_v^2 = \frac{1}{2}\frac{GM}{R_{\rm V}}\,, \tag{4.6}$$

and substituting the exact solution we find $\sin^2\eta_{\rm V} = 1 - \cos\eta_{\rm V}$, which is solved by $\eta_{\rm V} = k\frac{\pi}{2}$, $k = 0, 1, 2, \ldots$. Since virialization is reached after turnaround but before collapse, the correct solution is $\eta_{\rm V} = \frac{3}{2}\pi$, for which

$$R_{\rm V} = GM = \frac{R_M}{2} \qquad t_{\rm V} = GM\left(\frac{3}{2}\pi + 1\right) \tag{4.7}$$

The (non-linear) halo overdensity at virialization is usually calculated by taking the ratio of $\rho_{\rm V}$ to $\bar{\rho}$:

$$1 + \delta_{\rm V} = \frac{\rho_{\rm V}}{\bar{\rho}} = 18\pi^2 \simeq 178\,. \tag{4.8}$$

Indeed, in simulations and observations the halo radius is often defined as the radius within which the density is 200 times the background one.

4.2 The original Press & Schechter idea

Using the results from the spherical collapse model, Press & Schechter [67] derived a formula for the mass function of virialized objects as a mapping from the linear density field to the distribution of nonlinear objects.

Since we always observe overdensities smoothed with some resolution, the starting point of the calculation is to smooth the field on a scale R, using a filter function $W(x;R)$:

$$\delta(\mathbf{x};R) \equiv \int {\rm d}x' W(|\mathbf{x}-\mathbf{x}'|;R)\delta(\mathbf{x})\ . \tag{4.9}$$

The variance is then

$$\sigma^2(R) = \langle\delta(\mathbf{x};R)\rangle = \int_0^\infty \frac{{\rm d}k}{2\pi^2} k^2 P(k)|W^2(k;R)|\ . \tag{4.10}$$

There are several choices for the filter function. The most natural one is probably a top-hat in real space: $W(x;R) = 3\theta(R-x)/(4\pi R^3)$, which has the advantage that the mass enclosed within it is simply $M = 4\pi R^3/3$. However, such a window function has the undesirable property that it leads to power on all scales in Fourier space. Therefore, another popular choice is a sharp k-space filter: $W(k;R) = \theta(1/R-k)$, using which the variances on different scales are uncorrelated. Unfortunately, the volume associated with this filter function is infinite, and therefore one does not have a well-defined mass-radius relationship.

The basic assumption of the Press and Schechter approach is that objects will collapse when the smoothed overdensity field on some scale exceeds a threshold value, but that the collapse of small objects does not influence the collapse of larger ones. Following this assumption, the derivation is based on three ingredients: a characterization of the statistical properties of the primordial density fluctuations; the evolution of these fluctuations according to linear perturbation theory; the value of the threshold for collapse into a virialized object, which is based on the spherical collapse model.

The primordial density fluctuation field is assumed to be a Gaussian random variable; therefore, the smoothed field $\delta_R(\mathbf{x})$ is Gaussian as well. Thus, the probability of attaining a value between δ and $\delta + \mathrm{d}\delta$ is

$$P(\delta; R)\mathrm{d}\delta = (2\pi\sigma_R^2)^{-1/2} \exp\left[-\frac{\delta^2}{2\sigma_R^2}\right] . \tag{4.11}$$

The linear evolution of the density field is $\delta(\mathbf{x}, z) = D(z)\delta_0(\mathbf{x})$, where $D(z)$ is the growth function and $\delta_0(\mathbf{x})$ is the density contrast field linearly extrapolated to the present time.

An object of mass M will form when the smoothed density on the scale $R(M)$ will exceed the threshold $\delta_c \simeq 1.686$, given by the spherical collapse model. Any region that exceeds the critical density is assumed to meet that threshold when smoothed on some larger scale $R' > R$. Therefore, the fractional volume $\mathcal{F}(M)$ occupied by virialized objects larger than the smoothing scale $R(M)$ is the cumulative probability for a region to have a smoothed density above threshold:

$$\mathcal{F}(M) = \int_{\delta_c}^{\infty} \mathrm{d}\delta P(\delta;\, R) = \frac{1}{2}\,\mathrm{erfc}\left(\frac{\nu}{\sqrt{2}}\right) , \tag{4.12}$$

where $\mathrm{erfc}(x)$ is the complementary error function, and $\nu \equiv \delta_c/(D(z)\sigma(M))$ is the height of the threshold in units of the standard deviation of the smoothed overdensity. This formula, however, is inconsistent as it is. In fact, in hierarchical models the variance diverges as $R \to 0$, which means that all the mass of the Universe should be in virialized objects and $\mathcal{F}(0)$ should be unity. However, $\mathrm{erfc}(0) = 1$ so that, according to eq. (4.12), only half of the mass is contained in virialized objects. To correct for this issue, Press & Schechter multiplied their function by a factor of two. Proceeding with this extra factor of two, the number density of virialized objects with masses between M and $M + \mathrm{d}M$ is

$$\frac{\mathrm{d}n}{\mathrm{d}M}\mathrm{d}M = \frac{\bar{\rho}}{M}\left|\frac{\mathrm{d}\mathcal{F}(M)}{\mathrm{d}M}\right|\mathrm{d}M . \tag{4.13}$$

Substituting eq. (4.12), we get

$$\frac{\mathrm{d}n}{\mathrm{d}M}\mathrm{d}M = \sqrt{\frac{2}{\pi}}\frac{\bar{\rho}}{M^2}\nu\frac{\mathrm{d}\ln\nu}{\mathrm{d}\ln M}\exp\left(-\frac{\nu^2}{2}\right)\mathrm{d}M . \tag{4.14}$$

4.3 The excursion set theory

A major shortcoming of the Press & Schechter approach is the following: it can happen that the overdensity field $\delta_R(\mathbf{x})$ smoothed on some scale R may be less than the critical density δ_c, but it can exceed it at some *larger* scale $R' > R$. It is natural to assume that the larger region will collapse, overwhelming the less dense patches within it. Accounting for this effect provides the factor of 2

that Press & Schechter missed, since the fraction of mass in collapsed objects is increased. In the literature, this issue is referred to as the *cloud-in-cloud* problem. The correct way to proceed is to compute the *largest* value of the smoothing scale for which the density threshold is exceeded. The formalism to do this calculation constitutes the *excursion set theory* approach, beautifully explained in a paper by Bond *et al.* [69], using results in the classical review by Chandrasekhar [70]. Recently, this formalism has been put on firmer mathematical grounds by Maggiore & Riotto [71, 72, 73], in a way that allows to include effects due to diffusing threshold or non-Gaussianity of the density field (see chapters 6 and 7).

Consider the smoothed density fluctuation field $\delta(\mathbf{x}; R)$, extrapolated to the present using linear theory. Focus on a single point in space $\mathbf{x}$, which we can take as the origin by translational invariance, and study the smoothed overdensity field as a function of the smoothing scale, $\delta(R)$. For very large R, the standard deviation $\sigma(R) \to 0$, and the probability that the region exceeds the threshold is vanishingly small. Decreasing the smoothing scale R, the standard deviation $\sigma(R)$ increases, and sooner or later the smoothed overdensity will reach the threshold δ_c. Mathematically, the problem is to compute the probability of *first crossing* of the barrier δ_c at some scale R. Usually, one considers the evolution of $\delta(R)$ as a function of the variance $S(R) \equiv \sigma^2(R)$, which can be regarded as a "time" variable[1]. The trajectory $\delta(S)$ is a random walk, governed by the Langevin equation:

$$\frac{\mathrm{d}\delta(S)}{\mathrm{d}S} = \frac{4\pi}{S'(R)} \int_0^\infty \mathrm{d}k \frac{\partial \tilde{W}(k,R)}{\partial R} \tilde{\delta}(\mathbf{k}) \ . \tag{4.15}$$

The properties of the noise depend on the choice of the filter. If the filter is a top-hat in k-space (and the distribution of δ is Gaussian), we have a white noise and the resulting stochastic process is a Brownian random walk. The probability of a transition from δ to $\delta + \Delta\delta$ when the time is increased from S to $S + \Delta S$ obeys the integral equation[2]:

$$\Pi(\delta; S + \Delta S) = \int \mathrm{d}(\Delta\delta) \Psi(\Delta\delta, \Delta S) \Pi(\delta - \Delta\delta; S) \ , \tag{4.16}$$

which expresses the important fact that every step is uncorrelated with the previous history of the trajectory. It is possible to show that $\Pi(\delta, S)$ obeys the Fokker-Planck equation:

$$\frac{\partial \Pi}{\partial S} = \frac{1}{2} \frac{\partial^2 \Pi}{\partial \delta^2} \ . \tag{4.17}$$

[1] This is always possible in hierarchical models, as the variance is a strictly monotonic function of the smoothing scale.

[2] This is called Chapman-Kolomogorov or Einstein-Smoluchowski equation.

The distribution of trajectories which never exceed δ_c prior to a specific value of S is given by the solution of the Fokker-Planck equation with appropriate boundary and initial conditions. The initial condition is a Dirac delta centered on some starting point δ_0 at some scale S_0: $\Pi(\delta_0, S_0) = \delta_D(\delta_0)$. The boundary condition is that the barrier at δ_c is an absorbing barrier: $\Pi(\delta_c, S) = 0$. The solution can be obtained most elegantly by the method of images, and it is:

$$\Pi(\delta, S) = \frac{1}{\sqrt{2\pi\Delta S}} \left[\exp\left(-\frac{(\Delta\delta)^2}{2\Delta S} \right) - \exp\left(-\frac{(2(\delta_c - \delta_0) - \Delta\delta)^2}{2\Delta S} \right) \right] , \tag{4.18}$$

where $\Delta S = S - S_0$ and $\Delta\delta = \delta - \delta_0$. The first Gaussian represents the distribution of trajectories at time S without considering the barrier, and the second term removes the trajectories which have already crossed δ_c. The fraction of trajectories that have crossed the barrier at or prior to some time S is therefore

$$\mathcal{F}(S) = 1 - \int_{-\infty}^{\delta_c} \mathrm{d}\delta\, \Pi(\delta, S) = \mathrm{erfc}\left(\frac{\delta_c - \delta_0}{\sqrt{2}\Delta S} \right) . \tag{4.19}$$

Taking $S_0 = 0$ and $\delta_0 = 0$, this is precisely the Press-Schechter formula (4.12), but with the correct factor of 2. The differential probability of first crossing the barrier at time S ("first-crossing distribution") is obtained by differentiation:

$$\begin{aligned} F(S)\mathrm{d}S \equiv -\frac{\mathrm{d}\mathcal{F}(S)}{\mathrm{d}S}\mathrm{d}S = \left(\int_{-\infty}^{\delta_c} \mathrm{d}\delta \frac{\partial\Pi(\delta, S)}{\partial S} \right) \mathrm{d}S = \frac{1}{2} \left[\frac{\partial\Pi(\delta, S)}{\partial\delta} \right]_{-\infty}^{\delta_c} \mathrm{d}S \\ = \frac{\delta_c - \delta_0}{\sqrt{2\pi}\,\Delta S^{3/2}} \exp\left[-\frac{(\delta_c - \delta_0)^2}{2\Delta S} \right] \mathrm{d}S \end{aligned} \tag{4.20}$$

Eq. (4.20) is the fundamental result of excursion set theory, giving the probability of first crossing a barrier given any starting point δ_0 and any change in filtering scale ΔS. Therefore, the fraction of mass in virialized objects with mass between M and $M + \mathrm{d}M$ is

$$\frac{\mathrm{d}\mathcal{F}}{\mathrm{d}M} = \frac{1}{\sqrt{2\pi S}} \frac{\delta_c}{S} \exp\left[-\frac{\delta_c^2}{2S} \right] \left| \frac{\mathrm{d}S}{\mathrm{d}M} \right| , \tag{4.21}$$

where M is the mass which corresponds to the variance S, and the mass function follows from eq. (4.13):

$$\frac{\mathrm{d}n}{\mathrm{d}M} = \sqrt{\frac{2}{\pi}} \frac{\bar{\rho}_m}{M^2} \frac{\delta_c}{S} \left| \frac{\mathrm{d}\ln\sigma}{\mathrm{d}\ln M} \right| \exp\left[-\frac{\delta_c^2}{2\sigma^2} \right] . \tag{4.22}$$

In figure 4.1, the result (4.20) is compared to the predictions of a suite of cosmological N-body simulations. While the gross features of the mass function can be explained by the standard excursion set theory result (for instance, the power law at small masses and the exponential cutoff above the mass M_* where $\sigma(M_*) \simeq \delta_c$), the details are not. In fact, the extended Press-Schechter

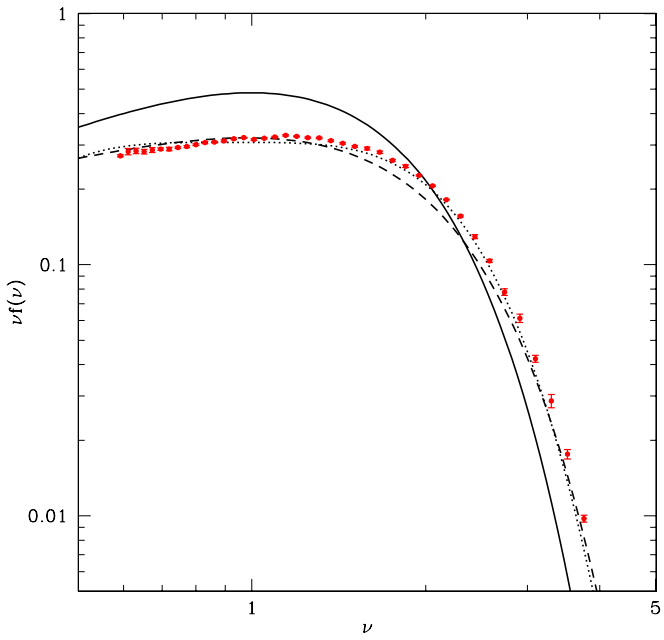


Figure 4.1: First crossing distributions. The solid line represents the standard extended Press-Schechter prediction, eq. (4.20). The dashed and dotted lines are the fits by Sheth and Tormen [74] and Jenkins *et al.* [75]. The points come from numerical data from a suite of N-body simulations by J. L. Tinker. The figure is taken from [64]

predicts too many low-mass haloes and too few high-mass ones, but, given the simplicity of the approach, the level of agreement with the data is indeed surprising and a good starting point for better analytical results.

The reason why the sharp k-space filter is so useful is that it is the only fiter for which steps are not correlated, i.e. the process is Markovian, and the solution for $\Pi(\delta, S)$ is very simple. In principle, it is possible to use other filter functions; in this case, however, it is necessary to numerically compute a large number of trajectories by solving the Langevin equation, and the probability distribution is obtained by finding the density of trajectories at each smoothing scale. This procedure is a bit cumbersome and time consuming, and moreover does not yield a closed-form solution for $\Pi(\delta, S)$ or for the first-crossing distribution. Recently, Maggiore and Riotto have shown that it is possible to take into account the effect of a filter different from a sharp k-space perturbatively in the difference between the 2-point functions calculated with a generic filter (actually they consider only the top-hat in real space) and with the sharp k one [71]. We will discuss their approach in chapter 6.

Chapter 5

The Spherical Collapse in Quintessence Cosmologies

One of the most important open questions for cosmology is whether dark energy is a dynamical component of the universe or a cosmological constant. A plethora of experiments and future observations are currently planned with the aim of improving our understanding of this question (see for instance [31] for a review). One of the most popular models of dynamical dark energy is quintessence, where the acceleration of the universe is driven by a minimally coupled scalar field with negative pressure. Standard quintessence is described by a minimally coupled canonical scalar field [76]. In this case, scalar fluctuations propagate at the speed of light and sound waves maintain quintessence homogeneous on scales smaller than the horizon scale [77]. Quintessence clustering takes place only on scales of order the Hubble radius, so that its effect is strongly limited by cosmic variance.

However, if the kinetic term is non-canonical, the speed of sound can be different from one. Indeed, as we showed in chapter 3, in an effective description of quintessence perturbations the speed of sound is a free parameter, subject to the constraints that the theory be free of ghost and gradient instabilities. In particular, if the equation of state is $w < -1$, the stability of the theory is guaranteed by the presence of higher derivative operators, provided that the speed of sound of perturbations is negligibly small: $|c_s| \lesssim 10^{-15}$.

Apart from these theoretical considerations, the fact that the speed of sound of quintessence may vanish opens up new observational consequences. Indeed, the absence of quintessence pressure gradients allows instabilities to develop on all scales, also on scales where dark matter perturbations become non-linear. Thus, we expect quintessence to modify the growth history of dark matter not

only through its different background evolution but also by actively participating to the structure formation mechanism, in the linear and non-linear regime, and by contributing to the total mass of virialized halos.

In the linear regime, a series of articles have investigated the observational consequences of a clustering quintessence. In particular, they have studied the different impact of quintessence with $c_s = 1$ or $c_s = 0$ on the cosmic microwave background [55, 56, 57, 59], galaxy redshift surveys [60], large neutral hydrogen surveys [61], or on the cross-correlation of the integrated Sachs-Wolfe effect in the cosmic microwave background with large scale structure [62, 63]. On non-linear scales, the dependence of the dark matter clustering on the equation of state of a homogeneous quintessence, i.e. with $c_s = 1$, has been investigated using N-body simulations in a number of articles (see for instance [78] and references therein for a recent account).

A popular analytical approach to study non-linear clustering of dark matter without recurring to N-body simulations is the spherical collapse model [66]. In this approach, one studies the collapse of a spherical overdensity and determines its critical overdensity for collapse as a function of redshift. Combining this information with the extended Press-Schechter theory [67, 69], one can provide a statistical model for the formation of structures which allows to predict the abundance of virialized objects as a function of their mass, as explained in chapter 4. Although it fails to match the details of N-body simulations, this simple model works surprisingly well and can give useful insigths into the physics of structure formation.

The spherical collapse can be generalized to include a cosmological constant (see for instance [79]) and quintessence with $c_s = 1$ [80] (see also [81, 82] for subsequent applications). If quintessence propagates at the speed of light, it does not cluster with dark matter but remains homogeneous. Indeed, pressure gradients contribute to maintain the same energy density of quintessence between the inner and outer part of the spherical overdensity. A study of the spherical collapse model with different quintessence potentials was performed in [83]. For a nice review on structure formation with homogeneous dark energy see also [84].

In this chapter, based on [85], we study the spherical collapse model in the case of quintessence with zero speed of sound. This represents the natural counterpart of the opposite case $c_s = 1$. Indeed, in both cases there are no characteristic length scales associated to the quintessence clustering[1], and the spherical collapse remains independent of the size of the object.

[1]The characteristic length scale associated to the quintessence clustering is the sound horizon scale, i.e., $L_s \equiv a \int c_s dt/a$. As mentioned above, this vanishes for $c_s = 0$ so that clustering takes place on all scales. For $c_s = 1$ we have $L_s = 2H_0^{-1}$, which is much larger than the scales associated to the spherical collapse.

As the spherical collapse occurs on length scales much smaller than the Hubble radius, we will describe it using a convenient coordinate system where the effect of the Hubble expansion can be treated as a small perturbation to the flat spacetime. In these "local" coordinates, the description of the spherical collapse becomes extremely simple and the well-known cases can be easily extended to quintessence with $c_s = 0$. In this case, pressure gradients are absent and quintessence follows dark matter during the collapse. Thus, in contrast with the non-clustering case $c_s = 1$ where quintessence and dark matter are not comoving, for $c_s = 0$ the collapsing region is described by an exact FLRW universe. Note that, even though the energy density of quintessence develops inhomogeneities as long as the collapse proceeds, the pressure inside and outside the overdense region remains the same. Thus, as explained below, our model does not give the same description of clustering quintessence as that proposed by [83] and studied, for instance, in [86, 87, 88].

We will see that, besides quantitative differences with respect to the $c_s = 1$ case – a different threshold for collapse and a different dark matter growth function – the $c_s = 0$ case has a remarkable qualitatively new feature. Quintessence clusters together with dark matter and participates in the total mass of the virialized object, contributing to their gravitational potential.

The chapter is organized as follows. In section 5.1 we describe quintessence models with $c_s = 0$. In section 5.2 we study spherical collapse solutions first in known cases (dark matter only, ΛCDM and $c_s = 1$ quintessence) and then in the case of a clustering dark energy, $c_s = 0$. It turns out to be much simpler to describe these solutions in coordinates for which the metric is close to Minkowski around a point in space. The equation for the evolution of the spherical collapse are solved in section 5.3 and the threshold for collapse is calculated in the various cases. This leads to the calculation of the dark matter mass function in section 5.4. In section 5.5 we study the accretion of quintessence to the dark matter haloes and its effect on the total mass function. The contribution of quintessence to the mass may be distinguished from the dark matter and baryon component in cluster measurements, as we briefly discuss in section 5.6.

5.1 The model: quintessence with $c_s^2 = 0$

Let us consider a k-essence field described by the action [39, 40]

$$S = \int \mathrm{d}^4x\sqrt{-g}\; P(\phi, X)\;, \qquad X = -g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi\;. \tag{5.1}$$

The evolution equation of ϕ derived from this equation is

$$\frac{1}{\sqrt{-g}}\partial_\mu(\sqrt{-g}\; 2P_{,X}\partial^\mu\phi) = -P_{,\phi}\;, \tag{5.2}$$

where $P_{,f} \equiv \partial P/\partial f$. The energy-momentum tensor of this field can be derived using

$$T_{\mu\nu} = -\frac{2}{\sqrt{-g}}\frac{\delta S}{\delta g^{\mu\nu}} \,, \tag{5.3}$$

and can be written in the perfect fluid form as [44]

$$T_{\mu\nu} = (\rho_Q + p_Q)u_\mu u_\nu + p_Q g_{\mu\nu} \,, \tag{5.4}$$

once we identify

$$\rho_Q = 2P_{,X}X - P \,, \qquad p_Q = P \,, \qquad u_\mu = -\frac{\partial_\mu \phi}{\sqrt{X}} \,. \tag{5.5}$$

Let us initially neglect perturbations of the metric and assume a flat FLRW universe with metric $\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2(t)\mathrm{d}\vec{x}^2$. The energy-momentum tensor of the field can be perturbed around a given background solution $\bar{\phi}(t)$ corresponding to a background energy density and pressure,

$$\bar{\rho}_Q = 2\bar{P}_{,X}\bar{X} - \bar{P} \,, \qquad \bar{p}_Q = \bar{P} \,, \tag{5.6}$$

where $\bar{X} = \dot{\bar{\phi}}^2$. To describe perturbations it is useful to write the scalar field as [32, 89]

$$\phi(t, \vec{x}) = \bar{\phi}(t + \pi(t, \vec{x})) \,, \tag{5.7}$$

where π describes the difference between the uniform time and scalar field hypersurfaces.[2] Then, eq. (5.5) can be expanded linearly in π using $\phi(t, \vec{x}) = \bar{\phi} + \dot{\bar{\phi}}\,\pi$ and $X(t, \vec{x}) = \bar{X} + \dot{\bar{X}}\pi + 2\bar{X}\dot{\pi}$. This yields, for the perturbations of the energy density, pressure and velocity,

$$\delta\rho_Q = \dot{\bar{\rho}}_Q\pi + (\bar{\rho}_Q + \bar{p}_Q + 4M^4)\dot{\pi} \,, \qquad \delta p_Q = \dot{\bar{p}}_Q\pi + (\bar{\rho}_Q + \bar{p}_Q)\dot{\pi} \,, \qquad u_i = -\partial_i\pi \,, \tag{5.8}$$

where we have used eq. (5.6) and defined $M^4 \equiv \bar{P}_{,XX}\bar{X}^2$, where M has the dimension of a mass.

To describe the evolution of perturbations we can expand the action (5.1) up to second order in π as done in chapter 3,

$$S = \int \mathrm{d}^4x\, a^3 \left[\dot{\bar{P}}\pi + 2\bar{P}_{,X}\bar{X}\dot{\pi} + \left(\bar{P}_{,X}\bar{X} + 2\bar{P}_{,XX}\bar{X}^2\right)\dot{\pi}^2 - \bar{P}_{,X}\bar{X}\frac{(\vec{\nabla}\pi)^2}{a^2} + \frac{1}{2}\ddot{\bar{P}}\pi^2 + 2\left(\bar{P}_{,X}\bar{X}\right)^{\cdot}\pi\dot{\pi}\right] . \tag{5.9}$$

The second term proportional to $\dot{\pi}$ can be integrated by parts so that the part of the action linear in π can be written using eq. (5.6) as $-\left[\dot{\bar{\rho}}_Q + 3H(\bar{\rho}_Q + \bar{p}_Q)\right]\pi$, where $H \equiv \dot{a}/a$ is the Hubble rate. This part cancels due to the background equation of motion. Furthermore, we can manipulate the last two terms of the action integrating by parts the last term, proportional to $\pi\dot{\pi}$, and making

[2] We assume that $\bar{\phi}$ is a monotonous function of t.

use of the background equation of motion, to rewrite them as $3\dot{H}\bar{P}_{,X}\bar{X}\,\pi^2$. Finally, it is convenient to rewrite the coefficients left in this expansion in terms of the background energy density and pressure using eq. (5.6). This yields

$$S = \int \mathrm{d}^4x\, a^3 \left[\frac{1}{2}\left(\bar{\rho}_Q + \bar{p}_Q + 4M^4\right)\dot{\pi}^2 - \frac{1}{2}(\bar{\rho}_Q + \bar{p}_Q)\frac{(\vec{\nabla}\pi)^2}{a^2} + \frac{3}{2}\dot{H}(\bar{\rho}_Q + \bar{p}_Q)\pi^2 \right] . \tag{5.10}$$

The coefficients of this quadratic action are completely specified by the background quantities $\bar{\rho}_Q + \bar{p}_Q$ and M^4. The latter is a function of time which we expect to vary slowly with a rate of order Hubble.[3] As shown in [33, 89], eq. (5.10) is the most general action describing quintessence in absence of operators with higher-order spatial derivatives. Note that this action is even more general than the starting Lagrangian (5.1) as it can be generically derived using only symmetry arguments [33]. An advantage of eq. (5.10) is that its coefficients are written in terms of observable quantities. Indeed, $\bar{\rho}_Q + \bar{p}_Q$ is proportional to $1 + w$, where $w \equiv \bar{p}_Q/\bar{\rho}_Q$ is the equation of state of quintessence, which we will take here and in the following to be constant. The parameter M^4 is related to the speed of sound of quintessence, given by

$$c_s^2 = \frac{\bar{\rho}_Q + \bar{p}_Q}{\bar{\rho}_Q + \bar{p}_Q + 4M^4} . \tag{5.11}$$

As can be seen from this equation, absence of ghost – i.e., positiveness of the time kinetic-term in eq. (5.10) – implies that c_s^2 has the same sign as $1 + w$ [32, 48, 89]. In particular, for $w < -1$ one has $c_s^2 < 0$, which signals the presence of gradient instabilities. As shown in [32, 89] stability can be guaranteed by the presence of higher derivative operators but requires that the speed of sound is extremely small, practically zero [89].

Regardless of the motivations expressed above on the stability of single field quintessence for $w < -1$, in the following we will be interested in considering the limit $c_s^2 \to 0$, which is obtained when $|\bar{\rho}_Q + \bar{p}_Q| \ll M^4$. We will see that what turns out to be physically relevant are the density and pressure perturbations on surfaces of constant ϕ, i.e. of constant π. These are the perturbations in the so-called velocity orthogonal gauge, and using eq. (5.8) they are given by

$$\delta\rho_Q^{(\mathrm{v.o.})} \equiv \delta\rho_Q - \dot{\bar{\rho}}_Q\pi = (\bar{\rho}_Q + \bar{p}_Q + 4M^4)\dot{\pi} , \qquad \delta p_Q^{(\mathrm{v.o.})} \equiv \delta p_Q - \dot{\bar{p}}_Q\pi = (\bar{\rho}_Q + \bar{p}_Q)\dot{\pi} . \tag{5.12}$$

Indeed, c_s^2 defined in eq. (5.10) can be written as [89]

$$c_s^2 = \frac{\delta p_Q^{(\mathrm{v.o.})}}{\delta\rho_Q^{(\mathrm{v.o.})}} . \tag{5.13}$$

[3] The time variation of M^4 is expected to be even slower than Hubble, i.e. of order $(1+w)H$, which is the typical time variation of ρ_Q.

Thus, the pressure perturbation is suppressed with respect to the energy density perturbation by the smallness of the speed of sound. As we will see, in the limit $c_s \to 0$ this implies that pressure forces are negligible and quintessence follows geodesics, remaining comoving with the dark matter.

In the limit $c_s \to 0$, the energy density perturbation on velocity orthogonal slicing becomes

$$\delta\rho_Q^{(\mathrm{v.o.})} = 4M^4\dot{\pi} \,. \tag{5.14}$$

Note that since $\dot{\pi} \sim H\pi$, the difference between $\delta\rho_Q^{(\mathrm{v.o.})}$ and $\delta\rho_Q$ is negigible for small speed of sound, $\delta\rho_Q^{(\mathrm{v.o.})} \simeq \delta\rho_Q$. All these conclusions hold independently of the value of $\delta\rho_Q/\bar{\rho}_Q$, provided that the effective theory described by action (5.10) remains valid, i.e. for $\dot{\pi} \ll 1$ [50, 89]. In particular, they hold also when perturbations in the energy density of quintessence become non-linear, i.e., for $\delta\rho_Q \gg \bar{\rho}_Q$.

Gravitational perturbations can be straightforwardly included in the action (5.10) as in [89]. As a warm-up exercise we will here, instead, study the evolution of quintessence in the spherical collapse solution. According to the spherical collapse model, the overdensity can be described as a closed FLRW universe with a scale factor R which is different from the one of the background a. This remains true also when we take into account quintessence with negligible speed of sound. Indeed, eq. (5.13) shows that there is no pressure difference between the inside and the outside of the overdense region. Therefore, inside the overdensity we can describe quintessence using eq. (5.9), where the time evolution of the metric is described by the scale factor R and we thus replace a^3 by R^3.

With this new action, the second term proportional to $\dot{\pi}$ can be integrated by parts and the coefficients of the linear part of the action rewritten in terms of $\bar{\rho}_Q$ and $\bar{p}_Q$ using eq. (5.6). However, now the linear part of the action does not cancel but can be written, using the background equation of motion, as $-\delta H(\bar{\rho}_Q + \bar{p}_Q)\pi$, where we have defined $\delta H = H_{\mathrm{in}} - H$, with $H_{\mathrm{in}} \equiv \dot{R}/R$. We are thus left with a linear term in the action, due to the difference between the rates of expansion inside and outside the overdensity. After manipulations of the last two terms in eq. (5.9), similarly to what was done to derive eq. (5.10), the action inside the overdensity becomes[4]

$$\begin{aligned} S = \int \mathrm{d}^4x\, R^3 \Bigg[& \frac{1}{2}\left(\bar{\rho}_Q + \bar{p}_Q + 4M^4\right)\dot{\pi}^2 - \frac{1}{2}(\bar{\rho}_Q + \bar{p}_Q)\frac{(\vec{\nabla}\pi)^2}{a^2} + \frac{3}{2}\dot{H}(\bar{\rho}_Q + \bar{p}_Q)\pi^2 \\ & - 3\delta H(\bar{\rho}_Q + \bar{p}_Q)\pi - \frac{3}{2}\delta H(\bar{\rho}_Q + \bar{p}_Q)^{\cdot}\,\pi^2 \Bigg] \,. \end{aligned} \tag{5.15}$$

[4]We will not include in the action the contribution to $\sqrt{-g}$ coming from the curvature of the closed FLRW universe. Indeed, as it is time independent, it does not affect our discussion.

5.1 The model: quintessence with $c_s^2 = 0$

Using that $\dot{\pi} \sim H\pi$, neglecting time variations of M and discarding terms suppressed when $|\bar{\rho}_Q + \bar{p}_Q| \ll M^4$ (i.e., in the limit $c_s^2 \to 0$) the equation of motion of π derived from this action reads

$$\ddot{\pi} + 3\frac{\dot{R}}{R}\dot{\pi} = -\frac{3\delta H}{4M^4}(\bar{\rho}_Q + \bar{p}_Q)\,. \tag{5.16}$$

As expected, the quintessence perturbation induced by the overdensity is proportional to $1 + w$, i.e. it vanishes in the limit of the cosmological constant. Note that the source term on the right-hand side of this equation can be written as $-3c_s^2\delta H$ and is suppressed by the smallness of c_s^2. This implies that, even for large overdensities, i.e. $\delta H \gtrsim H$, variations of π due to the gravitational potential well are extremely small, $\dot{\pi} \sim c_s^2$, inside the regime of validity of the effective theory. Furthermore, this also implies that the difference in π between the homogeneous and closed FLRW solutions is also tiny, $\Delta\pi \sim c_s^2 H^{-1}$. Thus, the quintessential scalar field practically lies on the same point of its potential.

Equation (5.16) can be written, using eq. (5.14) (and $\delta\rho_Q \simeq \delta\rho_Q^{(\mathrm{v.o.})}$), as

$$\dot{\delta\rho}_Q + 3\frac{\dot{R}}{R}\delta\rho_Q = -3\delta H(\bar{\rho}_Q + \bar{p}_Q)\,. \tag{5.17}$$

Note that, as δH is always negative, the sign of $\delta\rho_Q$ is the same as that of $1 + w$. Remarkably, this implies that for $w < -1$ dark matter halos accrete negative energy from quintessence, as was noticed at linear level in [56]. Combining eq. (5.17) with the background continuity equation, $\dot{\bar{\rho}}_Q + 3H(\bar{\rho}_Q + \bar{p}_Q) = 0$, we obtain

$$\dot{\rho}_Q + 3\frac{\dot{R}}{R}(\rho_Q + \bar{p}_Q) = 0\,. \tag{5.18}$$

This equation describes the evolution of the energy density of quintessence with $c_s^2 = 0$ inside a spherical overdensity dominated by dark matter. Note that the pressure perturbation is absent, as it is suppressed by $c_s^2 \to 0$. Indeed, this equation differs from the description currently given in the literature for clustering dark energy. In particular, the analogue of this equation given in [83, 86] includes the pressure perturbation $\delta p_Q = w\delta\rho_Q$. Including the pressure perturbation δp_Q leads to an incorrect description even in the linear regime, in contrast with eq. (5.18) which does match the linear theory for small overdensities.

In the following two sections we will make this analysis more complete and derive all the equations necessary to describe the spherical collapse with quintessence.

5.2 Spherical collapse in local coordinates

As we did in the former section, the spherical collapse is usually treated using FLRW coordinates, as in the simplest cases (Einstein-de Sitter or ΛCDM universe) the top-hat overdensity evolves as an independent closed universe. This somewhat obscures a crucial simplification of the problem, i.e. that the collapse of dark matter haloes occurs on scales much smaller than the Hubble radius. In this limit one can treat gravity as a small perturbation of Minkowski space[5]. As the dynamics of $c_s = 0$ quintessence is not completely intuitive, we want to make use of a coordinate system where this simplification is explicit; this will also make the dynamics of the other cases of spherical collapse clearer. We thus choose a coordinate system around a given point, such that the deviation of the metric from Minkowski is suppressed by H^2r^2, where r is the distance from the point, for any time[6]. Notice that we do not want to limit the validity of our approximation to times shorter than H^{-1} because this is also the typical time-scale of the evolution of a dark matter halo. These requirements define the so called Fermi coordinates. Note also that we are not taking any Newtonian limit: as we are interested in quintessence we cannot neglect pressure as source of gravity.

A particular choice of Fermi coordinates are the so-called Fermi normal coordinates [90], where the deviation of the metric from Minkowski can be written as a Taylor expansion around the origin whose leading coefficients are components of the Riemann tensor. These are (with the conventions of [91])

$$g_{00} = -1 - R_{0l0m}|_{\vec{0}}\, x^l x^m + \dots \,, \tag{5.19}$$

$$g_{0i} = 0 - \frac{2}{3} R_{0lim}|_{\vec{0}}\, x^l x^m + \dots \,, \tag{5.20}$$

$$g_{ij} = \delta_{ij} - \frac{1}{3} R_{iljm}|_{\vec{0}}\, x^l x^m + \dots \,. \tag{5.21}$$

Here we will be interested only in spherically symmetric solutions. As R_{0lim} vanishes because of rotational symmetry, g_{0i} must be of order higher than r^2, and we can neglect it in the following discussion. Furthermore, rotational symmetry implies that the corrections to g_{00} will be proportional to r^2 while those to g_{ij} will be proportional either to $r^2\delta_{ij}$ or to $x_i x_j$. It is possible to make a redefinition of the radial coordinate such as to get rid of the term $x_i x_j$ without affecting g_{00} and g_{0i} at $\mathcal{O}(r^2)$. In such a way we are using Fermi coordinates which are not of the normal form. In this

[5] For a recent use of this approximation in cosmology see [30]

[6] In the spherically symmetric case, the range of validity of this approximation goes to zero close to the collapse singularity. However, this is not relevant because the singularity is anyway an artifact of the spherical symmetry. In the real case the curvature of space remains small and the halo reaches virialization.

case the metric can be written in the Newtonian gauge (not to be confused with the cosmological perturbation theory Newtonian gauge) form as

$$ds^2 = -(1+2\Phi)dt^2 + (1-2\Psi)d\vec{x}^2 \,, \tag{5.22}$$

where Φ and Ψ are proportional to r^2. In this gauge the 00 component of the Einstein equation gives

$$\nabla^2\Psi = 4\pi G\rho \,. \tag{5.23}$$

The part of the ij Einstein equation proportional to the identity gives

$$6\ddot{\Psi} + 2\nabla^2(\Phi - \Psi) = 24\pi G p \,. \tag{5.24}$$

As the typical time scale is of order Hubble, the $\ddot{\Psi}$ term is suppressed with respect to $\nabla^2\Psi$ by $\mathcal{O}(H^2r^2)$, and can therefore be neglected. Thus, using eq. (5.23) we obtain

$$\nabla^2\Phi = 4\pi G(\rho + 3p) \,. \tag{5.25}$$

As a first step, let us show how one can use these coordinates to describe an unperturbed FLRW solution with non-vanishing curvature. In isotropic comoving coordinates this metric is written as

$$ds^2 = -d\tau^2 + a(\tau)^2 \frac{d\vec{y}^2}{(1+\frac{1}{4}K\vec{y}^2)^2} \,, \tag{5.26}$$

where K is the curvature parameter. Performing the change of coordinates $\tau = t - \frac{1}{2}Hr^2$ and $\vec{y} = \frac{\vec{x}}{a}(1+\frac{1}{4}H^2r^2)$ [30], with a and H evaluated at t rather than at τ, one gets at first order in H^2r^2,

$$ds^2 \simeq -\left[1-(\dot{H}+H^2)r^2\right]dt^2 + \left[1-\frac{1}{2}(H^2+K/a^2)r^2\right]d\vec{x}^2 \,, \tag{5.27}$$

which is indeed of the Fermi form (5.22), where the corrections from flat spacetime are given by

$$\Phi = -\frac{1}{2}(\dot{H}+H^2)r^2 \,, \qquad \Psi = \frac{1}{4}(H^2+K/a^2)r^2 \,. \tag{5.28}$$

Let us now show that the metric (5.27) is a solution of the Einstein equations (5.23) and (5.25). In the coordinates (t, x^i), $\rho(\tau)$ and $p(\tau)$ are not space independent; however, their space dependence is suppressed by H^2r^2, so that it can be neglected. With spherical symmetry, assuming regularity at the origin, the two Einstein equations (5.23) and (5.25) are then solved by

$$\Psi = \frac{8\pi G\rho}{3}\frac{r^2}{4} \,, \tag{5.29}$$

and

$$\Phi = \frac{4\pi G}{3}(\rho + 3p)\frac{r^2}{2} \ . \tag{5.30}$$

Comparing these expressions with (5.28) we recover the two Friedmann equations, respectively,

$$H^2 + \frac{K}{a^2} = \frac{8\pi G}{3}\rho \ , \tag{5.31}$$

and

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p) \ . \tag{5.32}$$

Note also that the traceless part of the ij Einstein equation, $(\partial_i\partial_j - \frac{1}{3}\delta_{ij}\nabla^2)(\Phi - \Psi) = 0$, is trivially satisfied by the expressions above. Matter stays at fixed $\vec{y}$ in the original FLRW coordinates; therefore it moves in the new coordinates as $\vec{x} \propto a$, i.e. with velocity $\vec{v} = H\vec{x}$. Finally, using this equality one can check that also the $0i$ component of the Einstein equations is satisfied.

Let us now look at the dynamical equations for the fluid. The time component of the conservation of the energy-momentum tensor gives (see for example [92]) the continuity equation

$$\dot{\rho} + \vec{\nabla}\cdot[(\rho + p)\vec{v}] = 0 \ . \tag{5.33}$$

This equation is the same as in Minkowski spacetime, as the gravitational corrections only induce terms suppressed by $\mathcal{O}(H^2r^2)$. When the velocity $\vec{v}$ is simply given by the unperturbed Hubble flow we obtain the standard conservation equation in expanding space, $\dot{\rho} + 3H(\rho + p) = 0$.

The spatial component of the conservation of the energy-momentum tensor gives the Euler equation,

$$\dot{\vec{v}} + (\vec{v}\cdot\vec{\nabla})\vec{v} = -\frac{1}{(\rho + p)}\left[\vec{\nabla}p + \vec{v}\,\frac{\partial p}{\partial t}\right] - \vec{\nabla}\Phi \ , \tag{5.34}$$

where we have assumed that $v \ll c$. At leading order in $\mathcal{O}(H^2r^2)$, gravitational perturbations enter only through the last term on the right-hand side of this equation. In the particular case of an isotropic and homogenous solution the first term on the right-hand side exactly cancels: as $p(\vec{x}, t) = p(t - Hx^2/2)$, $\vec{\nabla}p = -\dot{p}H\vec{x}$ and the gradient of the pressure cancels with the term coming from its time dependence. This is not surprising, as what matters in the Euler equation is the 4-dimensional gradient of pressure perpendicular to the fluid 4-velocity. In this case eq. (5.34) reduces to

$$\dot{\vec{v}} + (\vec{v}\cdot\vec{\nabla})\vec{v} = -\vec{\nabla}\Phi \ . \tag{5.35}$$

This equation is verified by the Hubble flow $\vec{v} = H\vec{x}$, since we get

$$\frac{\ddot{a}}{a}\vec{x} = -\vec{\nabla}\Phi \ , \tag{5.36}$$

which is clearly satisfied by the explicit expression for Φ, eq. (5.28).

We can now use these local coordinates to describe the spherical collapse in various cases, starting from the simplest.

- **Dark matter only**

Let us take a spherically symmetric distribution of matter around the origin. As both the gravitational potentials Φ and Ψ satisfy the Poisson equation, we do not need to know how the mass is radially distributed to solve for the gravitational background outside a given radius r: we just need the total mass inside the radius r. In particular (see figure 5.1), if inside a given radius $r_{\rm out}$ a distribution contains as much matter as the unperturbed cosmological solution, from the outside it will look exactly as the unperturbed background. This implies that we can smoothly glue this solution at $r = r_{\rm out}$ to the cosmological background, and that the latter will not be affected by the gravitational collapse inside. This is of course a linearized version of Birkhoff's theorem in General Relativity.

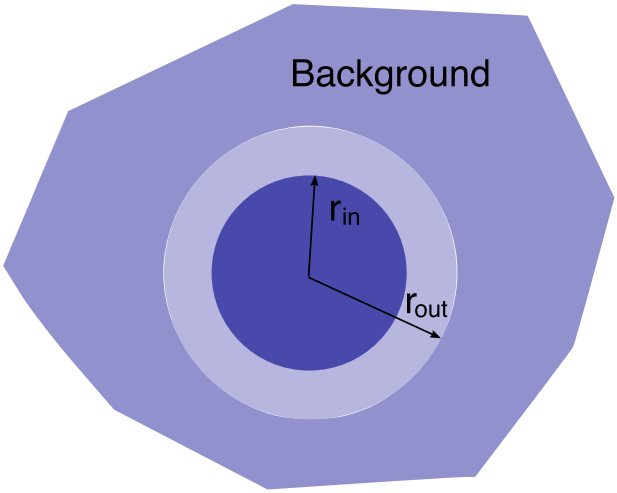


Figure 5.1: *Spherical collapse*

Conversely, the solution inside a given radius is not affected by what happens outside. In particular, if we assume a homogeneous initial condition inside a radius $r_{\rm in}$ (with $r_{\rm in} < r_{\rm out}$), this central region will evolve as if these homogeneous initial conditions were extended outside, i.e. as a complete FLRW solution [66]. The central overdense region will remain exactly homogeneous, reaching maximum expansion and then collapsing.

Without further assumptions, the evolution of the layer $r_{\rm in} < r < r_{\rm out}$ does not enjoy particular simplifications and its evolution must be computed as a function of the initial profile. In any case, this is usually irrelevant as we are only interested in the fate of the $r < r_{\rm in}$ region. If we assume that the layer $r_{\rm in} < r < r_{\rm out}$ is empty, the Poisson equations (5.23) and (5.25) give $1/r$ solutions for the potentials, like for a source localized at the origin. This is the linearization of the exact Schwarzschild solution.

- **Dark matter and a cosmological constant**

The considerations above also hold when we include a cosmological constant. Although Φ and Ψ now solve different equations (because $p_\Lambda \neq 0$), they are both of the Poisson form. Thus, we still have an unperturbed evolution outside $r_{\rm out}$ if the total matter inside matches the background value. Assuming initial homogeneity, the central region $r < r_{\rm in}$ will evolve like a complete FLRW universe [79]. Although now pressure does not vanish, it just comes from the cosmological constant which does not define a preferred frame and is therefore comoving with dark matter both inside and outside the overdensity.

- **Non-clustering quintessence: $c_s = 1$**

When quintessence has a speed of sound $c_s = 1$, it does not effectively cluster but it keeps on following the cosmological background solution, irrespective of the dark matter clustering [80].

As before, outside $r_{\rm out}$ there is an unperturbed cosmological background. What is new now is that the central region $r < r_{\rm in}$ does *not* behave as a complete FLRW solution, even if we start with a homogeneous overdensity. Indeed, quintessence and dark matter do not have a common velocity: while dark matter slows down and eventually starts collapsing, quintessence keeps following the external Hubble flow $\vec{v}_{\rm Q} = H_{\rm out}\vec{x}$. Note that, on the other hand, in the cosmological constant case one cannot define a dark energy 4-velocity as its energy-momentum tensor is proportional to the metric. To study the evolution of the dark matter overdensity one must use the Euler equation (5.36). Here, what defines the velocity flow of dark matter is the effective "scale factor" R, $\vec{v}_{\rm m} = \dot{R}/R\,\vec{x}$. This yields

$$\frac{\ddot{R}}{R}\vec{x} = -\vec{\nabla}\Phi \,. \tag{5.37}$$

Using the explicit solution (5.30) for Φ, this equation becomes [80]

$$\frac{\ddot{R}}{R} = -\frac{4\pi G}{3}(\rho_{\rm m} + \bar{\rho}_{\rm Q} + 3\bar{p}_{\rm Q}) \,, \tag{5.38}$$

where we have separated the contribution of dark matter and quintessence to Φ. Notice that, although this equation looks like one of the Friedmann equations, the dynamics of R is not the same as for a FLRW universe. Indeed, $\rho_{\rm m}$ evolves following the scale factor R, while the quintessence follows the external scale factor a. In a FLRW universe, from eq. (5.38) together with the continuity equation one can derive the first Friedmann equation, $(\dot{R}/R)^2 = \frac{8\pi G}{3}\rho - K/R^2$. Here, as the different components follow different scale factors, this is not longer possible and the first Friedmann equation does not hold.

- **Clustering quintessence: $c_s = 0$**

Let us now move to the subject of this chapter. We want to show that, in the limit of vanishing speed of sound, quintessence remains comoving everywhere with dark matter. In particular, this implies that in the region $r < r_{\rm in}$ quintessence follows dark matter in the collapse and the overdensity behaves as an exact FLRW solution so that, contrary to the $c_s = 1$ case, also the first Friedmann equation holds. The fact that quintessence remains comoving with dark matter can be understood both by using the fluid equations or directly from the scalar field equation of motion.

In the fluid language, the dynamics is described by the Euler equation (5.34). In general, in the presence of sizeable pressure gradients, a fluid does not remain comoving with dark matter, i.e. it does not follow geodesics. Since quintessence has a sizeable pressure, the fact that it moves following geodesics may be unexpected but it is obtained in the limit $c_s \to 0$. This can be easily seen by rewriting the Euler equation for quintessence in covariant form as

$$u^\mu \nabla_\mu u^\nu = -\frac{1}{(\rho_Q + p_Q)}(g^{\nu\sigma} + u^\nu u^\sigma)\nabla_\sigma p_Q \,, \tag{5.39}$$

where u^μ is the quintessence 4-velocity. When the right-hand side of this equation vanishes, the 4-velocity solves the geodesic equation. In fact, the pressure gradient is multiplied by the projector perpendicular to the fluid 4-velocity. This is the same as projecting on surfaces of constant ϕ, and it is equivalent to a gradient of the velocity-orthogonal pressure perturbation that appears in equation (5.12), which involves only $\dot{\pi}$, and not π. By eq. (5.13) this is negligible in the limit $c_s \to 0$ and thus the right-hand side of (5.39) vanishes.

This result is even clearer in the scalar field language Taking the derivative of the equation defining the quantity X in (5.1),

$$\partial^\nu(\partial^\mu\phi\partial_\mu\phi) = -\partial^\nu X \,, \tag{5.40}$$

and writing it in terms of the 4-velocity $u^\mu = -\partial^\mu\phi/\sqrt{X}$, we have

$$2u^\mu\sqrt{X}\nabla_\mu(\sqrt{X}u^\nu) = -\partial^\nu X \ , \tag{5.41}$$

and therefore

$$u^\mu\nabla_\mu u^\nu = -\frac{1}{2X}(g^{\nu\sigma} + u^\nu u^\sigma)\partial_\sigma X \ . \tag{5.42}$$

Equation (5.39) is recovered using eq. (5.5) and taking into account that $\partial_\sigma P(\phi, X) = \partial P/\partial\phi \cdot \partial_\sigma\phi + \partial P/\partial X \cdot \partial_\sigma X$ and that the first term vanishes when multiplied by the projector orthogonal to u^μ. From this, we clearly see that what matters is only the gradient of the pressure on $\phi = $ const hypersufaces. This vanishes in the limit $c_s \to 0$ and thus we have geodesic motion. We stress that, although quintessence with $c_s = 0$ follows geodesics, its dynamics is quite different from the one of dark matter. Pressure does not accelerate the quintessence 4-velocity but it does affect the energy conservation equation (5.33). Moreover, quintessence does not enjoy a conserved current, while dark matter particle number is conserved; this is related to the absence of the shift symmetry $\phi \to \phi + c$ in the scalar field Lagrangian (see for example [93]).

As discussed in section 5.1, the different dark matter evolution inside and outside the overdensity changes the quintessence solution by a very tiny amount $\Delta\pi \sim c_s^2 H^{-1}$: the quintessence field sits at the same position along its potential, $\phi = \bar{\phi}(t)$, apart from negligible c_s^2 corrections. Notice that this was derived using two different Friedmann coordinate systems, one following dark matter inside the overdensity and one following the unperturbed Hubble flow outside. Thus, in reality we have two solutions $\phi = \bar{\phi}(t_{\rm in})$ and $\phi = \bar{\phi}(t_{\rm out})$ respectively. Once these two solutions are written in the same local coordinates (5.27), the solution for ϕ becomes $\phi = \bar{\phi}(t - \frac{1}{2}H_{\rm in}r^2)$ for $r < r_{\rm in}$ and $\phi = \bar{\phi}(t - \frac{1}{2}H_{\rm out}r^2)$ for $r > r_{\rm out}$. This implies that in these coordinates π has to jump in the layer between the two regions, by an amount $\Delta\pi \sim \frac{1}{2}r^2\delta H$, and that this jump is not suppressed by c_s^2. One may expect that the scalar field would "react" to this gradient between the inside and the outside. However, this does not happen in the limit $c_s \to 0$ as the spatial kinetic term is very suppressed. Let us see this explicitly.

To study the scalar field equations in the local coordinates, one can start by writing the equations in Minkowski space and then check a posteriori that the deviation of the metric from flat space only gives relative corrections $\mathcal{O}(H^2r^2)$. The evolution equation for ϕ, eq. (5.2), reads in Minkowski space

$$-\partial_t(P_{,X}\dot{\phi}) + \partial_i(P_{,X}\partial_i\phi) = -\frac{1}{2}P_{,\phi} \ . \tag{5.43}$$

If we try a comoving solution of the form $\phi = \bar{\phi}(t - \frac{1}{2}Hr^2)$, we end up with the standard FLRW equation for $\bar{\phi}$, inclusive of the friction term,

$$-\partial_t(\bar{P}_{,X}\dot{\bar{\phi}}) - 3H\bar{P}_{,X}\dot{\bar{\phi}} = -\frac{1}{2}\bar{P}_{,\phi}\,. \tag{5.44}$$

Metric fluctuations give only a correction to this equation of order $\mathcal{O}(H^2r^2)$. As we discussed, the two homogeneous solutions for $r < r_{\rm in}$ and $r > r_{\rm out}$ are different so that we expect gradient terms to smooth out the initial top-hat profile. To estimate the thickness L of the layer over which the smoothing takes place, we can study perturbations around a top-hat profile and require the spatial and time kinetic term of the perturbation π in eq. (3.9) to be comparable:

$$M^4\dot{\pi}^2 \sim (\rho_Q + p_Q)\frac{\pi^2}{L^2}\,. \tag{5.45}$$

Using that $\dot{\pi} \sim H\pi$, from this comparison we obtain $L \sim |c_s|H^{-1}$. This makes perfect sense: our top-hat profiles are smoothed out over a distance comparable to the sound horizon[7].

In conclusion, the solutions outside and inside the overdensity are exact FLRW with quintessence comoving with dark matter. Gradient terms will smooth out this solutions on scales of order of the sound horizon, which vanishes for $c_s \to 0$. This discussion also tells us that taking $c_s = 0$ will be correct only for objects which are much bigger than the sound horizon $|c_s|H^{-1}$. In the opposite limit of an object which is much smaller than the sound horizon, one can treat quintessence as unperturbed as discussed above in the $c_s^2 = 1$ case. For example, if one is interested in objects larger than 1Mpc, one can neglect the speed of sound as long as $|c_s| \lesssim 10^{-4}$.

5.3 Solving the spherical collapse

In this section we derive the equations for the spherical collapse of dark matter in the presence of quintessence with vanishing speed of sound and we compute their solutions numerically.

- **The background universe**

The background is described by a flat FLRW metric with scale factor satisfying the Friedmann equation

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}(\bar{\rho}_m + \bar{\rho}_Q)\,, \tag{5.46}$$

[7] It is straightforward to check that this estimate is not altered by the higher derivative operators that are required for stability when $w < -1$ [32].

where $\bar{\rho}_m$ and $\bar{\rho}_Q$ are the background energy density of dark matter and quintessence, respectively. For later purposes, we express $\bar{\rho}_m$ and $\bar{\rho}_Q$ in terms of the fractional abundance of dark matter Ω_m,

$$\bar{\rho}_m \equiv \frac{3H^2}{8\pi G}\Omega_m \,, \qquad \bar{\rho}_Q = \frac{1-\Omega_m}{\Omega_m}\bar{\rho}_m \,. \tag{5.47}$$

Dark matter redshifts with the expansion as the physical volume, $\bar{\rho}_m \propto a^{-3}$, while the energy density of quintessence scales as $\bar{\rho}_Q \propto a^{-3(1+w)}$. The dark matter contribution to the critical density Ω_m can be written as a function of its value today, $\Omega_{m,0}$, and x, the scale factor normalized to unity today (at $t = t_0$):

$$x \equiv a/a_0 \,. \tag{5.48}$$

This yields

$$\Omega_m(x) = \left(1 + \frac{1-\Omega_{m,0}}{\Omega_{m,0}} x^{-3w}\right)^{-1} \,. \tag{5.49}$$

Equation (5.47) can then be rewritten as

$$\bar{\rho}_m = \frac{3H_0^2}{8\pi G}\frac{\Omega_{m,0}}{x^3} \,, \qquad \bar{\rho}_Q = \frac{1-\Omega_{m,0}}{\Omega_{m,0}} x^{-3w}\bar{\rho}_m \,, \tag{5.50}$$

where the second equation follows from (5.49). Furthermore, rescaling the time variable by defining

$$\eta \equiv \sqrt{\Omega_{m,0}}\, H_0 t \,, \tag{5.51}$$

one can rewrite the Friedmann equation as

$$\frac{\mathrm{d}x}{\mathrm{d}\eta} = (x\Omega_m(x))^{-1/2} \,. \tag{5.52}$$

The initial condition for x can be imposed at some small initial time η_i during matter dominance: $x_i = (3\eta_i/2)^{2/3}$. Then, eqs. (5.49) and (5.52) completely describe the background evolution of the metric and energy-momentum tensors.

- **The linear evolution**

Before studying the collapsing spherical overdensity we derive the evolution equations of perturbations of dark matter and quintessence in the linear regime. As we consider scales much smaller than the Hubble radius, the gauge dependence of perturbations is not important. We will thus perturb the continuity and Euler equations in local coordinates, eqs. (5.33) and (5.35), adding small inhomogeneous perturbations $\delta(t,\vec{x})$ and $\vec{u}(t,\vec{x})$ to the homogeneous energy density and Hubble flow velocity:

$$\rho = \bar{\rho}(1+\delta) \,, \qquad \vec{v} = H\vec{x} + \vec{u} \,. \tag{5.53}$$

Let us start from the dark matter. Perturbing at linear order eq. (5.33) with $p_m = 0$ yields

$$\left(\frac{\partial}{\partial t} + H\,\vec{x}\cdot\vec{\nabla}_x\right)\delta_m = -\vec{\nabla}_x\cdot\vec{u} \qquad \text{(local coords)}\,, \tag{5.54}$$

where we have specified that we are describing perturbations using local (physical) spatial coordinates $\vec{x}$. On the other hand, on the left-hand side of this equation one recognizes the time derivative at fixed comoving coordinates $\vec{y} = \vec{x}/a(t)$, i.e.,

$$\left(\frac{\partial}{\partial t}\right)_{\vec{y}} = \left(\frac{\partial}{\partial t} + H\,\vec{x}\cdot\vec{\nabla}\right)_{\vec{x}}\,. \tag{5.55}$$

Indeed, here we are interested in describing the evolution of an overdensity of dark matter contained in a comoving volume. Thus, we describe δ_m and $\vec{u}$ as a function of the comoving coordinates, which simply gives

$$\dot{\delta}_m + \frac{1}{a}\vec{\nabla}_y\cdot\vec{u} = 0\,. \tag{5.56}$$

To close this equation we need the evolution of the dark matter peculiar velocity $\vec{u}$. This can be obtained by perturbing at linear order the Euler equation (5.35). Using comoving coordinates, the perturbed Euler equation becomes

$$\dot{\vec{u}} + H\vec{u} + \frac{1}{a}\vec{\nabla}_y\delta\Phi = 0\,, \tag{5.57}$$

where $\delta\Phi$ is the perturbation of the Newtonian potential,

$$\delta\Phi = \Phi + \frac{1}{2}(\dot{H} + H^2)r^2\,. \tag{5.58}$$

Equations (5.56)–(5.58) have been derived for instance in [94] in the context of Newtonian mechanics described with expanding coordinates, for a pressureless fluid in the presence of vacuum energy. Here the Poisson equation for $\delta\Phi$ is sourced by both dark matter and quintessence perturbations:

$$\frac{1}{a^2}\nabla_y^2\delta\Phi = 4\pi G(\bar{\rho}_m\delta_m + \bar{\rho}_Q\delta_Q)\,, \tag{5.59}$$

where we have used that $\delta p_Q = 0$. The final step is to eliminate the peculiar velocity by subtracting the divergence of eq. (5.57) from the time derivative of eq. (5.56). With the Poisson equation (5.59) we obtain

$$\ddot{\delta}_m + 2H\dot{\delta}_m = 4\pi G(\bar{\rho}_m\delta_m + \bar{\rho}_Q\delta_Q)\,. \tag{5.60}$$

For quintessence we perturb the continuity equation (5.33) which gives, in comoving coordinates, using $\delta p_Q = 0$,

$$\dot{\delta}_Q - 3Hw\delta_Q + (1+w)\frac{1}{a}\vec{\nabla}\cdot\vec{u} = 0\,. \tag{5.61}$$

To eliminate the divergence of the peculiar velocity we can use eq. (5.56) taking quintessence to be comoving with dark matter. Indeed, as explained above, both the dark matter and quintessence follow geodesics and are dragged by the same potential well and the growing mode of their velocities is the same. Thus

$$\dot{\delta}_Q - 3Hw\delta_Q = (1+w)\dot{\delta}_m \ . \tag{5.62}$$

In matter dominance, when $\delta_m \propto a$, the solution of this equation is [89]

$$\delta_Q = \frac{1+w}{1-3w}\delta_m \ . \tag{5.63}$$

Note that the denominator on the right-hand side further suppresses the perturbation of quintessence with respect to the naïve $1+w$ estimate.

In terms of the dimensionless variables x and η, respectively defined in eqs. (5.48) and(5.51), equations (5.60) and (5.62) are rewritten as

$$\frac{\mathrm{d}^2\delta_m}{\mathrm{d}\eta^2} + \frac{2}{x}\frac{\mathrm{d}x}{\mathrm{d}\eta}\frac{\mathrm{d}\delta_m}{\mathrm{d}\eta} = \frac{3}{2x^3}\left(\delta_m + \frac{1-\Omega_{m,0}}{\Omega_{m,0}}x^{-3w}\delta_Q\right) \ , \tag{5.64}$$

where we have used eq. (5.50), and

$$\frac{\mathrm{d}\delta_Q}{\mathrm{d}\eta} - \frac{3}{x}\frac{\mathrm{d}x}{\mathrm{d}\eta}w\delta_Q = (1+w)\frac{\mathrm{d}\delta_m}{\mathrm{d}\eta} \ . \tag{5.65}$$

The initial conditions are set in terms of the initial dark matter density contrast $\delta_{m,i}$. In matter dominance $\dot{\delta}_m = H\delta_m$, i.e. $\mathrm{d}\delta_m/\mathrm{d}\eta|_i = 2\delta_{m,i}/(3\eta_i)$, while the value of $\delta_{Q,i}$ is fixed by $\delta_{m,i}$ through equation (5.63).

- **The spherical overdensity**

We now study the evolution of a spherical homogeneous overdensity of radius R in a FLRW background that satisfies the Friedmann equation (5.46). We denote the energy densities of dark matter and quintessence inside the collapsing ball by ρ_m and ρ_Q, respectively. Since dark matter is pressureless, $p_m = 0$ and since quintessence pressure perturbation is negligible, $\delta p_Q \ll \delta\rho_Q$, we can take quintessence pressure to be the unperturbed one $\bar{p}_Q$.

In local coordinates, the evolution of the scale factor R is described by the Euler equation (5.36). Using the appropriate scale factor – i.e., R instead of a – and replacing the potential Φ using eq. (5.30), the divergence of this equation can be written as

$$\frac{\ddot{R}}{R} = -\frac{4\pi G}{3}\left(\rho_m + \rho_Q + 3\bar{p}_Q\right) \ . \tag{5.66}$$

(Note that for a non-clustering quintessence the equation for R is the same with ρ_Q replaced by $\bar{\rho}_Q$ [80].)

For the evolution equations for ρ_m and ρ_Q we use the continuity equation (5.33). Inside the ball this reads, for dark matter,

$$\dot{\rho}_m + 3\frac{\dot{R}}{R}\rho_m = 0\,, \tag{5.67}$$

whose solution is simply

$$\rho_m = \rho_{m,i}\frac{R_i^3}{R^3}\,. \tag{5.68}$$

For dark energy eq. (5.33) becomes

$$\dot{\rho}_Q + 3\frac{\dot{R}}{R}(\rho_Q + \bar{p}_Q) = 0\,, \tag{5.69}$$

which can be rewritten in terms of the nonlinear density contrast $\Delta_Q \equiv \rho_Q/\bar{\rho}_Q - 1$ as

$$\dot{\Delta}_Q + 3\frac{\dot{R}}{R}\Delta_Q - 3\frac{\dot{a}}{a}(1+w)\Delta_Q + 3(1+w)\left(\frac{\dot{R}}{R} - \frac{\dot{a}}{a}\right) = 0\,. \tag{5.70}$$

To solve eqs. (5.66) and (5.70) numerically it is convenient to use y, the radius of the ball normalized to unity at the initial time:

$$y \equiv R/R_i\,, \tag{5.71}$$

and change a and t to the dimensionless variables x, η. Using eq. (5.50), eq. (5.66) can be rewritten as

$$\frac{\mathrm{d}^2 y}{\mathrm{d}\eta^2} + \frac{1}{2}\left[\frac{1+\delta_{m,i}}{x_i^3}\frac{1}{y^2} + (1+3w+\Delta_Q)\frac{1-\Omega_{m,o}}{\Omega_{m,o}}\frac{y}{x^{3(1+w)}}\right] = 0\,, \tag{5.72}$$

where we have used eq. (5.68) and that in the linear regime, where the initial conditions are set, $\rho_m/\bar{\rho}_m|_i = 1+\delta_{m,i}$. Equation (5.70) yields

$$\frac{\mathrm{d}\Delta_Q}{\mathrm{d}\eta} + 3(1+w)\left(\frac{1}{y}\frac{\mathrm{d}y}{\mathrm{d}\eta} - \frac{1+\Delta_Q}{x}\frac{\mathrm{d}x}{\mathrm{d}\eta}\right) + 3\frac{\Delta_Q}{y}\frac{\mathrm{d}y}{\mathrm{d}\eta} = 0\,. \tag{5.73}$$

As initial conditions we have $y_i = 1$ by definition; the expansion rate of a collapsing sphere with dark matter only and in the linear regime can be written as [95]

$$\frac{\dot{R}}{R} = \frac{2}{3t}\left(1 - \frac{1}{3}\delta_m\right)\,, \tag{5.74}$$

which fixes the first derivative of y: $\mathrm{d}y/\mathrm{d}\eta|_i = 2(1-\delta_{m,i}/3)/(3\eta_i)$. For the dark energy perturbation we use that $\Delta_{Q,i}$ is linear at early times, $\Delta_{Q,i} = \delta_{Q,i}$, and thus is fixed in terms of $\delta_{m,i}$ by eq. (5.63).

By solving numerically eq. (5.52) for the background evolution described by x and plugging the result into the coupled eqs. (5.72) and (5.73), one can compute the evolution of R as a function of time t, from the initial time t_i to the time of collapse t_c. The evolution of R is shown in figure 5.2 for four different models: CDM only, ΛCDM, $c_s = 1$ quintessence and $c_s = 0$ quintessence in the cases $w = -0.7$ and $w = -1.3$. We stress that quintessence models with $c_s^2 = 1$ and $w < -1$ are plagued by ghost instabilities and are thus very pathological on short scales; we study them here only for comparison with the $c_s^2 = 0$ case. We assume $\Omega_{m,0}$ and $\Omega_{Q,0}$ (or Ω_Λ) to be the WMAP5 best fit values [96]. We have taken $\delta_{m,i} = 3 \cdot 10^{-4}$ as initial dark matter overdensity at $\eta_i = 10^{-6}$. As expected, since the cosmological constant and the quintessence slow down the evolution of R, the collapse is faster in the pure CDM model. This effect takes place earlier for $w > -1$, as in this case quintessence is more important in the past than the cosmological constant. Thus for $w = -0.7$ the collapse happens later. On the contrary, for a quintessence with $w = -1.3$ collapse takes place earlier. For $w = -0.7$ the collapse is enhanced by the quintessence perturbations and it takes place faster when $c_s^2 = 0$. The opposite happens for $w < -1$, as in this case negative energy clusters, hindering the collapse (see eq. (5.17)).

In general, the time of collapse depends on the value of the initial dark matter overdensity. This is shown in figure 5.3, where the redshift of collapse z_c is plotted as a function of the initial density constrast $\delta_{m,i}$ at the same initial time. As expected, larger overdensities collapse earlier, at higher redshift. For large enough overdensities – and early enough collapse – the redshift of collapse becomes the same for all four different models, because the cosmological constant or the quintessence remain subdominant during the whole process. As expected, quintessence with $w > -1$ requires a larger initial overdensity to collapse and in this case quintessence perturbations with $c_s^2 = 0$ help the collapse. The opposite happens for $w < -1$.

An important quantity to compute in order to derive the mass function is the critical density contrast δ_c, i.e. the density contrast in the linear theory computed at the time when the spherical collapse solution reaches the singularity. Thus, we numerically solve the linear evolution equations for δ_m and δ_Q, eqs. (5.64) and 5.65, and we take δ_c to be δ_m at the time of collapse. In the standard CDM scenario δ_c is given by the well-known number 1.686 [66], independently of the redshift of collapse z_c. However, in the presence of a cosmological constant or quintessence, δ_c depends on the redshift of collapse. Indeed, as the relative abundance of dark matter and dark energy changes with time, the dynamics of the spherical collapse depends on when it takes place. This is shown in figure 5.4, where we plot δ_c as a function of z_c. This result generalizes to quintessence with $c_s^2 = 0$ the standard results obtained for CDM [66], ΛCDM [79] and smooth quintessence [76]. As

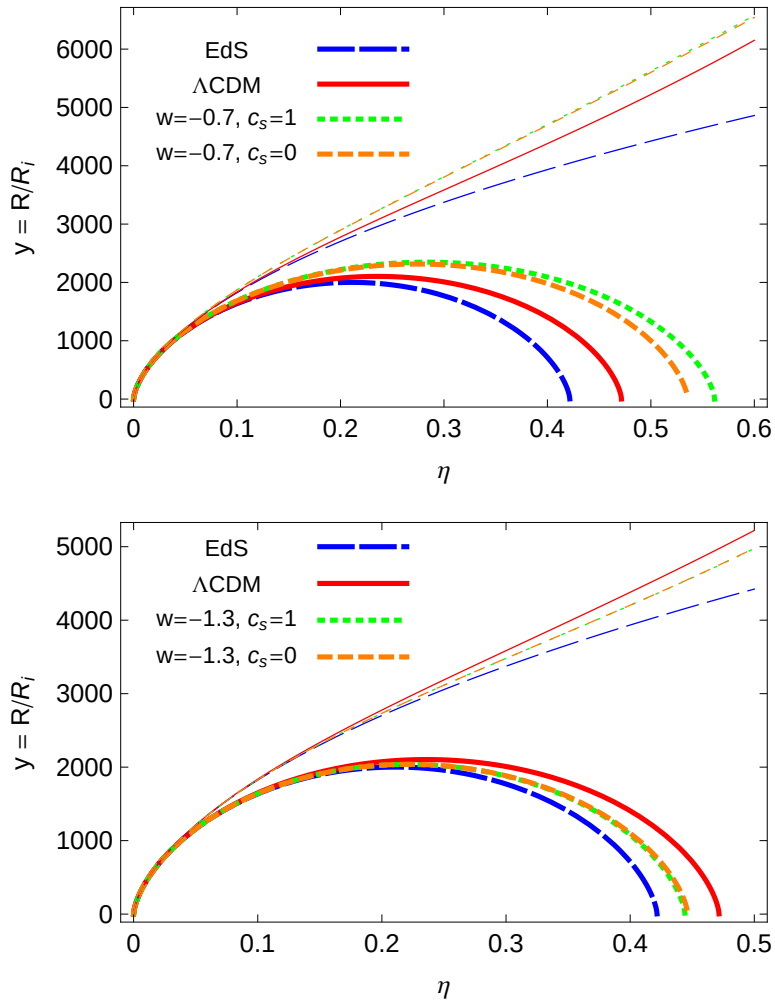


Figure 5.2: Thick lines: time evolution of the radius for a spherical collapse. Thin lines: time evolution following the linearized solutions. The quintessence equation of state is $w = -0.7$ (above) and $w = -1.3$ (below). Starting with the same overdensity, a model with CDM only is the first to collapse. In the upper figure ΛCDM collapses before the quintessence models as dark energy with $w = -0.7$ is more important in the past. The situation is reversed for $w = -1.3$. For $w = -0.7$ the $c_s = 0$ quintessence collapses before $c_s = 1$ as positive energy clusters together with dark matter. For $w = -1.3$ the situation is reversed as negative energy clusters and hinders the collapse. Note that quintessence models with $c_s^2 = 1$ and $w < -1$ are plagued by ghost instabilities and are thus very pathological on short scales. In this figure and in the following ones we study this case only for comparison with the $c_s^2 = 0$ case.

expected, if the collapse takes place early, when the cosmological constant or quintessence are not important, the critical density δ_c will be the same as for CDM. The cosmological constant decreases

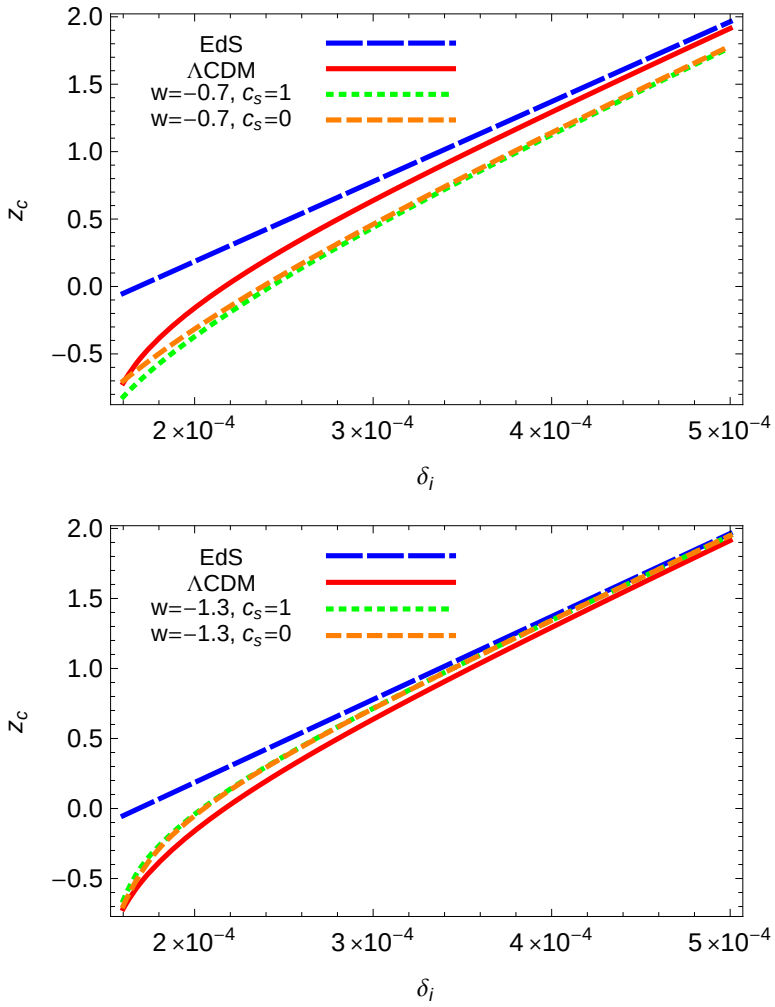


Figure 5.3: Redshift of collapse as a function of the initial overdensity. In the upper figure quintessence models have $w = -0.7$, while they have $w = -1.3$ in the lower one. The behavior follows that explained in figure 5.2.

the value of δ_c and quintessence with $w > -1$, becoming important earlier, decreases it even more. Quintessence perturbations enhance δ_c if $w > -1$. This effect is very mild for $w < -1$ because quintessence becomes important only at very late times.

The change of threshold is very small ($\lesssim 0.5\%$) [84] in all the cases and it is easy to understand why. Let us compare for example a universe with CDM only with a ΛCDM one and let us focus on objects that collapse at a given redshift, say $z = 0$. The initial overdensity must be rather bigger in the ΛCDM case to overcome the acceleration induced by Λ. But the threshold δ_c is obtained

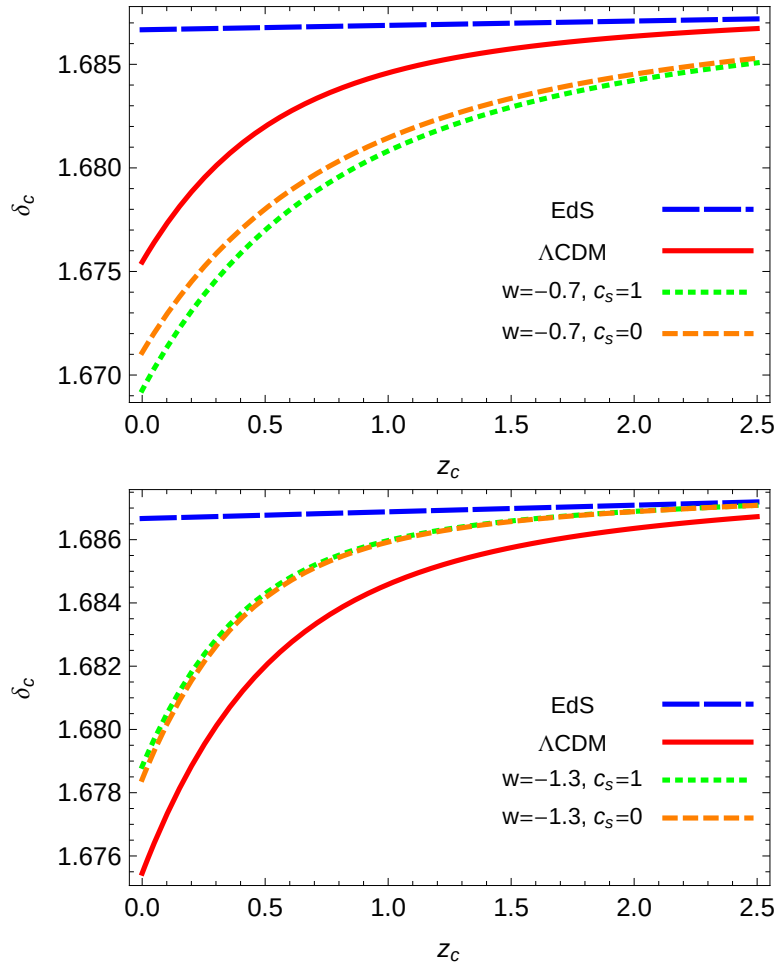


Figure 5.4: *Linear overdensity at collapse as a function of the redshift of collapse. In the upper figure the quintessence models have $w = -0.7$, while they have $w = -1.3$ in the lower one.*

evolving this initial value with the linear transfer functions and this will suppress the ΛCDM value, exactly for the same reason which required it to be bigger in the first place. In other words, the only effect comes from the difference between the linear and non-linear evolution and this causes only a small suppression with respect to the CDM case.

5.4 The mass function of dark matter haloes

We are now ready to discuss the predictions for the mass function using the Press-Schechter formalism [67, 69]. We will first concentrate on the mass of dark matter, leaving aside for the moment

the contribution of the quintessence mass to the halos. The volume density of dark matter halos of mass M is given by

$$\frac{dn_{\rm PS}}{dM}(M,z) = -\sqrt{\frac{2}{\pi}}\frac{\bar{\rho}_m}{M^2}\frac{\delta_c(z)}{D(z)\sigma_M}\frac{d\log\sigma_M}{d\log M}\exp\left[-\frac{\delta_c^2(z)}{2D^2(z)\sigma_M^2}\right] . \tag{5.75}$$

Here σ_M^2 is the smoothed variance of the density field today, which we define with a sharp cut-off in real space

$$\sigma_M^2 \equiv \frac{1}{2\pi^2}\int_0^\infty dk\, k^2 |W(kR)|^2 P_m(k) \qquad \text{with} \qquad M \equiv \frac{4\pi}{3}R^3\bar{\rho}_m\,, \tag{5.76}$$

where P_m is the matter power spectrum and $W(kR) \equiv 3(\sin kR - kR\cos kR)/(kR)^3$ is the Fourier transform of the top-hat window function in real space. Note that the Press-Schechter mass function (5.75) can be rigorously derived only using a sharp filter in Fourier space. Thus its use with a sharp filter in real space is just an approximation; for corrections to this approximation see for example [71] and the discussion in chapter 6. Notice that the redshift dependence of the threshold $\delta_c(z)$ only comes from the spherical collapse dynamics discussed in the previous Section and does not include the growth of the matter power spectrum, which is separately taken into account by the linear growth function $D(z)$.

The linear matter power spectrum is very similar in the cases $c_s = 1$ and $c_s = 0$. The difference comes from the contribution to the Poisson equation of quintessence perturbations as shown in eq. (5.64). The effect is independent of k and it is thus equivalent to a change in the growth function $D(z)$. The change in the growth function can be easily calculated as a function of z by solving eqs (5.64) and (5.65). The result is shown in fig. 5.5. Given that quintessence becomes relevant only recently and that perturbations are suppressed by $1+w$, the effect on dark matter does not exceed the percent level; this result is somewhat smaller than a naïve estimate one can get by comparing the two contributions on the right-hand side of eq. (5.64). As one can see in fig. 5.5, for $1+w > 0$ setting the speed of sound of quintessence to zero fosters the clustering and the dark matter spectrum is slightly enhanced; for $1+w < 0$ quintessence has negative energy and the dark matter spectrum is suppressed. The size of the effect is smaller for $1+w < 0$ as quintessence becomes relevant only very recently.

For the calculation of the mass function, we do not need only the growth rate, but also the complete matter power spectrum. We use for this the publicly available code CAMB [11], which allows to set to zero the speed of sound of quintessence. Apart from w all the other cosmological parameters are set to the WMAP5 best fit values [96]. We have checked that the effect of setting

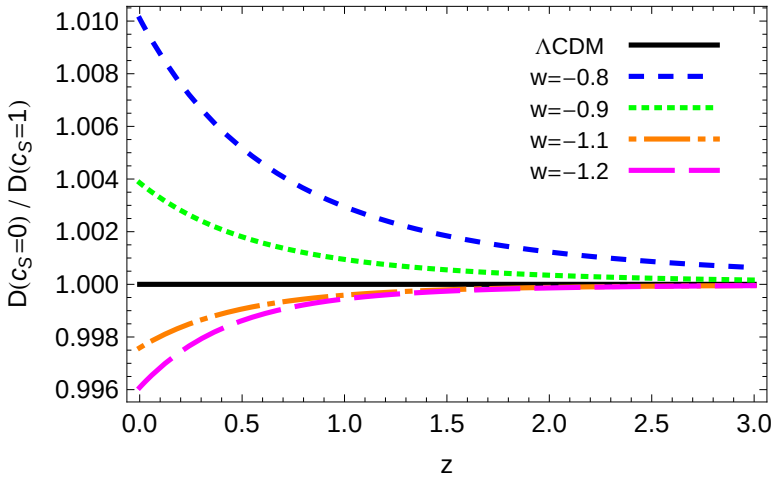


Figure 5.5: Ratio between the growth functions $D(z)$ for $c_s = 0$ and $c_s = 1$ as a function of the redshift z.

$c_s = 0$ instead of $c_s = 1$ in the code is compatible with what we got in fig. 5.5. In figure 5.6 we show the matter power spectrum with $c_s = 0$ and $c_s = 1$ for two different values of w. The speed of sound gives a small effect, much smaller than the modification of the growth rate induced by the different background: for $w > -1$ quintessence becomes relevant before and suppresses the spectrum in comparison with ΛCDM. The opposite effect is obtained in the case $w < -1$.

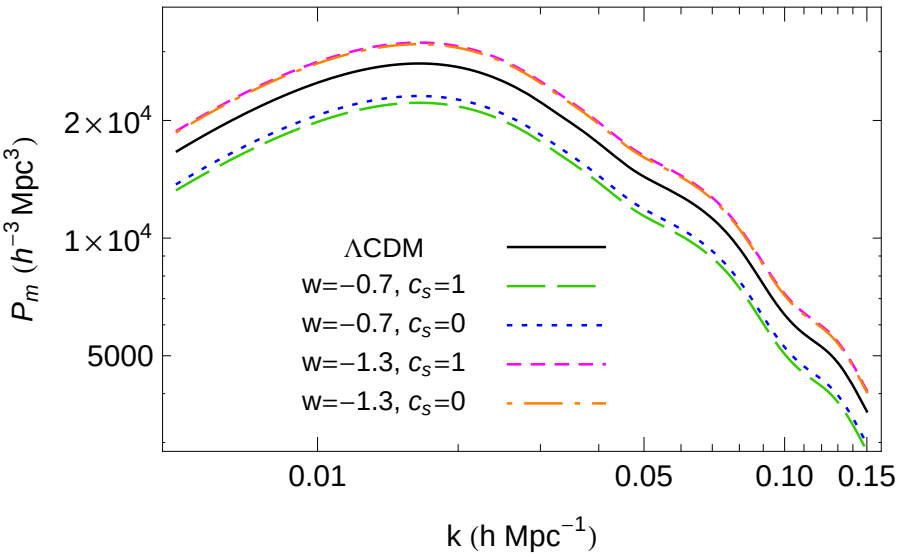


Figure 5.6: Matter power spectrum. The two lines with $w = -1.3$ are almost superimposed.

The equation of state w and the speed of sound of quintessence enter in the mass function eq. (5.75) modifying the growth function $D(z)$ and the threshold for collapse $\delta_c(z)$. Notice that only the combination $\delta_c(z)/D(z)$ enters in the Press-Schechter formula. As we discussed in the previous Section, the change in the threshold is very suppressed and much smaller than the correction of the growth function.

We want to focus on the effect of clustering quintessence on the collapse with respect to the case when quintessence remains unperturbed. This effect can be estimated taking the ratio of the Press-Schechter mass functions in the two cases,

$$\frac{dn_{\rm PS}/dM(w, c_s = 0)}{dn_{\rm PS}/dM(w, c_s = 1)} \,, \tag{5.77}$$

plotted in figure 5.7. We see that the effect is quite small: the ratio becomes large at high masses as a consequence of the exponential dependence of the number density on the mass. We do not dwell on the measurability of this small effect because, as we will discuss in the next section, the additional contribution to the mass of the halo coming from the clustered quintessence will give a comparable change in the mass function.

It is well known that the Press-Schechter formula does not fit in detail the mass function obtained by numerical simulations. A better fit, motivated by the ellipsoidal collapse model, is given by the Sheth-Tormen mass function [74]:

$$\frac{dn_{\rm ST}}{dM}(M, z) = -\sqrt{\frac{2a}{\pi}} A \left[1 + \left(\frac{a\delta_c^2}{D(z)^2\sigma_M^2}\right)^{-p}\right] \frac{\bar{\rho}}{M^2} \frac{\delta_c}{D(z)\sigma_M} \frac{d\log\sigma_M}{d\log M} \exp\left[-a\frac{\delta_c^2}{2D(z)^2\sigma_M^2}\right] \,, \tag{5.78}$$

with $a = 0.707$, $A = 0.322184$ and $p = 0.3$. Since, as discussed, the dependence of the threshold δ_c on the cosmology is very mild in all cases, δ_c is usually taken to be z independent and equal to the EdS value, $\delta_c = 1.686$. Notice that in this way the mss function is "universal", in the sense that the dependence on the cosmological parameters and redshift is only through the smoothed linear density field $D(z)\sigma_M$. It is reasonable to expect that the Sheth-Tormen formula gives a good description of the mass function in the case of non-clustering quintessence. In this case the only effect of quintessence is through the time dependence of the background and its effect on the growth function; this is not qualitatively different from the case of ΛCDM. In other words, we expect "universality" to hold also in this case. On the other hand, we can estimate the effect of clustering quintessence using the ratio described above: in taking the ratio we expect that the shortcomings of the Press-Schechter prescription will partially cancel. Therefore, in figure 5.8 we plot

$$\frac{dn}{dM} \equiv \frac{dn_{\rm ST}}{dM}(w, c_s = 1) \cdot \frac{dn_{\rm PS}/dM(w, c_s = 0)}{dn_{\rm PS}/dM(w, c_s = 1)} \,. \tag{5.79}$$

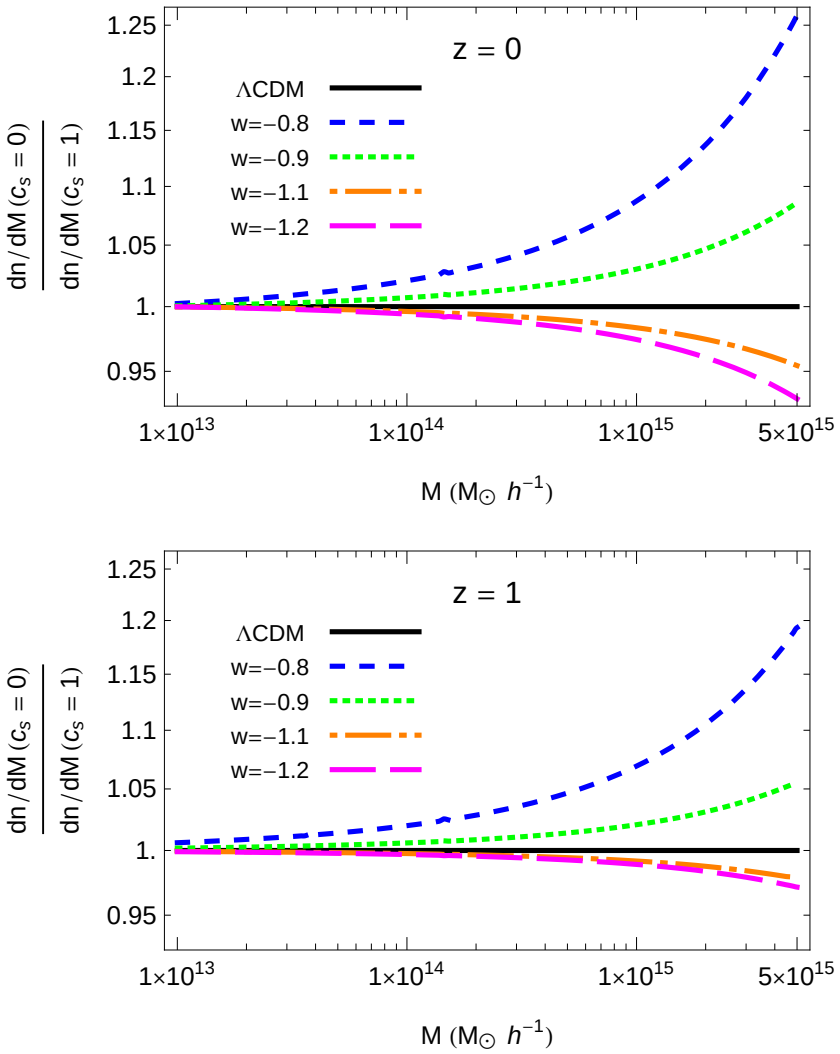


Figure 5.7: Ratio of the Press-Schechter mass function for $c_s = 0$ and $c_s = 1$ at $z = 0$ (above) and $z = 1$ (below).

As expected the main effect is at low redshift and high masses.

5.5 Quintessence contribution to the halo mass

So far we have been interested in the contribution of dark matter to the halo mass function. Given that quintessence with vanishing speed of sound participates in the collapse, one may wonder

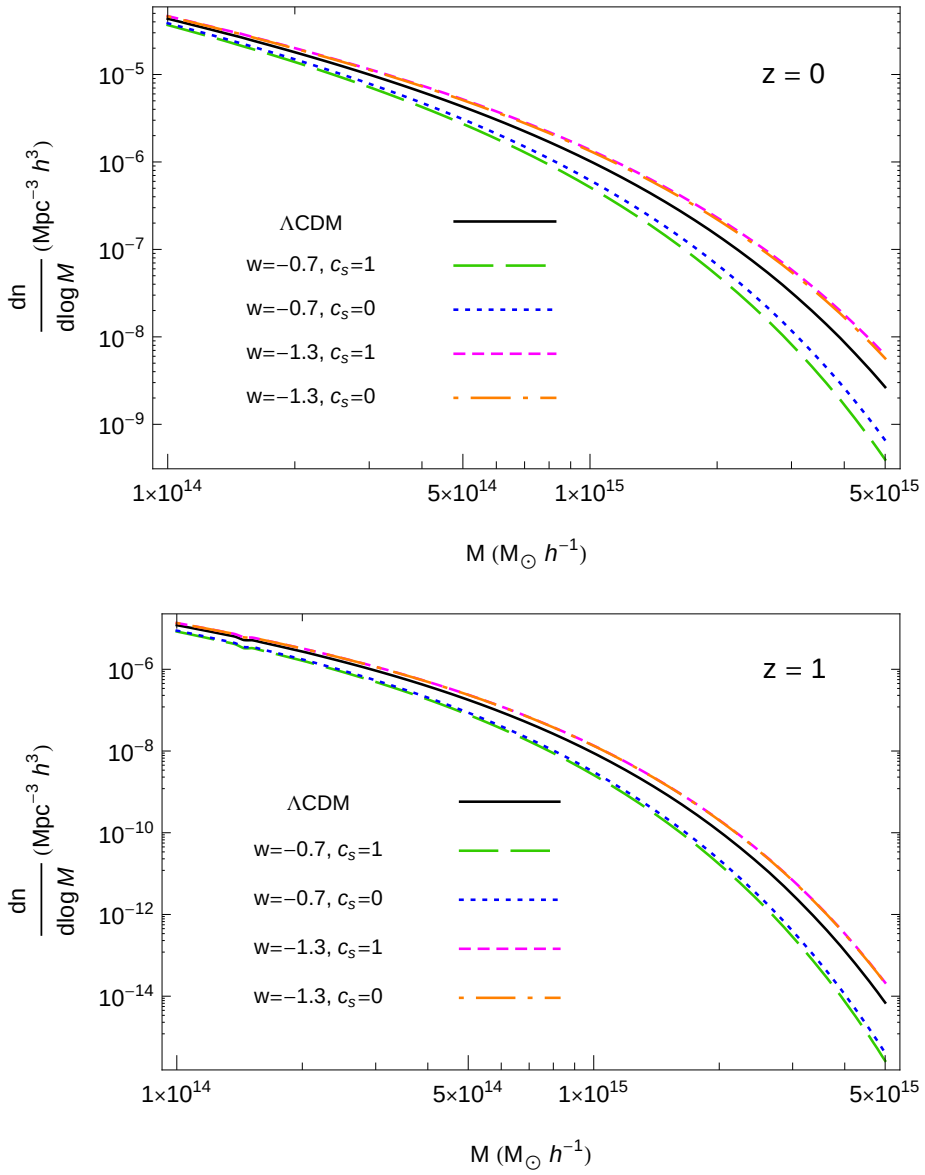


Figure 5.8: Mass function calculated using eq. (5.79) for $z = 0$ (above) and $z = 1$ (below).

whether quintessence will contribute a sizeable amount of mass to the dark matter halo. After all, most of the measurements will be sensitive to the total mass of the object, and not only to the fraction of it associated with dark matter.

5.5 Quintessence contribution to the halo mass

The quintessence contribution to the halo mass can be defined as

$$M_Q \equiv \int_{\rm M} {\rm d}^3x \; \delta\rho_Q \; , \tag{5.80}$$

where the integral is extended over the whole dark matter overdensity. If we stick to the spherical collapse model this just reduces to $(4\pi/3)R^3\delta\rho_Q$. Of course it makes sense to interpret this expression as a contribution to the halo mass, only if it stays practically constant over the time scales of interest, i.e. a Hubble time. In the spherically symmetric case $\delta\rho_Q$ follows equation (5.17):

$$\dot{\delta\rho}_Q + 3\frac{\dot{R}}{R}\delta\rho_Q = -3\delta H(\bar{\rho}_Q + \bar{p}_Q) \, . \tag{5.81}$$

When $|\delta\rho_Q| \gg |1+w|\bar{\rho}_Q$, the right-hand side is negligible. In this limit $\delta\rho_Q$ redshifts as matter so that the integral (5.80) becomes constant, as can be seen in figure 5.9. We expect this condition to be marginally satisfied at turn-around when $\delta\rho_Q \sim (1+w)\bar{\rho}_Q$. This allows us to estimate the quintessence contribution to the halo mass,

$$\frac{M_Q}{M_m} \sim (1+w)\frac{\Omega_Q}{\Omega_m} \; , \tag{5.82}$$

although for a precise estimate one cannot neglect the evolution of the quintessence mass after turn-around. Notice also that quintessence with $1+w<0$ contributes with a negative mass.

Of course, this is only the prediction of the idealized spherical collapse solution. In reality, dark matter haloes virialize with an overdensity ~ 200 times larger than the background. What happens to quintessence while dark matter virializes? Quintessence is exactly comoving with dark matter; eventually dark matter reaches shell crossing and the velocity field ceases to be single-valued. This corresponds to the formation of cusps in the quintessence field, similarly to what discussed in the ghost condensate case [50] and more recently in the context of Hořava-Lifshift gravity [97, 98]. The dynamics of the cusps will depend on higher derivative operators and possibly on the UV completion of the theory[8]. In any case, the dynamics of quintessence in this phase is very complicated [50] and its treatment is the subject of ongoing work. Let us assume however that the cusps are somehow regularized and try to draw some general conclusion that is independent of the details of virialization [9].

[8] For positive $1+w$ one has $c_s^2 > 0$. In this case, one would naïvely expect that quintessence remains smooth on very short scales thus preventing the formation of cusps. However, in our case the velocity of the quintessence fluid (which is the same as the dark matter velocity) exceeds the speed of sound, i.e. it is "supersonic". In this case, sound waves are too slow to prevent the formation of caustics.

[9] We are implicitly assuming that the process responsible for smoothing the cusps does not lead to an energy loss to infinity. For instance, this can happen if the smoothing excites new relativistic degrees of freedom which are radiated away.

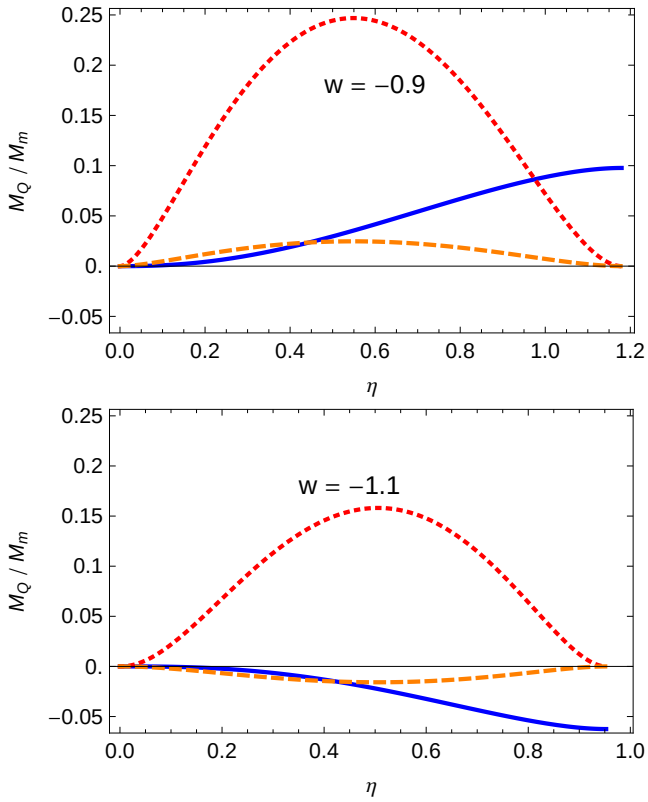


Figure 5.9: Spherical collapse for $w = -0.9$ (above) and $w = -1.1$ (below). Solid blue line: the quintessence contribution to the halo mass $(4\pi/3)R^3\delta\rho_Q$, normalized to the constant dark matter mass, as a function of time. Note that it becomes constant at late time. Red dotted line: the same, but for the background energy density $\bar{\rho}_Q$. Orange dashed: the same for $(1+w)\bar{\rho}_Q$.

As $\delta\rho_Q \simeq (1+w)\bar{\rho}_Q$ at turn-around, when matter starts virialization the inequality $|\delta\rho_Q| \gg |1+w|\bar{\rho}_Q$ is satisfied with rather good approximation: as the virial radius is approximately half of the turn-around radius, $\delta\rho_Q$ has grown by approximately 8 times from the beginning of the collapse while $\bar{\rho}_Q$ has remained approximately constant. The spherical collapse solution indicates that the quintessence mass is, with good approximation, constant at virialization. In the real case we expect the variation of the quintessence mass to be even smaller; in fact, the conservation of the stress-energy tensor tells us that the mass variation is related to the energy flux across a surface around the object,

$$\dot{M} = \oint dS_i \, T_0^i \, . \tag{5.83}$$

As during virialization the velocity field of quintessence will cease to be radial, we expect this

integral to be suppressed with respect to the spherically symmetric case. We conclude that, as the flux integral is negligible, *the mass associated to quintessence stays constant independently of the details of virialization*. Thus, a good estimate of the quintessence mass can be obtained from the spherical collapse model evaluating the quintessence contribution at the virialization radius, see figure 5.10. The ratio between the virialization and the turn-around radii is taken to be the same as in ΛCDM [79][10].

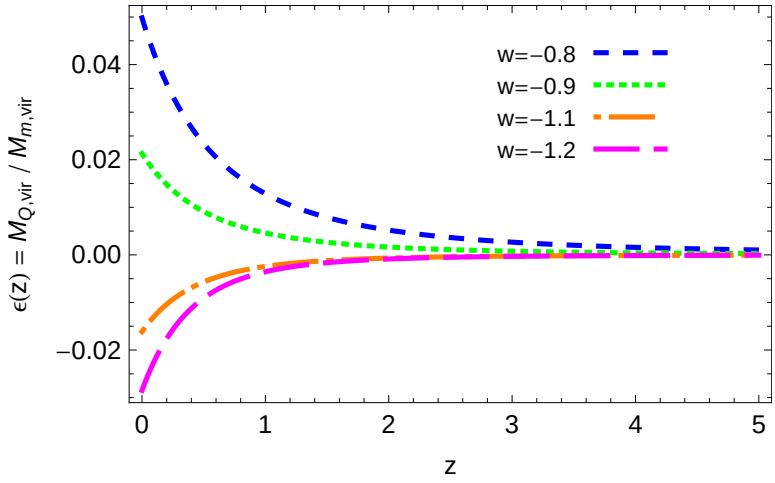


Figure 5.10: The mass contribution of quintessence to the total halo mass, calculated from the spherical collapse solution when the radius reaches the virialization value.

Let us now see how we can take into account the additional quintessence mass in the prediction for the mass function. A rigorous treatment would be quite challenging: the formation of an object of mass M at redshift z should be accompanied by an extra mass $M \to M[1+\epsilon(z)]$, where $\epsilon(z) \equiv M_{Q,vir}/M_{m,vir}$ is the quintessence to dark matter mass ratio at virialization. Then one should follow this extra mass as the halo merges to form larger objects, which in turn accrete extra quintessence as they form. However $\epsilon(z)$ is important only at low redshifts, so that we expect the main effect to be on large objects that formed very recently, the ones on the exponential tail of the mass function. Let us see how we can estimate the effect in this region. At any epoch, the largest objects are mostly created with a negligible rate of destruction through merging to form larger objects. Therefore, the formation rate for these objects can be accurately approximated by

$$-\frac{\partial}{\partial z}\frac{dn_{\rm PS}}{d\log M}(M,z)\ . \tag{5.84}$$

[10]Quintessence clustering will modify the virial radius with respect to ΛCDM with corrections $\mathcal{O}(|1+w|)$. The effect of these corrections on the amount of clustered quintessence is $\mathcal{O}(|1+w|^2)$ and can therefore be safely neglected.

As the effect of quintessence is to rescale the mass of the object as it forms, it is more convenient to use the mass function per logarithmic mass interval. Using this expression, when positive, as an approximation for the formation rate gives

$$\frac{\partial}{\partial z}\frac{dn_{\rm PS,m+Q}}{d\log M}(M,z) = \frac{\partial}{\partial z}\frac{dn_{\rm PS}}{d\log M}\left(M(1-\epsilon(z)),z\right) \ . \tag{5.85}$$

Expanding the right-hand side of this expression and integrating it over the redshift enables us to take into account the extra mass associated with quintessence accreted from an inizial redshift z_i:

$$\begin{aligned}\frac{dn_{\rm PS,m+Q}}{d\log M}(M,z) = &\frac{dn_{\rm PS}}{d\log M}(M,z) \\ &+\int_{z_i}^{z} d\tilde{z}\ \epsilon(\tilde{z})\left[-\frac{\partial}{\partial\log M}\frac{\partial}{\partial\tilde{z}}\frac{dn_{\rm PS}}{d\log M}(M,\tilde{z})\right]\cdot\theta\left(-\frac{\partial}{\partial\tilde{z}}\frac{dn_{\rm PS}}{d\log M}(M,\tilde{z})\right) \ ,\end{aligned} \tag{5.86}$$

where $\theta(x)$ is the Heaviside theta function.

In figure 5.11 we plot the mass functions including the quintessence mass contribution, using the new Press-Schechter mass formula eq. (5.86) into eq. (5.79). Notice that the effect of quintessence mass is to bring the $c_s = 0$ lines closer to the ΛCDM one. To better visualize this effect, in figure 5.12 we plot the ratio between the $c_s = 0$ and $c_s = 1$ case. For $w > -1$ ($w < -1$), setting to zero the speed of sound of quintessence not only does enhance (diminish) the clustering of dark matter, as discussed in the previous sections, but it also adds positive (negative) mass to the halo: the two effects therefore pile up and the second is quantitatively slightly dominant. The sum of the two effects is rather large: for values of w still compatible with the present data and for large masses, the difference between the predictions of the $c_s = 0$ and the $c_s = 1$ cases is of order one.

We stress that the new mass function is not universal, in the sense that there is an explicit red-shift dependence besides the one implicit in the growth of σ_R[11]. This z dependence is quite remarkable: the fact that the modification takes place only at very low red-shift is quite distinctive and a clear link of the effect with the onset of acceleration.

5.6 Three contributions to the mass

In the previous section we saw that a distinctive signature of quintessence with $c_s = 0$ is the extra contribution of quintessence to the mass of virialized objects[12]. Although a detailed study of how

[11] The universality of the mass function, even in the presence of scale independent non-Gaussianity, has been tested with good accuracy in N-body simulations [99].

[12] In this section we will assume this mass to be positive, which is the case for $1 + w > 0$. Similar considerations apply to the $1 + w < 0$ case.

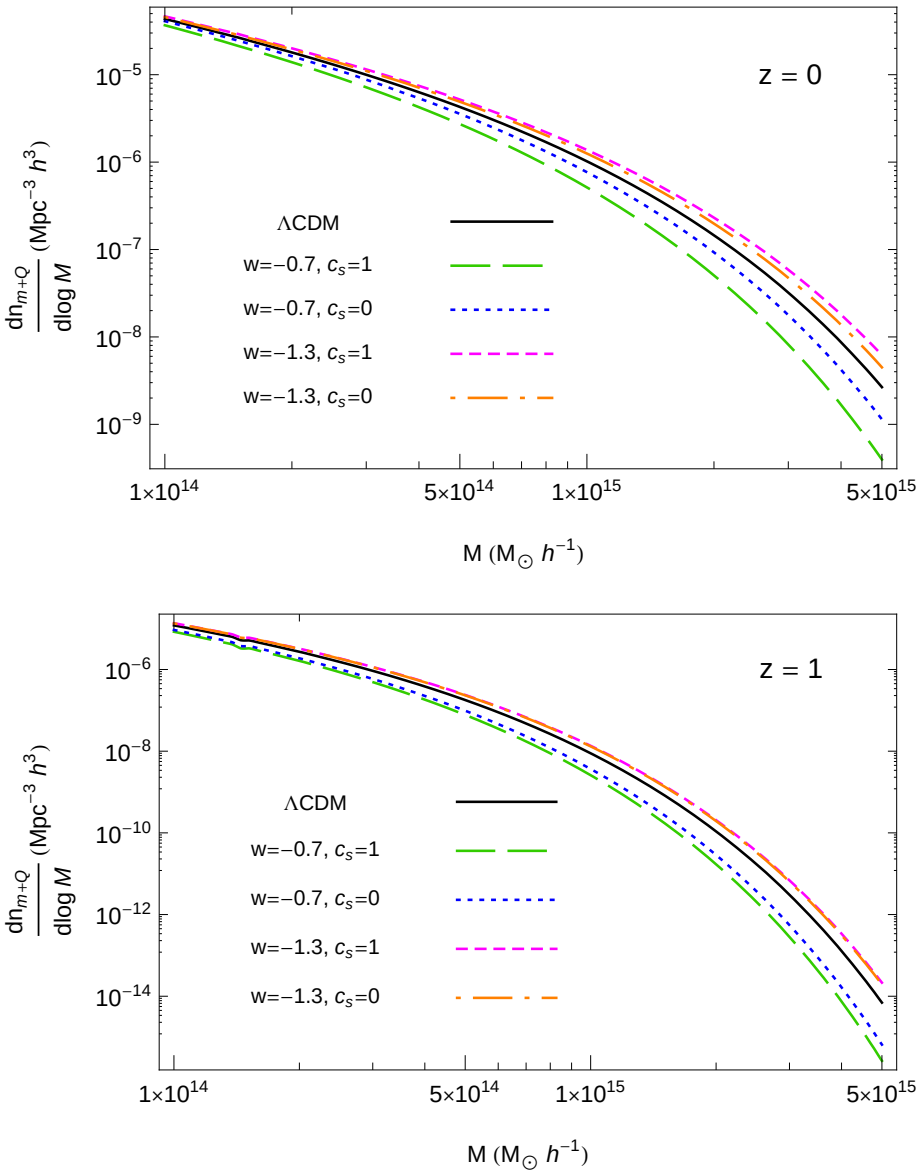


Figure 5.11: Mass function for $z = 0$ (above) and $z = 1$ (below), including the quintessence mass contribution, calculated using eq. (5.86) and (5.79).

to distinguish this extra contribution goes beyond the scope of this thesis, few remarks are in order. Let us focus on galaxy clusters: as discussed in the previous section, the mass of these large objects is significantly affected by the quintessence mass. Moreover, clusters are mostly dominated

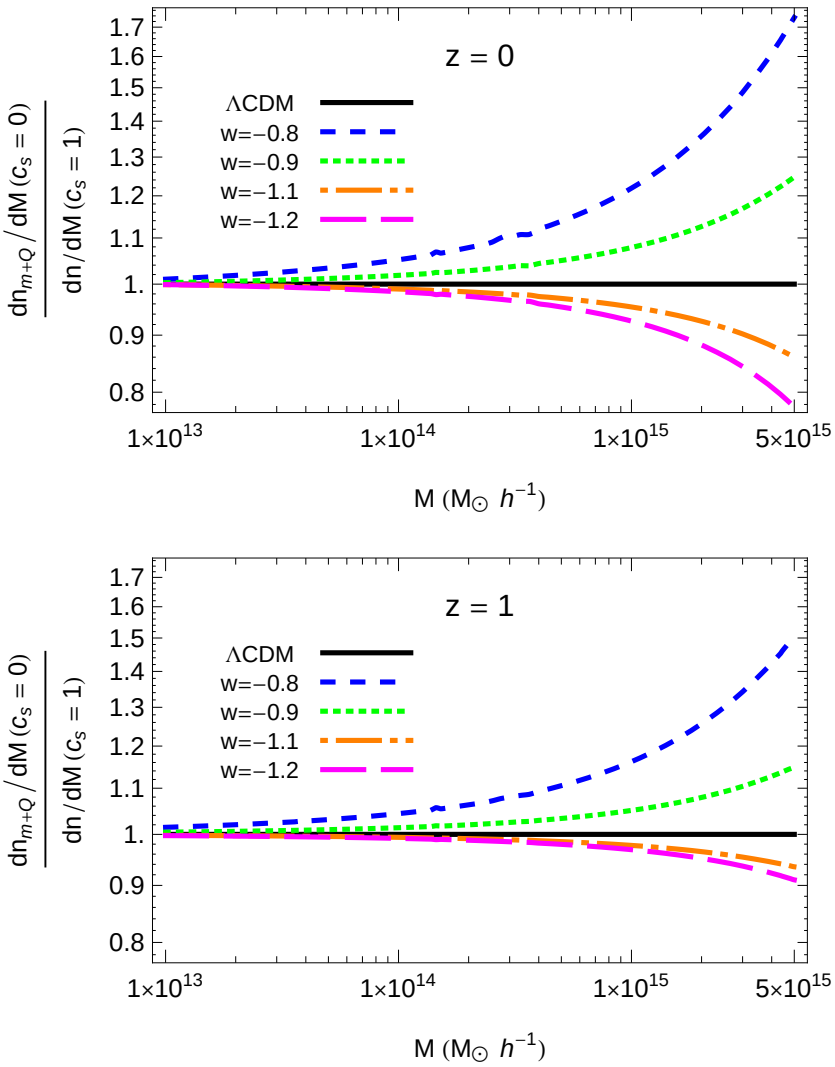


Figure 5.12: Ratio of the Press-Schechter mass function, including the quintessence contribution eq. (5.86), for $c_s = 0$ and $c_s = 1$ at $z = 0$ (above) and $z = 1$ (below).

by gravitational physics (for an introduction to the subject see [100]). If we neglect quintessence for the moment, a cluster is characterized by its baryon mass, mostly in the form of gas, and by the dark matter mass. For sufficiently large clusters, in first approximation one expects the ratio between these two components to be close to the cosmological baryon to dark matter ratio [101].

The various techniques to study clusters have different sensitivities to the two components (for a

review on cluster detection tecniques see [102]). Weak lensing and galaxy velocity measurements are sensitive to the gravitational mass of the object, i.e. to the sum of the dark matter and baryon mass. On the other hand the optical richness, a measure of the light emitted by galaxies inside the cluster, is a probe of the baryon fraction of the mass. X-ray and Sunyaev-Zeldovich (SZ) measurements are sensitive both to the dark matter component and to the baryonic one. While these effects are caused by the gas in the cluster, the temperature and the equilibrium of the intergalactic medium depend on the dark matter distribution. Indeed, assuming hydrostatic equilibrium, one can use the X-ray image of a cluster to reconstruct the baryonic and dark matter radial profiles from its temperature and luminosity distribution, as explained for instance in [103, 104]. The combination of all these techniques should allow a robust reconstruction of the dark matter and baryon components of the clusters.

These standard techniques assume that there is no extra contribution to the cluster mass. However, the crucial property of $c_s = 0$ models of quintessence is that a third kind of mass, associated to quintessence, enters in the game. How are the different tecniques sensitive to these new mass? Lensing and galaxy velocity measurements will still trace the total mass, now including, after dark matter and baryons, also quintessence. On the other hand, optical light associated to galaxies should be rather independent of the accreted quintessence. X-ray experiments should indicate a higher gravitational mass – both from studying the hydrostatic equilibrium and from temperature measurements which are sensitive to the gravitational potential well – but the same baryon component. This effect is similar to a reduction of the $f_{\rm gas}$ parameter, the ratio between baryonic and dark matter mass. Similar considerations apply also to SZ measurements.

A useful lever arm for distinguishing the accretion of dark energy from the other uncertainties in the description of a cluster is the strong red-shift dependence. Quintessence should be relevant only at very low redshift. It would be useful to find out the best combination of observables which is able to constrain the presence of extra mass in the cluster. Besides a smoking gun of clustering quintessence, it would be a useful consistency check between the various measurements.

5.7 Conclusions

Using the spherical collapse model, we have studied how the clustering of quintessence with negligible speed of sound can affect the prediction for the mass function of dark haloes. As quintessence does not develop pressure gradients, it follows geodesic motion remaining comoving with the dark matter. In contrast to the case where quintessence remains smooth, spherical collapsing regions

behave as exact closed FLRW solutions with the quintessence contributing to the overdensity. To study the spherical collapse we found it useful to use Fermi coordinates where the effect of the expansion can be treated as a perturbation around Minkowski spacetime locally around a spatial point. In contrast to comoving coordinates, a unique coordinate system can be employed to describe both the interior and the exterior of the collapsing region.

Quintessence with zero speed of sound modifies dark matter clustering with respect to the smooth quintessence case through the linear growth function and the linear threshold for collapse. Besides these conventional effects there is a more important and qualitatively new phenomenon: quintessence mass adds to the one of dark matter, contributing to the halo mass by a fraction of order $(1+w)\Omega_Q/\Omega_m$. This effect is quite difficult to model and in this work we have adopted a simplified treatment, which gives an accurate estimate of the high mass tail of the distribution where the effect is more relevant.

As dark energy plays an active role in the formation of structures, the distinction between what we call dark matter and dark energy becomes fuzzy. The distinction between the two components can be probed by the different redshift dependence. It is quite remarkable that our knowledge of structure formation still allows a completely inhomogeneous dark energy component on short scales.

In this thesis we do not attempt to study experimental constraints and forecasts. However, we have seen that for values of w which are not too close to the cosmological constant one, let's say $|1+w| \gtrsim 0.1$, the predictions for the high mass tail of the mass function in the $c_s = 0$ and the $c_s = 1$ cases are quite distinctive. The effect of clustering quintessence gives order one changes in the expected number of clusters. Whether the effect will be measurable or not will depend on the possibility of breaking the degeneracy with cosmological parameters, most notably σ_8 and Ω_m, and on the good recostruction of the limiting mass of the survey. Similar concerns have been addressed in the context of the effects of primordial non-Gaussianity on the mass function in [105].

This line of work can be continued in various directions. A better theoretical treatment of quintessence accretion would strengthen our predictions for the mass function, although at a certain point only a (challenging) numerical simulation can make a fully reliable prediction. On the observational side, it would be interesting to understand what is the minimum value of $|1+w|$ that allows a distinction between $c_s = 0$ and $c_s = 1$, using the forthcoming cluster mass function measurements. One should also explore whether other probes, lensing for example, are better suited to study the non-linear clustering of quintessence. It would also be interesting to explore the effect of quintessence on the halo dynamics, but this most probably requires a way to deal with the

formation of cusps in the quintessence field. In a more model-independent way, one could try to use data on clusters to constrain the presence of extra gravitational mass besides the dark matter mass.

Chapter 6

Path integral approach to excursion set theory

We have described the excursion set theory for the mass function calculation, and we used the formalism to predict the modifications of the mass function in quintessence cosmologies in chapter 5. However, the excursion set theory has some problems. Apart from its being or not a fair physical model for the formation and distribution of haloes, there are issues concerning its implementation. In fact, also in the case where only dark matter clusters, we can identify two basic classes of concerns.

The first one stems from the fact that, even assuming that the spherical (or ellipsoidal) collapse is a correct description of halo dynamics, the use of the Fokker-Planck equation to describe the distribution of trajectories is an oversimplification. Moreover, it is not clear how to generalize the approach to different filter functions, which introduce non-Markovianities in the random walk, or to non-Gaussianity of the primordial density fluctuations, which is a generic prediction (and a powerful discriminator) of inflationary models.

The second problem is represented by the use of the simple spherical or ellipsoidal collapse model to describe halo formation. Indeed, dark matter haloes grow through all kinds of processes like smooth accretion, violent encounters and fragmentarions, and even the very definition of what is a halo (both in simulations and in observations) is not unique.

These are the basic motivations that led Maggiore and Riotto (hereafter, MR) to put the excursion set theory on firmer mathematical grounds in a series of papers [71, 72, 73]. The idea is to rewrite the mathematical problem of finding the first passage distribution in terms of a path integral with boundaries, and then taking the (subtle) continuum limit. The formalism just reduces

to the results of Bond *et al.* in the case of a sharp k-space filter and fixed barrier δ_c. However, for a generic filter, it is possible to give a perturbative treatment of the non-Markovian contributions. To cope with the physical limitations of the spherical collapse model and take into account (at least at an effective level) some of the physical complications of halo formation, one can treat the threshold value as a stochastic variable itself.

We will use the path integral formalism to study the effects of primordial non-Gaussianity in chapter 7, improving upon previous approaches in the literature [73, 105, 106]. As an introduction, we will now illustrate the basic ideas and formulae of the path integral approach, for the case of Gaussian density fluctuations, referring to [71, 72].

6.1 General formalism

We consider a stochastic process $\hat{\delta}(S)$, where the density fluctuation field δ will play the role of a "position" variable, while the variance S is to be regarded as a "time" variable. We choose to denote stochastic quantities with a $\hat{}$ to distinguish them from the values they take. The mean is $\langle\hat{\delta}(S)\rangle = 0$ and, since we are considering the Gaussian case, the only non-vanishing correlator is $\langle\hat{\delta}(S_1)\hat{\delta}(S_2)\rangle_c$, where the subscript c stands for connected. A generic trajectory starts at "time" $S_0 = 0$ from an initial position $\hat{\delta}(0) = \delta_0$, and we follow the trajectory until S. The interval $[0, S]$ is discretized in n intervals $\Delta S = S/n$, and we adopt the notation $S_k = k\Delta S$, with $S_n \equiv S$. The trajectory is defined by the set of values it takes at each timestep, $\{\delta_1, \dots, \delta_n\}$, such that $\hat{\delta}(S_k) = \delta_k$, and the δ_i are allowed to range in the whole interval $[-\infty, \infty]$. The probability density in the space of trajectories is

$$W(\delta_0; \delta_1, \dots, \delta_n; S_n) = \langle \delta_D(\hat{\delta}(S_1) - \delta_1) \dots \delta_D(\hat{\delta}(S_n) - \delta_n)\rangle \ , \tag{6.1}$$

where δ_D is the Dirac delta distribution. In terms of W, we can define

$$\Pi_{\Delta S}(\delta_0; \delta_n; S_n) \equiv \int_{-\infty}^{\delta_c} \mathrm{d}\delta_1 \dots \int_{-\infty}^{\delta_c} \mathrm{d}\delta_{n-1} W(\delta_0; \delta_1, \dots, \delta_n; S_n) \ , \tag{6.2}$$

which is the probability that a trajectory, starting from δ_0 at $S = 0$, will be at δ_n at time S_n having never exceeded the threshold δ_c. We will be interested in the continuum limit of $\Pi_{\Delta S}$ as $\Delta S \to 0$, and in the first crossing rate

$$F(S) = -\int_{-\infty}^{\delta_c} \mathrm{d}\delta_n \partial_S \Pi_{\Delta S=0}(\delta_0; \delta_n; S_n) \ . \tag{6.3}$$

In (6.1), one can write the Dirac deltas using the integral representation $\delta_{\rm D}(x) = \int_{-\infty}^{\infty} {\rm d}\lambda e^{-i\lambda x}/2\pi$, to obtain

$$W(\delta_0;\delta_1,\dots,\delta_n;S_n) = \int_{\infty}^{\infty} \frac{{\rm d}\lambda_1}{2\pi}\dots\frac{{\rm d}\lambda_n}{2\pi}\langle e^{-i\sum_j \lambda_j\hat{\delta}(S_j)}\rangle e^{i\sum_j\lambda_j\delta_j} \,. \tag{6.4}$$

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 [107]

$$\langle e^{-i\sum_j\lambda_j\hat{\delta}_j}\rangle = \exp\left[\sum_{p=2}^{\infty}\frac{(-i)^p}{p!}\sum_{j_1,..,j_p=1}^{n}\lambda_{j_1}\dots\lambda_{j_p}\langle\hat{\delta}_{j_1}\dots\hat{\delta}_{j_p}\rangle_c\right] \,, \tag{6.5}$$

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 $\hat{\delta}_j = \hat{\delta}(S_j)$.

In the Gaussian case, all connected n-point correlators vanish except for $n=2$, so that

$$W(\delta_0=0;\delta_1,\dots,\delta_n;S_n) = \int \mathcal{D}\lambda\, e^{i\sum_{i=1}^n\lambda_i\delta_i}e^{-\frac{1}{2}\sum_{i,j=1}^n\lambda_i\lambda_j\langle\hat{\delta}(S_i)\hat{\delta}(S_j)\rangle_c} \,, \tag{6.6}$$

where $\int\mathcal{D}\lambda = \int_{-\infty}^{\infty}{\rm d}\lambda_1/(2\pi)\dots\int_{-\infty}^{\infty}{\rm d}\lambda_n/(2\pi)$. The distribution of trajectories is

$$\Pi_{\Delta S}(\delta_0=0;\delta_n;S_n) = \int_{-\infty}^{\delta_c}{\rm d}\delta_1\dots\int_{-\infty}^{\delta_c}{\rm d}\delta_{n-1}\int\mathcal{D}\lambda\exp\left[i\sum_{i=1}^n\lambda_i\delta_i - \frac{1}{2}\sum_{i,j=1}^n\lambda_i\lambda_j\langle\hat{\delta}(S_i)\hat{\delta}(S_j)\rangle_c\right] \,. \tag{6.7}$$

6.1.1 Sharp k-space filter

If the filter function is chosen to be a sharp filter in k-space, things simplify considerably as $\hat{\delta}(S)$ obeys a Langevin equation with a white noise:

$$\frac{\partial\hat{\delta}(S)}{\partial S} = \hat{\eta}(S) \qquad \langle\hat{\eta}(S_1)\hat{\eta}(S_2)\rangle = \delta_D(S_1-S_2) \,. \tag{6.8}$$

In this case, the path integral can be performed exactly, and the result can be written as

$$W^{\rm gm}(\delta_0;\delta_1,\dots,\delta_n;S_n) = \Psi_{\Delta S}(\delta_n-\delta_{n-1})W^{\rm gm}(\delta_0;\delta_1,\dots,\delta_{n-1};S_{n-1}) \,, \tag{6.9}$$

where

$$\Psi_{\Delta S}(\Delta\delta) = \frac{1}{\sqrt{2\pi\Delta S}}\exp\left[-\frac{(\Delta\delta)^2}{2\Delta S}\right] \,. \tag{6.10}$$

Eq. (6.9) expresses the Markovianity property of the stochastic process, i.e. the fact that the probability of jumping from δ_{n-1} to δ_n depends only on the previous step, but not on the past history of the trajectory. Integrating eq. (6.9) over $\delta_1,\dots,\delta_{n-1}$ we get the important relation

$$\Pi^{\rm gm}_{\Delta S}(\delta_0;\delta_n;S_n) = \int_{-\infty}^{\delta_c}{\rm d}\delta_{n-1}\Psi_{\Delta S}(\delta_n-\delta_{n-1})\Pi^{\rm gm}_{\Delta S}(\delta_0;\delta_{n-1};S_{n-1}) \,, \tag{6.11}$$

which generalizes the well known Chapman-Kolmogorov equation to the discrete time case.

We will now go on to find the continuum limit $\Delta S \to 0$ of the distribution $\Pi^{\rm gm}$, and we will show that this reduces to the results of Bond *et al.*. It is mathematically very difficult to directly perform the path integral in eq. (6.2), since the integrals over $\delta_1, \dots, \delta_{n-1}$ run only up to δ_c, and already the innermost integral yields an error function whose argument involves the next integration variable[1]. A better strategy is to make use of eq. (6.11). Denoting $S_{n-1} = S$, $S_n = S + \Delta S$, $\delta_n = \delta$, $\delta_{n-1} = \delta - \Delta\delta$, it can be written as

$$\Pi^{\rm gm}_{\Delta S}(\delta_0; \delta; S + \Delta S) = \int_{\delta - \delta_c}^{+\infty} {\rm d}(\Delta\delta) \Psi_{\Delta S}(\Delta\delta) \Pi^{\rm gm}_{\Delta S}(\delta_0; \delta - \Delta\delta; S) \; . \tag{6.12}$$

In the limit $\Delta S \to 0$, $\Psi_{\Delta S}(\Delta\delta) \to \delta_D(\Delta\delta)$, so that the zeroeth order in ΔS of eq. (6.12) reads

$$\Pi^{\rm gm}_{\Delta S=0}(\delta_0; \delta; S) = \int_{\delta - \delta_c}^{+\infty} {\rm d}(\Delta\delta) \delta_D(\Delta\delta) \Pi^{\rm gm}_{\Delta S=0}(\delta_0; \delta - \Delta\delta; S) \; . \tag{6.13}$$

If $\delta - \delta_c < 0$, the integral includes the support of the delta function, so that we have the trivial identity that $\Pi^{\rm gm}_{\Delta S=0}(\delta_0; \delta; S)$ is equal to itself. However, if $\delta \geq \delta_c$, we get

$$\Pi^{\rm gm}_{\Delta S=0}(\delta_0; \delta; S) = 0 \qquad \text{if} \quad \delta \geq \delta_c \; , \tag{6.14}$$

which is precisely the absorbing boundary condition postulated in the classical excursion set theory, which naturally emerges in this "microscopic" approach.

For $\delta < \delta_c$, the continuum limit is more involved. If $\delta - \delta_c$ is finite, it is possible to show that the continuum limit gives

$$\frac{\partial \Pi^{\rm gm}(\delta_0; \delta; S)}{\partial S} = \frac{1}{2} \frac{\partial^2 \Pi^{\rm gm}(\delta_0; \delta; S)}{\partial \delta^2} \; , \tag{6.15}$$

which is precisely the Fokker-Planck equation. Since $\Pi^{\rm gm}$ obeys the absorbing boundary condition (6.13), we recover the Bond *et al.* result, which gives the mass function

$$f_{\rm PS}(S) = \sqrt{\frac{2}{\pi}} \frac{\delta_c}{S^{1/2}} \, e^{-\delta_c^2/(2S)} \; . \tag{6.16}$$

[1] A possible approach is to integrate over the δ_i from $-\infty$ to $+\infty$, introducing Heaviside theta functions. Then one should use the integral representation of the theta function, and one gets a double path integral. In principle, this method of calculation can help in the more general case of different filter function or non Gaussian distribution, but it is not clear whether this approach leads to some practical results.

6.2 Generic filter functions

We now examine the case of Gaussian fluctuations, but using a generic filter function. The variance S will be the one calculated with the filter function at hand. In the Markovian case, i.e. when one uses a sharp-k filter, the distribution function of trajectories that never cross the barrier satisfies a local partial differential equation, namely the Fokker-Planck equation. This is not possible when the stochastic process is non-Markovian, as the distribution function depends on the whole history of the trajectories. In fact, take the S_n derivative of eq. (6.7):

$$\frac{\partial}{\partial S_n}\Pi_{\Delta S}(\delta_0;\delta_n;S_n) = \frac{1}{2}\sum_{i,j=1}^{n}\frac{\partial\langle\hat{\delta}_i\hat{\delta}_j\rangle_c}{\partial S_n}\int_{-\infty}^{\delta_c}\mathrm{d}\delta_1\ldots\int_{-\infty}^{\delta_c}\mathrm{d}\delta_{n-1}\partial_i\partial_j W(\delta_0;\delta_1,\ldots,\delta_n;S_n)\,, \tag{6.17}$$

where $\delta_i \equiv \partial/\partial\delta_i$, and we used the fact that ∂_i gives $i\lambda_i$ when acting on $\exp[i\sum_j\lambda_j\delta_j]$. Upon separating the term with $i=j=n$ from the rest, we get

$$\begin{aligned}\frac{\partial}{\partial S_n}\Pi_{\Delta S}(\delta_0;\delta_n;S_n) = \frac{1}{2}\frac{\partial^2}{\partial\delta_n^2}\Pi_{\Delta S}(\delta_0;\delta_n;S_n)\\ +\sum_{j=1}^{n-1}\frac{\partial\langle\hat{\delta}_n\hat{\delta}_j\rangle_c}{\partial S_n}\partial_n\int_{-\infty}^{\delta_c}\mathrm{d}\delta_1\ldots\int_{-\infty}^{\delta_c}\mathrm{d}\delta_{n-1}\partial_j W(\delta_0;\delta_1,\ldots,\delta_n;S_n)\,,\end{aligned} \tag{6.18}$$

Now, since the upper limit of the integrals is δ_c and not $+\infty$, the integral of the derivative $\partial_j W$ with $j<n$ gives a non-vanishing boundary term. The sharp-k filter is special because the 2-point correlator is $\langle\hat{\delta}_n\hat{\delta}_j\rangle_c = \min(S_n,S_j) = S_j$, which is independent of S_n for $j<n$. Therefore $\partial/\partial S_n\langle\hat{\delta}_n\hat{\delta}_j\rangle_c = 0$, and we are left with a local partial differential equation. For a generic form of the 2-point correlator, however, the second term on the RHS of eq. (6.18) does not vanish. Moreover, in the continuum limit the sum over S_j becomes an integral over the intermediate "time" variable S_k, which means that this term is non-local in "time", depending on the whole past history of the trajectory.

In the following, we focus on the real space top-hat filter, since it is the one most commonly used in the literature, but the method can be applied more generally. The two point function can be written as

$$\langle\hat{\delta}(S_i)\hat{\delta}(S_j)\rangle = \min(S_i,S_j) + \Delta(S_i,S_j)\,, \tag{6.19}$$

and an excellent approximation to the function $\Delta(S_i,S_j)$ is

$$\Delta_{jk} \simeq \kappa\min(S_j,S_k)\left(1-\frac{\min(S_j,S_k)}{\max(S_j,S_k)}\right)\,, \tag{6.20}$$

where $\kappa(R) \simeq 0.464+0.002R$, with R measured in h^{-1} Mpc, using the 7-year WMAP parameters [1] and the transfer function given in [108], as explained in chapter 7. Using (6.19) in eq. (6.7), and expanding to first order in Δ_{ij}, we find

$$\Pi^{\Delta_1}_{\Delta S}(\delta_0;\delta_n;S_n) = \frac{1}{2}\sum_{i,j}\Delta_{ij}\int_{-\infty}^{\delta_c}\mathrm{d}\delta_1\ldots\int_{-\infty}^{\delta_c}\mathrm{d}\delta_{n-1}\partial_i\partial_j W^{\mathrm{gm}}(\delta_0;\delta_1,\ldots,\delta_n;S_n)\,. \tag{6.21}$$

When passing to the continuum limit, the sum becomes an integral over the past history of the trajectories. There are some subtleties in taking the continuum limit of the resulting expressions, because one needs to compute the behaviour close to the barrier of the function $\Pi^{\mathrm{gm}}_{\Delta S}$, which is not analytic in ΔS. The final result can be split into a "memory" term Π^{mem}, which depends on one integral over one intermediate time variable, and a "memory-of-memory" term $\Pi^{\mathrm{mem-mem}}$, which depends on a double integral over two intermediate times:

$$\Pi^{\mathrm{mem}}_{\Delta S=0} = \kappa\partial_n\left[\frac{\delta_c(\delta_c-\delta_n)}{S_n}\,\mathrm{erfc}\left(\frac{2\delta_c-\delta_n}{\sqrt{2S_n}}\right)\right]\,, \tag{6.22}$$

$$\Pi^{\mathrm{mem-mem}}_{\Delta S=0} = \frac{\kappa\delta_c}{\sqrt{2\pi S_n}}\partial_n\left[e^{-(\delta_c-\delta_n)^2/(2S_n)}\int_0^{S_n}\frac{\mathrm{d}S_i}{S_i}e^{-\delta_c^2/(2S_i)}\,\mathrm{erfc}\left((\delta_c-\delta_n)\sqrt{\frac{S_i}{2S_n(S_n-S_i)}}\right)\right]\,. \tag{6.23}$$

Observe that neither of the two functions Π^{mem} and $\Pi^{\mathrm{mem-mem}}$ vanishes in $\delta_n=\delta_c$. However, their sum does, so that Π^{Δ_1} still satisfies the absorbing barrier boundary condition $\Pi_{\Delta S=0}(\delta_0;\delta_c;S)=0$.

It is now straightforward to compute the first crossing rate and the halo mass function up to first order in the non-Markovianities. Using (6.3), we find for the first crossing rate

$$F^{\mathrm{mem}}(S) = 0\,, \tag{6.24}$$

$$F^{\mathrm{mem-mem}}(S) = -\frac{\kappa\delta_c}{\sqrt{2\pi}S^{3/2}}\left[e^{-\delta_c/(2S)} - \frac{1}{2}\Gamma\left(0,\frac{\delta_c^2}{2S}\right)\right]\,, \tag{6.25}$$

where $\Gamma(0,z)$ is the incomplete Gamma function. Putting together eqs. (4.20, 6.24, 6.25) and using the definition of the halo mass function eq. (4.13), we find, up to first order in the non-Markovian corrections,

$$f(\nu) = (1-\kappa)\left(\frac{2}{\pi}\right)^{1/2}\nu\,e^{-\nu^2/2} + \frac{\kappa}{\sqrt{2\pi}}\nu\Gamma\left(0,\frac{\nu^2}{2}\right)\,. \tag{6.26}$$

A first check of the result eq.(6.26) is that it satisfies the normalization condition

$$\int_0^\infty\frac{\mathrm{d}\sigma}{\sigma}f(\sigma) = 1\,. \tag{6.27}$$

In fact, it is easy to see that, integrating eq. (6.26), the κ dependence cancels out and the normalization is correct.

6.3 The diffusing barrier

The excursion set theory based on the spherical collapse model fails at the quantitative level. Taking into account a generic filter function, as in eq. (6.26), does not improve much the agreement with N-body simulations. In fact, eq. (6.26) is everywhere lower than the PS mass function, and therefore, while it improves the agreement with N-body simulations at low masses, it is an even worse result at high masses. One can think that the ellipsoidal collapse model [109, 110] will help improving the agreement with simulations. In this case, the threshold acquires a dependence on σ:

$$B(\sigma) \simeq \delta_c \left[1 + 0.47 \left(\frac{\delta_c}{\sigma} \right)^{1.23} \right] , \tag{6.28}$$

which reflects the fact that low mass haloes (large σ) have larger deviations from sphericity and significant shear, that opposes collapse. Therefore, low-mass haloes require an higher density to collapse, and the barrier recedes away from δ_c at larger σ. On the contrary, high-mass haloes tend to be more spherical. Therefore, it is apparent from eq. (6.28) that the ellipsoidal collapse model will not help in improving the agreement with simulations at high masses: the barrier is always higher than δ_c, and the mass function will be sistematically lower than the PS one.

There is some crucial physical ingredient missing in the model. Actually, already in the simple ellipsoidal model, the value of the barrier is given by (6.28) only if one uses the most probable values for the eigenvalues of the collapsing ellipsoid. In general, the ellipsoidal collapse predicts a "fuzzy" threshold [111, 112], with a distribution that can extend even to values lower than δ_c. Actually, the possible sources of stochasticity of the barrier are different. For instance, haloes are subject to tidal effects from their environment, which give a distribution of values for the barrier [113]. Substructures within a collapsing halo can also influence the critical value for collapse. Last but not least, the very definition of what is a dark matter halo is a delicate problem, both in observations and in simulations. For instance, in simulations haloes are usually identified either with the friends-of-friends (FOF) algorithm, which track isodensity profiles, or with the spherical overdensity (SO) algorithm, which singles out spherical regions with a given overdensity. There is no fundamental physical reason to prefer one or the other algorithm, the choice being largely a matter of convenience for the particular problem at hand. This introduces an additional source of uncertainty in the threshold value, which will fluctuate around the value predicted by the ellipsoidal model with fluctuations due to different reasons, some of which are not simply predictable.

Therefore, a sensible thing to do is to extend excursion set theory to take into account a barrier which is itself a stochastic variable. We will consider an "effective" approach, supposing that the

barrier performs a random walk with a diffusion coefficient D_B. All our ignorance of the halo formation dynamics and our uncertainties in the definition of a halo are buried into D_B, which can then be fitted using N-body simulations, or treated as a nuisance parameter when comparing to observations. For the sake of simplicity, we will consider in the following a barrier which fluctuates around the spherical collapse value $\langle B\rangle = \delta_c$. We expect this to be a good approximation for large masses, and in particular this will be useful for the computation of non-Gaussian effects in 7. The variance is taken to be the one of a Brownian motion, and for this reason this model will be referred to as the diffusing barrier model:

$$\Sigma_B \equiv \langle (B - \langle B\rangle)^2\rangle^{1/2} = \sqrt{D_B}\sigma \ . \tag{6.29}$$

We will first neglect the filter effects, which will be taken into account later. We denote by $\Pi(\delta_0, \delta; B_0, B; S)$ the joint probability distribution that, at "time" S, the density contrast has reached the value δ starting from the initial value δ_0 and the barrier has reached the value B starting from the initial value $B_0 = \delta_c$. The density contrast and the barrier are uncorrelated stochastic variables and they diffuse independently: this means that Π satisfies the 2-dimensional Fokker-Planck equation

$$\frac{\partial \Pi}{\partial S} = \frac{1}{2}\frac{\partial^2 \Pi}{\partial \delta^2} + \frac{D_B}{2}\frac{\partial^2 \Pi}{\partial B^2} \ . \tag{6.30}$$

The initial conditions are $\Pi(S = 0) = \delta_D(\delta - \delta_0)\delta_D(B - \delta_c)$, while the boundary condition is that $\Pi(S)$ vanishes when $\delta(S) = B(S)$. The solution to the problem (6.30) can be found using the method of images [114]. The final distribution for the density is marginalized over the barrier, by integrating $\mathrm{d}B$ from δ to $+\infty$. The result is

$$\Pi^{\mathrm{gm}}(\delta; S) = \frac{1}{2\sqrt{2\pi S}}\left[e^{-\delta^2/(2S)}\,\mathrm{erfc}\left(-\frac{\delta_c - \delta}{\sqrt{2D_B S}}\right) - e^{-(2a\delta_c - \delta)^2/(2S)}\,\mathrm{erfc}\left(-\frac{2aD_B\delta_c - (\delta_c - \delta)}{\sqrt{2D_B S}}\right)\right] , \tag{6.31}$$

where $a \equiv 1/(1 + D_B)$. The first crossing rate is now given by

$$F^{\mathrm{gm}}(S) = \int_{-\infty}^{+\infty} \mathrm{d}\delta \frac{\partial \Pi^{\mathrm{gm}}}{\partial S} = \frac{\sqrt{a}\delta_c}{\sqrt{2\pi}S^{3/2}} e^{-a\delta_c^2/(2S)} \ . \tag{6.32}$$

The mass function $f(\sigma) = 2\sigma^2 F(\sigma^2)$ is then

$$f^{\mathrm{gm}}(\sigma) = \sqrt{\frac{2}{\pi}}\frac{\sqrt{a}\delta_c}{\sigma} e^{-a\delta_c^2/(2\sigma^2)} \ , \tag{6.33}$$

and it is apparent that the effect of a diffusing barrier is simply to replace $\delta_c \to \sqrt{a}\delta_c$. Notice that this is the same modification postulated by [115] in order to better fit the N-body simulations.

A simpler way to understand this result, which can be immediately generalized to the non-Markovian case, is to observe that the problem of a barrier with coordinate x_1 diffusing with a diffusion coefficient D_1, and a particle with coordinate x_2 diffusing with a diffusion coefficient D_2 is equivalent to a one degree of freedom problem [114], introducing the relative coordinate $x = x_2 - x_1$. The resulting stochastic motion is governed by an effective diffusion coefficient $D_{\rm eff} = D_1 + D_2$. In fact, consider the Langevin equations for both variables:

$$\dot{\hat{x}}_i = \hat{\eta}_i(t) \ , \qquad \text{with} \quad \langle \hat{\eta}_i(t)\hat{\eta}_i(t')\rangle = D_i \delta_D(t-t') \tag{6.34}$$

where $i = 1, 2$. Then, the relative coordinate satisfies $\dot{\hat{x}} = \hat{\eta}(t)$ with $\hat{\eta}(t) = \hat{\eta}_2(t) - \hat{\eta}_1(t)$. If $\hat{\eta}_1$ and $\hat{\eta}_2$ are uncorrelated, we find

$$\langle \hat{\eta}(t)\hat{\eta}(t')\rangle = \langle \hat{\eta}_1(t)\hat{\eta}_1(t')\rangle + \langle \hat{\eta}_2(t)\hat{\eta}_2(t')\rangle = (D_1 + D_2)\delta_D(t-t') \ , \tag{6.35}$$

which shows that the effective coefficient is $D_1 + D_2$. In our case, $D_1 = D_B$ and $D_2 = 1$, so that the first crossing distribution is the usual result (4.20), with the replacement $\delta_c \to \delta_c/(1+D_B)$.

For the non-Markovian case, using the method of the previous section and reasoning in an analogous way, the first crossing distribution is

$$f(\nu) = (1-\tilde{\kappa})\frac{2}{\pi}\frac{\sqrt{a}\delta_c}{\sigma} e^{-a\delta_c^2/(2\sigma^2)} + \frac{\tilde{\kappa}}{\sqrt{2\pi}}\frac{\sqrt{a}\delta_c}{\sigma}\Gamma(0, \frac{a\delta_c^2}{2\sigma^2}) \ , \tag{6.36}$$

where $\tilde{\kappa} \equiv \kappa/(1+D_B)$.

Chapter 7

An Improved Calculation of the Non-Gaussian Mass Function

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 [14, 15] 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 [16] one finds the following constraints: $-29 < f_{NL}^{loc} < +69$, already comparable with those obtained from CMB measurements in [1]: $-10 < f_{NL}^{loc} < +74$. The future is even more promising, with precisions of

$\Delta f_{NL}^{loc} \sim 10$ [116] and $\Delta f_{NL}^{loc} \sim 1$ [117] 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 [118]. The mass function, which is the focus of this chapter, has been used for example in [116] together with the scale dependent bias to produce forecasts for future surveys, and in [119] in an attempt to explain the presence of a very massive cluster at a large redshift as an indication of a large NG. Some reviews summarizing recent results on these topics are [120, 121].

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 $M + dM$:

$$\frac{dn}{dM} = \frac{\bar{\rho}}{M^2} f(\sigma) \left| \frac{d \ln \sigma}{d \ln M} \right| , \tag{7.1}$$

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. In the following, we will refer to $f(\sigma)$ itself as the mass function. Our calculations will be based on the path integral formulation of the excursion set theory, as introduced in chapters 4 and 6.

Turning to non-Gaussianities, the most popular non-Gaussian mass functions are those due to Matarrese, Verde and Jimenez [106] (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 [115]. It is not clear however that this is the correct way to proceed. Indeed, in a series of papers [71, 72, 73], Maggiore & Riotto (MR) presented a rigorous approach to the first-passage problem in terms of path integrals, and in [73] 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 $f_{\rm NL}$. Since the NG are assumed to be small, in the sense that the parameter $\epsilon \equiv \langle \delta^3 \rangle / \sigma^3$ 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, which allowed them to *non-perturbatively* account for the $\epsilon\nu^3$ term (which appears in an exponential in their approach).

It appears to us, therefore, that there is considerable room for improvement in the theoretical calculation of the mass function. The goal of our work is twofold. Firstly, we present a rigorous calculation of the mass function in the following way : we use the techniques developed by MR in [71, 72, 73], which allow us to track the complex multi-scale correlations involved in the calculation, and 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 [73, 105, 106]. 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 different expressions.

In this work we will not explicitly account for effects of the ellipsoidal collapse model [109, 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 [122].

This chapter is organized as follows. In section 7.1 we fix some notation and briefly introduce the two most popular shapes of primordial NG, i.e. the local and equilateral ones. In section 7.2 we present our calculation of the mass function. In section 7.3 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 7.4 we discuss the effects induced by some additional complications introduced in the problem due to the specific choice of the filter function [71], 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 [72]. In section 7.5 we compare our final result (7.60) 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 Appendices C and D.

7.1 Models of non-Gaussianity

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

$$\Phi(\mathbf{k}, z) = -\frac{3}{5} T(k) \frac{D(z)}{a} \mathcal{R}(\mathbf{k}) \,, \tag{7.2}$$

where $\mathcal{R}(\mathbf{k})$ is the (comoving) curvature perturbation, which stays constant on superhorizon 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

$$\begin{aligned} \delta(\mathbf{k}, z) &= -\frac{2ak^2}{3\Omega_m H_0^2} \Phi(\mathbf{k}, z) = \frac{2k^2}{5\Omega_m H_0^2} T(k) D(z) \mathcal{R}(\mathbf{k}) \\ &\equiv \mathcal{M}(k, z) \mathcal{R}(\mathbf{k}) \,, \end{aligned} \tag{7.3}$$

where we substituted eq. (7.2). Here, Ω_m is the present time fractional density of matter (cold dark matter and baryons), and $H_0 = 100\, h\, \mathrm{km\, s^{-1} Mpc^{-1}}$ 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 [1], 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.* [109]:

$$T_{\mathrm{BBKS}}(x) \equiv \frac{1}{2.34x} \ln(1 + 2.34x) \left(1 + 3.89x + (16.1x)^2 + (5.46x)^3 + (6.71x)^4\right)^{-1/4} \,, \tag{7.4}$$

where $x \equiv k(h\,\mathrm{Mpc}^{-1})/\Gamma$ with a shape parameter $\Gamma = \Omega_m h \exp\left[-\Omega_b(1 + \sqrt{2h}/\Omega_m)\right]$ that accounts for baryonic effects as described in [108]. For more accurate results, one could use a numerical transfer function, as obtained by codes like CMBFAST [10] or CAMB [11]; 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 $W_R(|\mathbf{x}|)$, and consider the smoothed density field (around one point, which we take as the origin):

$$\delta_R = \int \frac{\mathrm{d}^3k}{(2\pi)^3}\widetilde{W}(kR)\delta(\mathbf{k})\,, \tag{7.5}$$

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

$$\widetilde{W}(y) = \frac{3}{y^3}\left(\sin y - y\cos y\right)\,. \tag{7.6}$$

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, as discussed in chapter 6, which we will comment on later. By using (7.5) and (7.3) we have, for the 3-point function,

$$\langle\delta_{R_1}\delta_{R_2}\delta_{R_3}\rangle_c = \int \frac{\mathrm{d}^3k_1}{(2\pi)^3}\frac{\mathrm{d}^3k_2}{(2\pi)^3}\frac{\mathrm{d}^3k_3}{(2\pi)^3}\widetilde{W}(k_1R_1)\widetilde{W}(k_2R_2)\widetilde{W}(k_3R_3)\times \\ \times\,\mathcal{M}(k_1)\mathcal{M}(k_2)\mathcal{M}(k_3)\langle\mathcal{R}(\mathbf{k}_1)\mathcal{R}(\mathbf{k}_2)\mathcal{R}(\mathbf{k}_3)\rangle_c\,, \tag{7.7}$$

where the subscript c denotes the connected part; analogous formulae are valid for the higher order correlations.

7.1.1 Shapes of non-Gaussianity

The function $\langle\mathcal{R}(\mathbf{k}_1)\mathcal{R}(\mathbf{k}_2)\mathcal{R}(\mathbf{k}_3)\rangle_c$ encodes information about the physics of the inflationary epoch. By translational invariance, it is proportional to a momentum-conserving delta function:

$$\langle\mathcal{R}(\mathbf{k}_1)\mathcal{R}(\mathbf{k}_2)\mathcal{R}(\mathbf{k}_3)\rangle_c = (2\pi)^3\delta_D(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)B_{\mathcal{R}}(k_1,k_2,k_3)\,, \tag{7.8}$$

where the (reduced) bispectrum $B_{\mathcal{R}}(k_1,k_2,k_3)$ 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 [123]. We will now describe these two models in a bit more detail.

7.1.1.1 The local model

The local bispectrum is produced when the NG is generated outside the horizon, for instance in the curvaton model [124, 125] or in the inhomogeneous reheating scenario [126]. In these models, the curvature perturbation can be written in the following form:

$$\mathcal{R}(\mathbf{x}) = \mathcal{R}_g(\mathbf{x}) + \frac{3}{5} f_{\rm NL}^{\rm loc} \left(\mathcal{R}_g^2(\mathbf{x}) - \langle \mathcal{R}_g^2 \rangle \right) + \frac{9}{25} g_{\rm NL} \mathcal{R}_g^3(\mathbf{x}) \,, \tag{7.9}$$

where $\mathcal{R}_g$ 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

$$B_{\mathcal{R}}(k_1, k_2, k_3) = \frac{6}{5} f_{\rm NL}^{\rm loc} \left[P_{\mathcal{R}}(k_1) P_{\mathcal{R}}(k_2) + \text{cycl.} \right] \,, \tag{7.10}$$

where "cycl." denotes the 2 cyclic permutations of the wavenumbers, and $P_{\mathcal{R}}(k)$ is the power spectrum given by $P_{\mathcal{R}}(k) = A k^{n_s - 4}$. The trispectrum is given by

$$\begin{aligned} \langle \mathcal{R}(\mathbf{k}_1) \mathcal{R}(\mathbf{k}_2) \mathcal{R}(\mathbf{k}_3) \mathcal{R}(\mathbf{k}_4) \rangle_c &= (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4) \\ &\times \left[\frac{36}{25} f_{\rm NL}^2 \sum_{\substack{b<c \\ a \neq b,c}} P_{\mathcal{R}}(|\mathbf{k}_a + \mathbf{k}_b|) P_{\mathcal{R}}(k_b) P_{\mathcal{R}}(k_c) + \frac{54}{25} g_{\rm NL} \sum_{a<b<c} P_{\mathcal{R}}(k_a) P_{\mathcal{R}}(k_b) P_{\mathcal{R}}(k_c) \right] \,. \end{aligned} \tag{7.11}$$

7.1.1.2 The equilateral model

Models with derivative interactions of the inflaton field [127, 128, 129] 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 [130] (see also [131]):

$$B_{\mathcal{R}}(k_1, k_2, k_3) = \frac{18}{5} f_{\rm NL}^{\rm equil} A^2 \left[\frac{1}{2 k_1^{4-n_s} k_2^{4-n_s}} + \frac{1}{3 (k_1 k_2 k_3)^{2(4-n_s)/3}} - \frac{1}{(k_1 k_2^2 k_3^3)^{(4-n_s)/3}} + 5 \text{ perms.} \right] . \tag{7.12}$$

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.

7.2 Random walks and the halo mass function

We now turn to the main calculation of our work. The non-Gaussian halo mass function can be obtained by calculating the barrier first crossing rate 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 [71, 73] and the mass function can be shown to be $f(\sigma) = 2\sigma^2 F(\sigma)$. 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 (7.6) 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, 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 7.4.

Extending the results of chapter 6, 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

$$f = -2S\,\frac{\partial}{\partial S}\bigg|_{\delta_c} \int_{-\infty}^{\delta_c} \mathrm{d}\delta_1 \ldots \mathrm{d}\delta_n \exp\bigg[-\frac{1}{3!}\sum_{j,k,l=1}^{n} \langle \hat{\delta}_j \hat{\delta}_k \hat{\delta}_l \rangle_c \partial_j \partial_k \partial_l + \frac{1}{4!}\sum_{j,k,l,m=1}^{n} \langle \hat{\delta}_j \hat{\delta}_k \hat{\delta}_l \hat{\delta}_m \rangle_c \partial_j \partial_k \partial_l \partial_m + \ldots \bigg] W^{\mathrm{gm}}\,, \tag{7.13}$$

where W^{gm} is given by the expression (6.9), and 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 (δ_c, S) to (ν, S), in which case the partial derivative becomes

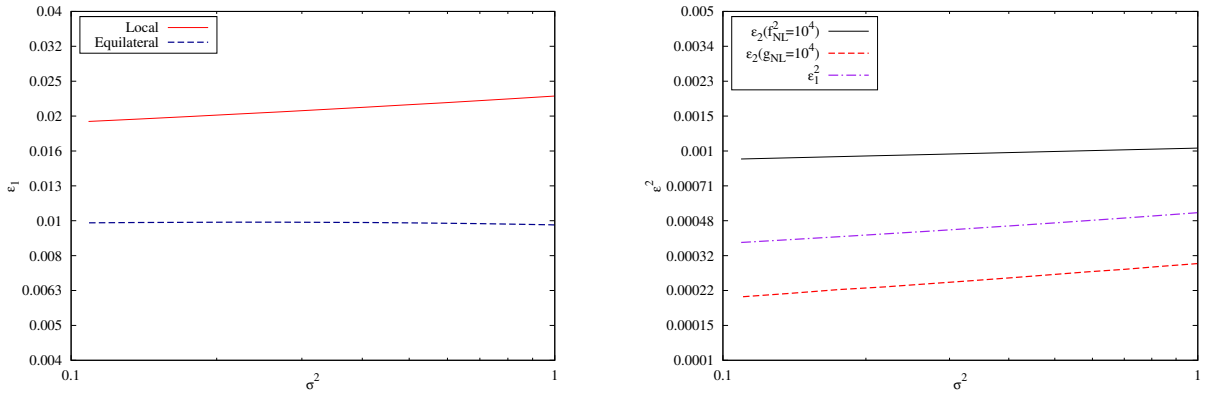
$$-2S(\partial/\partial S)|_{\delta_c} = \nu(\partial/\partial\nu)|_S - 2S(\partial/\partial S)|_\nu \equiv \nu\partial_\nu - 2S\partial_S\,. \tag{7.14}$$

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

$$\varepsilon_{n-2} \equiv \frac{\langle \hat{\delta}_R^n \rangle_c}{\sigma_R^n} \quad ; \quad n \geq 3\,, \tag{7.15}$$

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 $\varepsilon_1 \sim f_{\mathrm{NL}} A^{1/2}$, $\varepsilon_2 \sim g_{\mathrm{NL}} A$, etc[1]. Typically we expect $\epsilon \lesssim 10^{-2}$ for $f_{\mathrm{NL}} \lesssim 100$. Figure 7.1 shows the behaviour of ε_1 and ε_2 in the local and equilateral models, as a

[1]Notationally we distinguish the order parameter ϵ from the specific NG functions ε_1 and ε_2.



(a) (b)

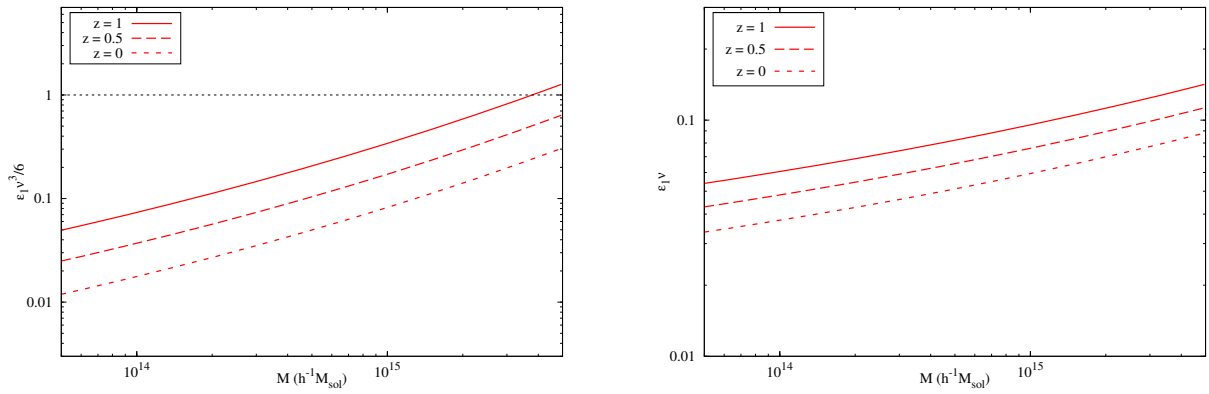
Figure 7.1: Scale dependence of the ε_n. Panel (a) : Behaviour of ε_1 vs. σ^2 in the local and equilateral models, for $f_{\rm NL} = 100$ in each case. Panel (b) : Behaviour of ε_2 for the local model with $f_{\rm NL} = 100$ and $g_{\rm NL} = 10^4$. The terms proportional to $f_{\rm NL}^2$ and $g_{\rm NL}$ are shown separately. Also shown is ε_1^2 for the same model. The axes are logscale.

function of $S = \sigma_R^2$. 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, where they are expected to be scale-independent. However, as we see from figure 7.1, 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 $\epsilon\nu^3 \sim \mathcal{O}(1)$. Figure 7.2 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 $f_{\rm NL}^{\rm loc} = 100$. The behaviour for the equilateral NG is qualitatively similar. The redshift dependence of these quantities comes from the definition of ν,

$$\nu(M,z) \equiv \sqrt{a}\frac{\delta_{c0}}{\sigma(M)}\frac{D(0)}{D(z)} \equiv \frac{\delta_c(z)}{\sigma(M)}\,, \tag{7.16}$$

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) (b)

Figure 7.2: Panel (a) : Behaviour of $\varepsilon_1\nu^3/6$ vs. mass in the local non-Gaussian model, for $f_{\rm NL} = 100$. 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).

a different from unity (specifically $\sqrt{a} \simeq 0.89$) can be motivated by allowing the collapse threshold to vary stochastically [72], as we will discuss in section 7.4. 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 7.3 we will discuss the effects of relaxing this condition and probing smaller length scales.

We now turn to the "unequal time" correlators appearing in (7.13). Since we are concerned with large scales, we are in the small S limit, and following MR we expand the n-point correlators around the "final time" S. We can define the Taylor coefficients

$$\mathcal{G}_3^{(p,q,r)}(S) \equiv \left[\frac{d^p}{dS_j^p}\frac{d^q}{dS_k^q}\frac{d^r}{dS_l^r}\langle\hat{\delta}(S_j)\hat{\delta}(S_k)\hat{\delta}(S_l)\rangle_c\right]_{S_j=S_k=S_l=S}, \tag{7.17}$$

and then expand

$$\langle\hat{\delta}_j\hat{\delta}_k\hat{\delta}_l\rangle_c = \sum_{p,q,r=0}^{\infty}\frac{(-1)^{p+q+r}}{p!q!r!}\mathcal{G}_3^{(p,q,r)}(S)(S-S_j)^p(S-S_k)^q(S-S_l)^r\,. \tag{7.18}$$

For the 4-point function we will have an analogous expression involving coefficients $\mathcal{G}_4^{(p,q,r,s)}$.

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[2]

$$\langle\hat{\delta}_j\hat{\delta}_k\hat{\delta}_l\rangle_c = \varepsilon_1(S_jS_kS_l)^{1/2} \;\;;\;\; \langle\hat{\delta}_j\hat{\delta}_k\hat{\delta}_l\hat{\delta}_m\rangle_c = \varepsilon_2(S_jS_kS_lS_m)^{1/2}\,. \tag{7.19}$$

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:

$$\begin{aligned}\mathcal{G}_3^{(1,0,0)} &= \frac{1}{2}\varepsilon_1(S)c_1(S)S^{1/2} \;\;;\;\; \mathcal{G}_3^{(2,0,0)} = -\frac{1}{4}\varepsilon_1(S)c_2(S)S^{-1/2}\,,\\ \mathcal{G}_3^{(1,1,0)} &= \frac{1}{4}\varepsilon_1(S)c_3(S)S^{-1/2} \;\;;\;\; \mathcal{G}_4^{(1,0,0,0)} = \frac{1}{2}\varepsilon_2(S)c_4(S)S\,,\end{aligned} \tag{7.20}$$

where the coefficients $c_n(S)$ 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 (7.19). Figure 7.3 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 $f_{\rm NL}$ 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.

Using the first few terms of the unequal time expansions, in our toy model one can write

$$\begin{aligned}\sum_{j,k,l=1}^{n}\langle\hat{\delta}_j\hat{\delta}_k\hat{\delta}_l\rangle_c\,\partial_j\partial_k\partial_l = \varepsilon_1 S^{3/2}\Bigg(&\sum_{j,k,l=1}^{n}\partial_j\partial_k\partial_l - \frac{3}{2}\sum_{j=1}^{n}(1-\frac{S_j}{S})\partial_j\sum_{k,l=1}^{n}\partial_k\partial_l\\ &-\frac{3}{8}\sum_{j=1}^{n}(1-\frac{S_j}{S})^2\partial_j\sum_{k,l=1}^{n}\partial_k\partial_l + \frac{3}{4}\sum_{j,k=1}^{n}(1-\frac{S_j}{S})(1-\frac{S_k}{S})\partial_j\partial_k\sum_{l=1}^{n}\partial_l + \dots\Bigg)\,,\end{aligned} \tag{7.21}$$

$$\sum_{j,k,l,m=1}^{n}\langle\hat{\delta}_j\hat{\delta}_k\hat{\delta}_l\hat{\delta}_m\rangle_c\,\partial_j\partial_k\partial_l\partial_m = \varepsilon_2 S^2\Bigg(\sum_{j,k,l,m=1}^{n}\partial_j\partial_k\partial_l\partial_m - 2\sum_{j=1}^{n}(1-\frac{S_j}{S})\partial_j\sum_{k,l,m=1}^{n}\partial_k\partial_l\partial_m + \dots\Bigg)\,. \tag{7.22}$$

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 (7.13) 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

[2]Throughout the paper we will consider at most 4-point correlators. This truncation is justified given our assumptions, as we will see later.

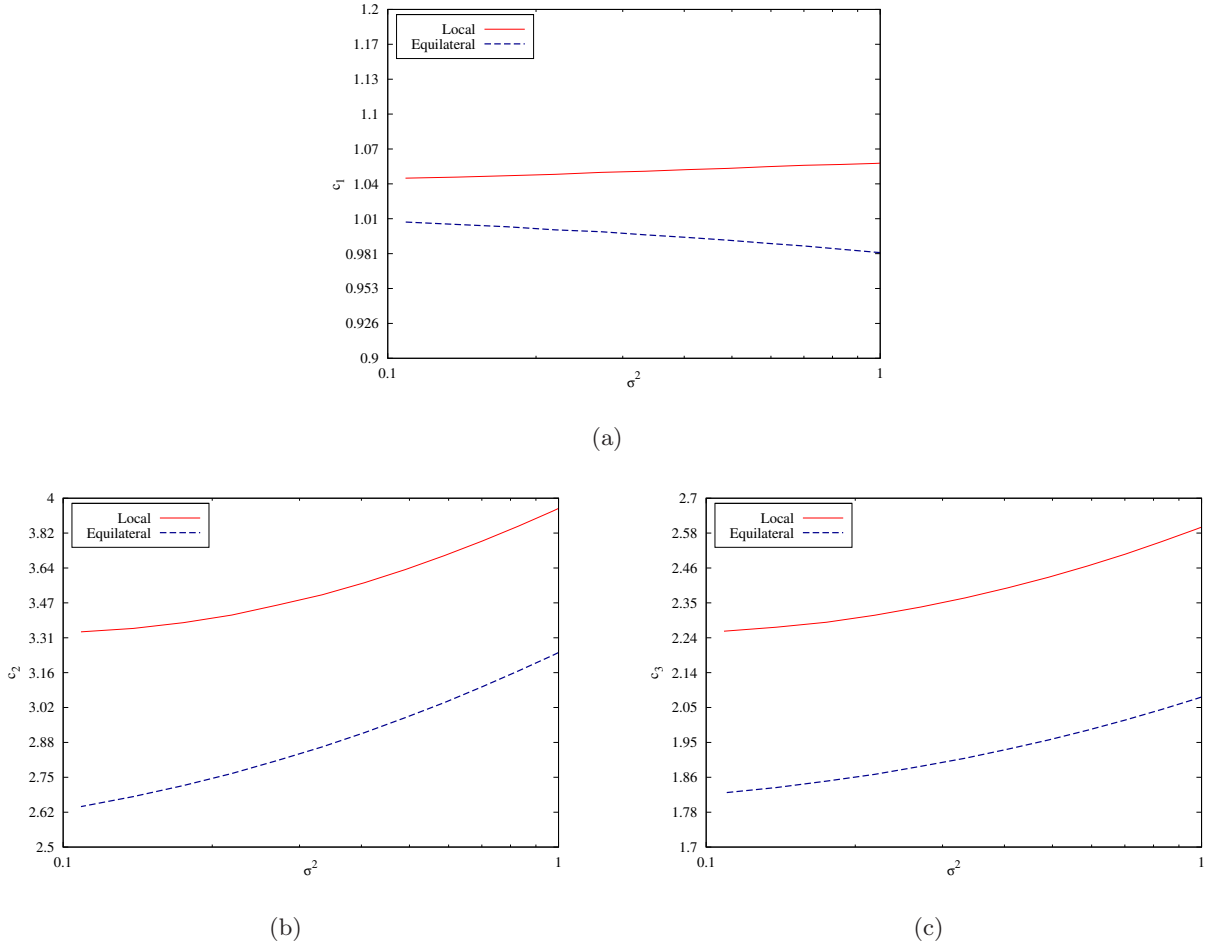


Figure 7.3: The derivative coefficients c_1 (panel (a)), c_2 (panel (b)) and c_3 (panel (c)), as a function of σ^2, for local and equilateral NG models. These quantities are independent of redshift and the NG amplitudes $f_{\rm NL}$ and $g_{\rm NL}$. The axes are logscale.

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 7.3 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 (7.13) involves the multiple integral of $W^{\rm gm}$. The operator $\nu\partial_\nu$ acts on the error function to give the Gaussian rate. 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

$$\int_{-\infty}^{\delta_c}\mathrm{d}\delta_1\dots\mathrm{d}\delta_n\sum_{j=1}^{n}\partial_j g=\frac{\partial}{\partial\delta_c}\int_{-\infty}^{\delta_c}\mathrm{d}\delta_1\dots\mathrm{d}\delta_n g\,, \tag{7.23}$$

Using this, and the fact that $S^{1/2}(\partial/\partial\delta_c)|_S=\partial_\nu|_S$, we see that the leading term in (7.21) (i.e. the term with no powers of $(1-S_j/S)$), leads to a term involving

$$\varepsilon_1\nu\partial_\nu(\partial_\nu)^3\,\mathrm{erf}\left(\nu/\sqrt{2}\right)\sim f_{\rm PS}\,\varepsilon_1\,\nu^3(1+\mathcal{O}(\nu^{-2}))\,,$$

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 (7.23). We see that the entire exponential operator $\exp[-(\varepsilon_1 S^{3/2}/3!)\sum_{j,k,l=1}^{n}\partial_j\partial_k\partial_l]$ in (7.13) can be pulled across the multiple integral and converted to $\exp\left[-(\varepsilon_1/3!)\partial_\nu^3\right]$ acting on the remaining integral. Similarly, the operator $\exp\left[(\varepsilon_2 S^2/4!)\sum_{j,k,l,m=1}^{n}\partial_j\partial_k\partial_l\partial_m\right]$ can be pulled out and converted to $\exp\left[(\varepsilon_2/4!)\partial_\nu^4\right]$, 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

$$\begin{aligned}f=\nu\,e^{-(\varepsilon_1/3!)\partial_\nu^3+(\varepsilon_2/4!)\partial_\nu^4+\dots}\partial_\nu\int_{-\infty}^{\delta_c}\mathrm{d}\delta_1\dots\mathrm{d}\delta_n\exp\Bigg[\frac{1}{3!}\varepsilon_1 S^{3/2}\Bigg(\frac{3}{2}\sum_{j=1}^{n}(1-\frac{S_j}{S})\partial_j\sum_{k,l=1}^{n}\partial_k\partial_l\\ +\frac{3}{8}\sum_{j=1}^{n}(1-\frac{S_j}{S})^2\partial_j\sum_{k,l=1}^{n}\partial_k\partial_l-\frac{3}{4}\sum_{j,k=1}^{n}(1-\frac{S_j}{S})(1-\frac{S_k}{S})\partial_j\partial_k\sum_{l=1}^{n}\partial_l+\dots\Bigg)\\ -\frac{1}{4!}\varepsilon_2 S^2\Bigg(2\sum_{j=1}^{n}(1-\frac{S_j}{S})\partial_j\sum_{k,l,m=1}^{n}\partial_k\partial_l\partial_m+\dots\Bigg)\Bigg]W^{\rm gm}\,.\end{aligned} \tag{7.24}$$

Now consider the action of the individual terms in the remaining exponential under the integrals, but without exponentiation. From MR [73], we have the following results[3],

$$\sum_{j=1}^{n}(1-\frac{S_j}{S})\sum_{k,l=1}^{n}\int_{-\infty}^{\delta_c}\mathrm{d}\delta_1\dots\mathrm{d}\delta_n\partial_j\partial_k\partial_l W^{\rm gm}=\left(\frac{2}{\pi}\right)^{1/2}\frac{1}{S^{3/2}}e^{-\nu^2/2}\,, \tag{7.25a}$$

$$\sum_{j=1}^{n}(1-\frac{S_j}{S})^2\sum_{k,l=1}^{n}\int_{-\infty}^{\delta_c}\mathrm{d}\delta_1\dots\mathrm{d}\delta_n\partial_j\partial_k\partial_l W^{\rm gm}=\left(\frac{2}{\pi}\right)^{1/2}\frac{3}{S^{3/2}}h(\nu)\,, \tag{7.25b}$$

$$\sum_{j,k=1}^{n}(1-\frac{S_j}{S})(1-\frac{S_k}{S})\sum_{l=1}^{n}\int_{-\infty}^{\delta_c}\mathrm{d}\delta_1\dots\mathrm{d}\delta_n\partial_j\partial_k\partial_l W^{\rm gm}=\left(\frac{2}{\pi}\right)^{1/2}\frac{4}{S^{3/2}}h(\nu)\,, \tag{7.25c}$$

[3]The terms in (7.25a), (7.25b) and (7.25c) are, upto prefactors, the integrals of what MR denote as $\Pi^{(3,\rm NL)}$, $\Pi^{(3,\rm NNLa)}$ and $\Pi^{(3,\rm NNLb)}$ respectively in [73].

where we have defined

$$h(\nu) \equiv e^{-\nu^2/2} - \left(\frac{\pi}{2}\right)^{1/2} \nu \,\mathrm{erfc}\,\frac{\nu}{\sqrt{2}} = \frac{\nu}{2^{3/2}}\Gamma\left(-\frac{1}{2}, \frac{\nu^2}{2}\right) , \tag{7.26}$$

where $\Gamma(-1/2, \nu^2/2)$ is an incomplete Gamma function. Let us focus on the term in (7.25a). If we linearize in ε_1 in (7.24), then this term appears with $\varepsilon_1 t^{3/2}\partial_\nu$ 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 (7.25b) and (7.25c) 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 $\varepsilon_1 \,\mathrm{erfc}\,\nu/\sqrt{2}$ 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 C. 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 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 (7.25b) and (7.25c) 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:

$$\sum_{j=1}^{n}(1-\frac{S_j}{S})\sum_{k,l,m=1}^{n}\int_{-\infty}^{\delta_c}\mathrm{d}\delta_1\ldots\mathrm{d}\delta_n\partial_j\partial_k\partial_l\partial_m W^{\rm gm} = -\left(\frac{2}{\pi}\right)^{1/2}\frac{1}{S^2}\nu\, e^{-\nu^2/2}\,, \tag{7.27a}$$

$$\sum_{j,k=1}^{n}(1-\frac{S_j}{S})(1-\frac{S_k}{S})\sum_{l,l_1,l_2,l_3=1}^{n}\int_{-\infty}^{\delta_c}\mathrm{d}\delta_1\ldots\mathrm{d}\delta_n\partial_j\partial_k\partial_l\partial_{l_1}\partial_{l_2}\partial_{l_3} W^{\rm gm} = -\left(\frac{2}{\pi}\right)^{1/2}\frac{4}{S^3}\nu\, e^{-\nu^2/2}\,, \tag{7.27b}$$

where we have used the result (7.23), and in (7.27b) also the identity

$$\partial_\nu^3 h(\nu) = -\nu\, e^{-\nu^2/2}\,. \tag{7.28}$$

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

$$f = \left(\frac{2}{\pi}\right)^{1/2} \nu\, e^{-(\varepsilon_1/3!)\partial_\nu^3+(\varepsilon_2/4!)\partial_\nu^4+\ldots}\left[e^{-\nu^2/2} - \frac{1}{4}\varepsilon_1\nu\, e^{-\nu^2/2} + \frac{5}{16}\varepsilon_1\left(\frac{\pi}{2}\right)^{1/2}\mathrm{erfc}\,\frac{\nu}{\sqrt{2}} + \frac{1}{8}\left(\varepsilon_1^2 - \frac{2}{3}\varepsilon_2\right)e^{-\nu^2/2}\left(\nu^2-1\right) + \mathcal{O}(\epsilon^3\nu^3)\right], \tag{7.29}$$

where we ignore terms like $\varepsilon_1\nu\mathcal{O}(\nu^{-4})$.

To compute the action of the exponentiated derivative operators, we start by writing the expression in square brackets in (7.29) in terms of its Fourier transform, using the relations[4]

$$\begin{aligned} e^{-\nu^2/2} &= \int_{-\infty}^{\infty}\frac{\mathrm{d}\lambda}{\sqrt{2\pi}}e^{i\lambda\nu}e^{-\lambda^2/2}\,,\\ -\nu\, e^{-\nu^2/2} &= \int_{-\infty}^{\infty}\frac{\mathrm{d}\lambda}{\sqrt{2\pi}}(i\lambda)e^{i\lambda\nu}e^{-\lambda^2/2}\,,\\ \nu^2 e^{-\nu^2/2} &= -\int_{-\infty}^{\infty}\frac{\mathrm{d}\lambda}{\sqrt{2\pi}}(\lambda^2-1)e^{i\lambda\nu}e^{-\lambda^2/2}\,,\\ \left(\frac{\pi}{2}\right)^{1/2}\mathrm{erfc}\,\frac{\nu}{\sqrt{2}} &= \int_{-\infty}^{\infty}\frac{\mathrm{d}\lambda}{\sqrt{2\pi}}\frac{i}{\lambda}e^{i\lambda\nu}e^{-\lambda^2/2}\,. \end{aligned} \tag{7.30}$$

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, this gives

$$f(\nu) = \left(\frac{2}{\pi}\right)^{1/2}\nu\int_{-\infty}^{\infty}\frac{\mathrm{d}\lambda}{\sqrt{2\pi}}e^{i\lambda\nu}e^{-\lambda^2/2+(-i\lambda)^3\varepsilon_1/6+(-i\lambda)^4\varepsilon_2/24+\ldots}\mathcal{P}(\lambda) \tag{7.31}$$

where $\mathcal{P}(\lambda)$ is the truncated series given by

$$\mathcal{P}(\lambda) = 1 + \frac{1}{4}i\varepsilon_1\lambda + \frac{5}{16}\frac{i\varepsilon_1}{\lambda} - \frac{1}{4}\lambda^2\left(\frac{\varepsilon_1^2}{2} - \frac{\varepsilon_2}{3}\right) + \ldots \tag{7.32}$$

The integral in eq. (7.31) can be performed using the saddle point approximation. We write it as

$$f(\nu) = \left(\frac{2}{\pi}\right)^{1/2}\nu\int\frac{\mathrm{d}\lambda}{2\pi}e^{\phi(\lambda)}\,, \tag{7.33}$$

where

$$\phi(\lambda) \equiv i\lambda\nu - \frac{1}{2}\lambda^2 + \frac{i\varepsilon_1}{6}\lambda^3 + \frac{\varepsilon_2}{24}\lambda^4 + \ln\mathcal{P}(\lambda) + \ldots \tag{7.34}$$

The location of the saddle point, $\lambda = \lambda_*$, is the solution of $\phi'(\lambda_*) = 0$, and the saddle point approximation then tells us that

$$\int_{-\infty}^{\infty}\frac{\mathrm{d}\lambda}{\sqrt{2\pi}}e^{\phi(\lambda)} = e^{\phi(\lambda_*)}(|\phi''(\lambda_*)|)^{-1/2} \tag{7.35}$$

[4]We are using a regulator which shifts the pole at $\lambda = 0$ in the last expression in (7.30), to $\lambda = -i\alpha$ where α is real, positive and small.

(see Appendix D 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 $\epsilon\nu$. The expression for ϕ' at the relevant order is given by

$$\phi'(\lambda) = i\nu - \lambda + \frac{i\varepsilon_1}{2}\lambda^2 + \dots \,, \tag{7.36}$$

and solving for λ_* perturbatively up to order $\epsilon\nu$, we find

$$\lambda_* = i\nu\left[1 - \frac{1}{2}\varepsilon_1\nu + \mathcal{O}(\epsilon^2\nu^2)\right] . \tag{7.37}$$

The expression for the mass function $f(\nu)$ then works out to

$$f(\nu) = \left(\frac{2}{\pi}\right)^{1/2} \nu \exp\left[-\frac{1}{2}\nu^2\left(1 - \frac{\varepsilon_1}{3}\nu + \frac{1}{4}\left(\varepsilon_1^2 - \frac{\varepsilon_2}{3}\right)\nu^2 + \mathcal{O}(\epsilon^3\nu^3)\right)\right] \times \left(1 - \frac{1}{4}\varepsilon_1\nu\left(3 - \frac{5}{4\nu^2}\right) + \left(\varepsilon_1^2 - \frac{\varepsilon_2}{3}\right)\nu^2 + \mathcal{O}(\epsilon^3\nu^3)\right) , \tag{7.38}$$

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, (7.24) is replaced with

$$\begin{aligned} f &= (\nu\partial_\nu - 2S\partial_S)e^{-(\varepsilon_1(S)/3!)\partial_\nu^3 + (\varepsilon_2(S)/4!)\partial_\nu^4 + \dots} g(\nu, S) \\ &= \left[\nu + \frac{1}{3}\dot{\varepsilon}_1\varepsilon_1\partial_\nu^2 - \frac{1}{12}\dot{\varepsilon}_2\varepsilon_2\partial_\nu^3\right] e^{-(\varepsilon_1(S)/3!)\partial_\nu^3 + (\varepsilon_2(S)/4!)\partial_\nu^4 + \dots}\, \partial_\nu g(\nu, S) \\ &\qquad - 2te^{-(\varepsilon_1(S)/3!)\partial_\nu^3 + (\varepsilon_2(S)/4!)\partial_\nu^4 + \dots}\, \partial_t g(\nu, S) \,, \end{aligned} \tag{7.39}$$

where, for any function $v(S)$, the dot is defined as

$$\dot{v}(S) \equiv \frac{d\ln v}{d\ln S} \,, \tag{7.40}$$

and the function $g(\nu, S)$ can be shown to be

$$g(\nu, S) = \left(\frac{2}{\pi}\right)^{1/2}\left[\left(\frac{\pi}{2}\right)^{1/2} \operatorname{erf}\frac{\nu}{\sqrt{2}} + \frac{1}{4}\varepsilon_1 c_1 e^{-\nu^2/2} + \frac{\varepsilon_1}{4}\left(\frac{3}{4}c_2 - 2c_3\right)h(\nu) - \frac{1}{8}\varepsilon_1^2 c_1^2 \nu e^{-\nu^2/2} + \frac{1}{12}\varepsilon_2 c_4 \nu e^{-\nu^2/2} + \dots\right] , \tag{7.41}$$

The expression in (7.39) can be evaluated analogously to (7.29), 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

$$f(\nu,S) = \left(\frac{2}{\pi}\right)^{1/2} \nu \exp\left[-\frac{1}{2}\nu^2\left(1-\frac{\varepsilon_1}{3}\nu+\frac{1}{4}\left(\varepsilon_1^2-\frac{\varepsilon_2}{3}\right)\nu^2+\mathcal{O}(\epsilon^3\nu^3)\right)\right] \times\left\{1-\frac{1}{4}\varepsilon_1\nu\left[\left(c_1+2-\frac{4}{3}\dot{\varepsilon}_1\right)+\frac{1}{\nu^2}\left(\frac{3}{4}c_2-2c_3+\frac{4}{3}\dot{\varepsilon}_1+2c_1(\dot{\varepsilon}_1+\dot{c}_1)\right)\right]\right. \left.+\frac{1}{8}\nu^2\left[\varepsilon_1^2\left(c_1^2+2c_1+5-\frac{2}{3}\dot{\varepsilon}_1(c_1+6)\right)-2\varepsilon_2\left(1+\frac{1}{3}c_4-\frac{1}{3}\dot{\varepsilon}_2\right)\right]+\mathcal{O}(\epsilon^3\nu^3)\right\}, \quad (7.42)$$

which reduces to (7.38) 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 $\sigma^2(R)=Ad_1(R)+A(Af_{\rm NL}^2)d_2(R)$ 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 $\epsilon\sim f_{\rm NL}A^{1/2}$. 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 $\varepsilon\nu^3\ll 1$, this correction can still be consistently ignored. Hence we can safely set $\nu=\delta_c/\sigma_{\rm g}$ in all of our expressions.

7.3 Consistency of the truncation

7.3.1 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

$$f\sim e^{-\frac{1}{2}\nu^2\left(1+\epsilon\nu+\epsilon^2\nu^2+\mathcal{O}(\epsilon^2,\epsilon^3\nu^3)\right)}\left[1+\epsilon\nu+\frac{\epsilon}{\nu}+\epsilon^2\nu^2+\mathcal{O}(\epsilon\nu^{-3},\epsilon^2,\epsilon^3\nu^3)\right], \quad (7.43)$$

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 (7.43) 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 up to 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

$$f \sim e^{-\frac{1}{2}\nu^2\left(1+\epsilon\nu+\epsilon^2\nu^2\right)}\left[1+\epsilon\nu+\mathcal{O}\left(\nu^{-4}\right)\right] . \tag{7.44}$$

Clearly, similar arguments can be applied at smaller scales where, for instance, one might have $\epsilon\nu^3 \simeq \nu^{-1}, \nu^{-2}$, etc. It is then important to ask which mass scales correspond to these "transition points". In figure 7.4 we plot $\nu(M, z)$ given by (7.16) 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 $f_{\rm NL} = 100$ we have $\varepsilon_1 \simeq 0.02$ (see fig. 7.1), 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

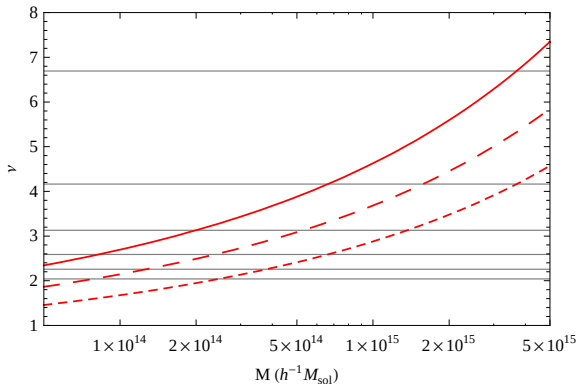


Figure 7.4: $\nu \equiv \delta_c(z)/\sigma(M)$ in the range $5\cdot10^{13} < (M/h^{-1}M_{\rm sol}) < 5\cdot10^{15}$ for three different redshifts, with $\epsilon = 1/300$. The solid, long dashed and short dashed curves correspond to redshifts $z = 1$, 0.5 and 0 respectively. The horizontal lines mark 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}.

reader may skip to the end of the present subsection, where we present such a single consistent expression.

7.3.1.1 $\epsilon\nu^3 \simeq \nu^{-1}$

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

$$\epsilon\nu \simeq \nu^{-3} \quad ; \quad \epsilon\nu^{-1} \simeq \nu^{-5} \quad ; \quad \epsilon^2\nu^2 \simeq \nu^{-6}\,,$$

while discarding

$$\mathcal{O}(\epsilon\nu^{-3}) = \mathcal{O}(\nu^{-7}) \quad ; \quad \mathcal{O}(\epsilon^2) = \mathcal{O}(\nu^{-8}) \quad ; \quad \mathcal{O}(\epsilon^3\nu^3) = \mathcal{O}(\nu^{-9})\,.$$

Our expression (7.43) 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

$$f \sim e^{-\frac{1}{2}\nu^2\left(1+\epsilon\nu+\epsilon^2\nu^2\right)}\left[1+\epsilon\nu+\frac{\epsilon}{\nu}+\mathcal{O}\left(\nu^{-6}\right)\right] \tag{7.45}$$

is also consistent at this transition point, and has a form which is identical to the ones we will see next.

7.3.1.2 $\epsilon\nu^3 \simeq \nu^{-2}$

As we mentioned earlier, this transition point is observationally quite interesting. The terms we retain in the exponential are

$$\epsilon\nu^3 \simeq \nu^{-2} \quad ; \quad \epsilon^2\nu^4 \simeq \nu^{-6}\,,$$

while discarding $\mathcal{O}(\epsilon^3\nu^5) = \mathcal{O}(\nu^{-10})$, and in the polynomial we retain

$$\epsilon\nu \simeq \nu^{-4} \quad ; \quad \epsilon\nu^{-1} \simeq \nu^{-6} \quad ; \quad \epsilon^2\nu^2 \simeq \nu^{-8}\,,$$

while discarding

$$\mathcal{O}(\epsilon\nu^{-3}) = \mathcal{O}(\nu^{-8}) \quad ; \quad \mathcal{O}(\epsilon^2) = \mathcal{O}(\nu^{-10}) \quad ; \quad \mathcal{O}(\epsilon^3\nu^3) = \mathcal{O}(\nu^{-12})\,.$$

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 thus read

$$f \sim e^{-\frac{1}{2}\nu^2\left(1+\epsilon\nu+\epsilon^2\nu^2\right)}\left[1+\epsilon\nu+\frac{\epsilon}{\nu}+\mathcal{O}\left(\nu^{-8}\right)\right]\,. \tag{7.46}$$

7.3.1.3 $\epsilon\nu^3 \simeq \nu^{-3}$

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

$$f \sim e^{-\frac{1}{2}\nu^2\left(1+\epsilon\nu+\epsilon^2\nu^2\right)}\left[1+\epsilon\nu+\frac{\epsilon}{\nu}+\mathcal{O}\left(\nu^{-9}\right)\right]\,. \tag{7.47}$$

7.3.1.4 $\epsilon\nu^3 \simeq \nu^{-4}$ 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

$$f \sim e^{-\frac{1}{2}\nu^2(1+\epsilon\nu)}\left[1+\epsilon\nu+\frac{\epsilon}{\nu}+\ldots\right] \tag{7.48}$$

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 7.2 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* the 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

$$f \sim e^{-\frac{1}{2}\nu^2\left(1+\epsilon\nu+\epsilon^2\nu^2\right)}\left[1+\epsilon\nu+\frac{\epsilon}{\nu}+\mathcal{O}(\epsilon^3\nu^5,\epsilon^2\nu^2,\epsilon\nu^{-3})\right], \tag{7.49}$$

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 7.4):

$$\begin{aligned} f(\nu,S) = f_{\rm PS}(\nu)\,\exp\left(\frac{1}{6}\varepsilon_1\nu^3-\frac{1}{8}\left(\varepsilon_1^2-\frac{\varepsilon_2}{3}\right)\nu^4\right) \\ \times\left\{1-\frac{1}{4}\varepsilon_1\nu\left(\left(c_1+2-\frac{4}{3}\dot{\varepsilon}_1\right)+\frac{1}{\nu^2}\left(\frac{3}{4}c_2-2c_3+\frac{4}{3}\dot{\varepsilon}_1+2c_1(\dot{\varepsilon}_1+\dot{c}_1)\right)\right) \right. \\ \left. +\mathcal{O}(\epsilon^3\nu^5,\epsilon^2\nu^2,\epsilon\nu^{-3})\right\}. \end{aligned} \tag{7.50}$$

7.3.2 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 [73, 105, 106]. If one considers the range of validity of the perturbative expansion, the strongest result so far has been due to MVJ [106], who explicitly retain the exponential dependence on ε_1. Their expression for f can be written as[5]

$$f_{\rm MVJ} = f_{\rm PS}(\nu)\frac{e^{\varepsilon_1\nu^3/6}}{\sqrt{1-\varepsilon_1\nu/3}}\left[1-\frac{1}{2}\varepsilon_1\nu\left(1-\frac{2}{3}\dot{\varepsilon}_1\right)\right]. \tag{7.51}$$

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 Sheth & Tormen one [115]. 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 (7.50). When one

[5]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 (7.51), which is also the one used by most other authors (see e.g. [22, 132]).

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 [122]). Indeed, our calculations in section 7.2 were based on techniques discussed by MR in [71, 73]. 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

$$f_{\rm MR} = f_{\rm PS}(\nu)\left\{1+\frac{1}{6}\varepsilon_1\nu^3\left[1-\frac{3}{2\nu^2}\left(c_1+2-\frac{4}{3}\dot{\varepsilon}_1\right)-\frac{3}{2\nu^4}\left(\frac{3}{4}c_2-2c_3+4\dot{\varepsilon}_1+2c_1(\dot{\varepsilon}_1+\dot{c}_1)\right)\right]\right\}. \tag{7.52}$$

This expression is precisely what one obtains by linearizing our expression (7.50) 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 [133]. The result most often quoted in the literature is their expression linear in ε_1 (and hence in $\varepsilon_1\nu^3$), which is

$$f_{\rm LMSV,lin} = f_{\rm PS}(\nu)\left\{1+\frac{1}{6}\varepsilon_1\nu^3\left[1-\frac{1}{\nu^2}\left(3-2\dot{\varepsilon}_1\right)-\frac{2}{\nu^4}\dot{\varepsilon}_1\right]\right\}. \tag{7.53}$$

In Appendix B.3 of [105], LMSV also give an expression involving ε_1^2 and ε_2, which can be written as

$$\begin{aligned} f_{\rm LMSV,quad} = f_{\rm PS}(\nu)\Bigg[1&+\frac{1}{6}\varepsilon_1\left(H_3(\nu)+\frac{2}{\nu}\dot{\varepsilon}_1H_2(\nu)\right)\\ &+\frac{1}{72}\varepsilon_1^2\left(H_6(\nu)+\frac{4}{\nu}\dot{\varepsilon}_1H_5(\nu)\right)+\frac{1}{24}\varepsilon_2\left(H_4(\nu)+\frac{2}{\nu}\dot{\varepsilon}_2H_3(\nu)\right)\Bigg], \end{aligned} \tag{7.54}$$

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

$$f \sim e^{-\nu^2/2}\left[1+\epsilon\nu^3+\epsilon\nu+\frac{\epsilon}{\nu}+\ldots\right], \tag{7.55}$$

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 (7.55) 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

$$\epsilon\nu^3 \simeq \nu^{-2} \; ; \quad \epsilon\nu \simeq \nu^{-4} \; ; \quad \epsilon\nu^{-1} \simeq \nu^{-6} \, ,$$

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 (7.55). The exponential also contributes a term $\epsilon^2\nu^4 \simeq \nu^{-6}$, which in fact involves the *trispectrum* of NG, again at the order retained by MR and LMSV. The error in the expression (7.55) 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 (7.55), as it stands, becomes exactly consistent (in the sense discussed in the previous subsection, see below (7.49)) 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 (7.55) 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 (7.50), in which the error (at least on large scales) is always parametrically *smaller* than the smallest terms we retain.

7.4 Effects of the diffusing barrier and the filter

In [72], MR showed that the agreement between a Gaussian mass function calculated using the statistics of random walks and the mass functions observed in numerical simulations with Gaussian initial conditions can be improved by allowing the barrier itself to perform a random walk. The width of this scatter was found by Robertson *et al.* [134] 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 [134].

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][6]. As MR argue in [73], 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 (7.16), setting $\sqrt{a} = 0.89$.

In [71], 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 7.2 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. A very good analytical approximation for the symmetric object Δ_{jk} is

$$\Delta_{jk} \simeq \kappa \min(S_j, S_k) \left(1 - \frac{\min(S_j, S_k)}{\max(S_j, S_k)} \right) , \tag{7.56}$$

where $\kappa(R) \simeq 0.464 + 0.002R$, with R measured in h^{-1} Mpc. The mass function is then obtained

[6] 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 [134]. 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.

by perturbatively expanding in Δ_{ij}, with the leading effect being due to the integral

$$\int_{-\infty}^{\delta_c} \mathrm{d}\delta_1 \ldots \mathrm{d}\delta_n \frac{1}{2} \sum_{j,k=1}^{n} \Delta_{jk} \partial_j \partial_k W^{\mathrm{gm}} \,, \tag{7.57}$$

which on evaluation leads to

$$f_{\mathrm{g,sharp-x}}(\nu, S) = \left(\frac{2}{\pi}\right)^{1/2} \nu \left[(1-\kappa) e^{-\nu^2/2} + \frac{\kappa}{2} \Gamma\left(0, \frac{\nu^2}{2}\right) + \mathcal{O}(\kappa^2)\right] \,, \tag{7.58}$$

where the subscript stands for Gaussian noise with the top-hat filter in real space, and κ introduces a weak explicit S dependence. In [73], MR proposed an extension of this result to the non-Gaussian case, by assuming that all the non-Gaussian terms 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 [73]) is

$$f_{\mathrm{ng,sharp-x}}(\nu, S) \sim \nu \left[(1-\kappa) e^{-\nu^2/2} \left(1 + \epsilon\nu^3 + \epsilon\nu + \epsilon\nu^{-1}\right) + \frac{\kappa}{2} \Gamma\left(0, \frac{\nu^2}{2}\right)\right] \,, \tag{7.59}$$

with the specific coefficients of the $\epsilon\nu^3$, $\epsilon\nu$ and $\epsilon\nu^{-1}$ terms being identical to those in (7.52). 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 $\sim \varepsilon_1 S^{1/2} \sum_j (S - S_j)\partial_j \sum_{k,l} \partial_k\partial_l$, 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[7]. Our calculations, which are 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. At present, 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 in [72], 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.* [134].

[7] Notice that this issue is completely decoupled from the subtleties in truncation discussed in section 7.3 – this problem remains even at scales where the MR expression is formally consistent.

Our final expression for the mass function, corrected for effects of the diffusing barrier and the top-hat real space filter, is

$$f(\nu,S) = f_{\rm PS}(\nu)\left(1-\tilde{\kappa}+\mathcal{O}(\tilde{\kappa}^2)\right)\exp\left[\frac{1}{6}\varepsilon_1\nu^3-\frac{1}{8}\left(\varepsilon_1^2-\frac{\varepsilon_2}{3}\right)\nu^4\right] \times\left\{1+\frac{\left(1-2\dot{\tilde{\kappa}}\right)}{1-\tilde{\kappa}}\tilde{\kappa}\nu^{-2}\left(1-2\nu^{-2}\right)-\frac{1}{4}\varepsilon_1\nu\left(\frac{c_1}{1-\tilde{\kappa}}+2-\frac{4}{3}\dot{\varepsilon}_1\right) -\frac{1}{4}\varepsilon_1\nu^{-1}\left(\frac{3}{4}c_2-2c_3+\frac{4}{3}\dot{\varepsilon}_1+2c_1(\dot{\varepsilon}_1+\dot{c}_1)\right) +\mathcal{O}(\tilde{\kappa}^2\nu^{-2},\tilde{\kappa}\epsilon\nu,\tilde{\kappa}\nu^{-6})+\mathcal{O}(\epsilon^2\nu^2,\epsilon^3\nu^5,\epsilon\nu^{-3})\right\}, \quad (7.60)$$

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 $f_{\rm PS}(\nu)=\sqrt{2/\pi}\,\nu\,e^{-\nu^2/2}$ is the Bond *et al.* result, with $\nu(M,z)$ defined in (7.16).

To summarize, (7.60) 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 7.1 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, 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 $f_{\rm NL}$ 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:

$$\dot{\varepsilon}_1=\frac{3}{2}\left(c_1-1\right)\ ;\ \dot{c}_1=1-\frac{3}{2}c_1+\frac{1}{c_1}\left(c_3-\frac{1}{2}c_2\right). \quad (7.61)$$

For completeness, in table 7.1 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.

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
$\varepsilon_2(f_{\rm NL}^2)$	-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 7.1: Analytical fits for the various NG parameters, in the local and equilateral cases, as a function of $t = \sigma^2$, in the range $2 \cdot 10^{13} < M/(h^{-1}M_{\rm sol}) < 5 \cdot 10^{15}$, for $f_{\rm NL} = 100$ and $g_{\rm NL} = 10^4$. We have $\varepsilon_1 \propto f_{\rm NL}$ in both cases, and for ε_2 in the local case we give separate fits for the terms proportional to $f_{\rm NL}^2$ 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 S, in the same mass range. The errors on all the fits are less than 1%, except for $\varepsilon_2(f_{\rm NL}^2)$ where the error is $\sim 6\%$, because of numerical difficulties in calculating this object. These fits of course depend on our choice of cosmological parameters.

7.5 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, recall that for MVJ and LMSV one has to multiply an analytically predicted ratio $R_{\rm ng} = f(\nu, M, f_{\rm NL})/f(\nu, M, f_{\rm NL} = 0)$ with a suitable Gaussian mass function based on fits to simulations, and 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 figures 7.5 and 7.6 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 (7.60) with the expressions of MR (7.52), LMSV linear (7.53) and quadratic (7.54), and of MVJ (7.51). 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 $f_{\rm NL} = 0$ (i.e. without NG but with filter effects when present). We use the local model of non-Gaussianity, setting $f_{\rm NL} = 100$ and $g_{\rm NL} = 0$, and the reference ΛCDM cosmology described in section 7.1. 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 (7.60) and (7.52) due to the barrier diffusion. We set $\sqrt{a} = 0.89$, which is the value inferred in [72], using the simulations of Robertson *et al.* [134].

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 section7.3, 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 (7.53) 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 quadratic LMSV expression (7.54) 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 (7.51) is the maximum of $\epsilon\nu$ (unequal time terms) and $\tilde{\kappa}\nu^{-2}$ (filter effects). Finally, the error for the MR expression (7.52) 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. 7.6, as explained above.

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 the correct way to proceed is 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 figure 7.6.

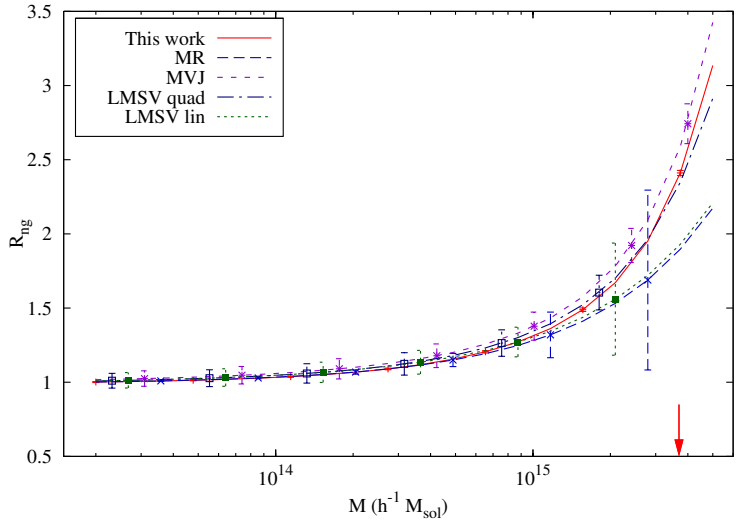


Figure 7.5: Theoretical comparison of the different mass functions at $z = 1$, without the filter effects, i.e. setting $\tilde{\kappa} = 0$. We plot the ratio $R_{\rm ng}$ of the non-Gaussian and Gaussian mass functions, in the local model with $f_{\rm NL} = 100$ and $g_{\rm NL} = 0$. See main text 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.

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 [71].

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 top-hat filter, which generates non-Markovianities in the random walk. We recognize two small parameters in which we perturb: ϵ, defined in (7.15), which measures the magnitude of the primordial NG, and $\nu^{-1} = \sigma_R/\delta_c$, which is small on 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. 7.3. We then used saddle point techniques 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 the filter function and to deviations from spherical collapse, as explained in sec. 7.4. Our final result is

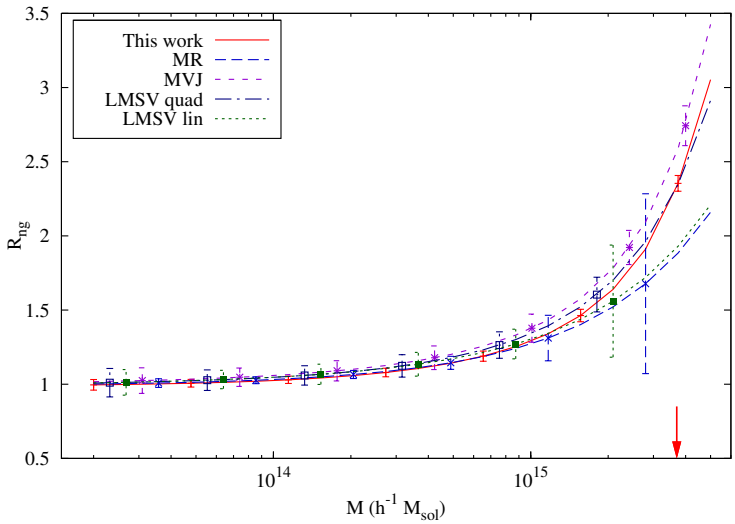


Figure 7.6: Same as fig. 7.5, but including filter effects. These affect only the error bars for MVJ and LMSV, and they affect both the curve and the error bars for MR and our result. For MR and our result, the Gaussian mass function used to construct the ratio $R_{\rm ng}$, is taken as the non-Gaussian result at $f_{\rm NL} = 0$, and hence includes filter effects.

presented in (7.60), which we reproduce here:

$$f(\nu,S) = f_{\rm PS}(\nu)\left(1-\tilde{\kappa}+\mathcal{O}(\tilde{\kappa}^2)\right)\exp\left[\frac{1}{6}\varepsilon_1\nu^3-\frac{1}{8}\left(\varepsilon_1^2-\frac{\varepsilon_2}{3}\right)\nu^4\right]$$
$$\times\left\{1+\frac{\left(1-2\dot{\tilde{\kappa}}\right)}{1-\tilde{\kappa}}\tilde{\kappa}\nu^{-2}\left(1-2\nu^{-2}\right)-\frac{1}{4}\varepsilon_1\nu\left(\frac{c_1}{1-\tilde{\kappa}}+2-\frac{4}{3}\dot{\varepsilon}_1\right)\right.$$
$$-\frac{1}{4}\varepsilon_1\nu^{-1}\left(\frac{3}{4}c_2-2c_3+\frac{4}{3}\dot{\varepsilon}_1+2c_1(\dot{\varepsilon}_1+\dot{c}_1)\right)$$
$$\left.+\mathcal{O}(\tilde{\kappa}^2\nu^{-2},\tilde{\kappa}\epsilon\nu,\tilde{\kappa}\nu^{-6})+\mathcal{O}(\epsilon^2\nu^2,\epsilon^3\nu^5,\epsilon\nu^{-3})\right\}. \quad (7.62)$$

In table 7.1 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 figures 7.5 and 7.6. In our work we have improved over the calculations of MVJ [106] and LMSV [105] (who ignore unequal time correlations) and of MR [73] (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 (7.62) 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 [73], and the saddle point approximation as used by MVJ [106].

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. Third, it would be interesting to study how to account for the effects of ellipsoidal collapse, in a framework such as the one employed in this paper. 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 [135]. Since voids probe larger length/mass scales than halos, they constitute a promising future probe of primordial NG [136].

Chapter 8

Conclusions

In this Thesis, we studied the modifications induced by generic single-field dark energy and by primordial non-Gaussianity on the large scale structure, in particular on the abundance and evolution of high-mass galaxy clusters.

First we studied generic single-field dark energy models, by a parametrization of the most general theory of their perturbations around a given background, including higher derivative terms. In appropriate limits this approach reproduces standard quintessence, k-essence and ghost condensation. We find no general pathology associated to an equation of state $w_Q < -1$ or in crossing the phantom divide $w_Q = -1$, but stability requires that the $w_Q < -1$ side of dark energy behaves, on cosmological scales, as a k-essence fluid with a virtually zero speed of sound. This implies that one should set the speed of sound to zero when comparing with data models with $w_Q < -1$ or crossing the phantom divide.

In particular, we can expect modifications on the evolution and abundance of massive galaxy clusters, and we studied the spherical collapse model in the presence of quintessence with negligible speed of sound (and generic values of w). As pressure gradients are negligible, quintessence follows dark matter during the collapse. The spherical overdensity behaves as a separate closed FLRW universe, so that its evolution can be studied exactly. We derive the critical overdensity for collapse and we use the extended Press-Schechter theory to study how the clustering of quintessence affects the dark matter mass function. The effect is dominated by the modification of the linear dark matter growth function, but a larger effect occurs on the total mass function, which includes the quintessence overdensities. Indeed, here quintessence constitutes a third component of virialized objects, together with baryons and dark matter, and contributes to the total halo mass by a fraction $\sim (1+w)\Omega_Q/\Omega_m$. This gives a distinctive modification of the total mass function at low redshift, in

coincidence with the onset of acceleration. We also expect to observe a modification of the baryon to dark matter ratio in clusters with respect to the cosmological value.

Finally, we use the halo mass function as a 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 improve 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 account for the various scale dependent combinations of small parameters which appear, since 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 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, and we also discuss 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. We present a detailed comparison between the various expressions for the mass functions, exploring the accuracy and range of validity of each.

The lines of work developed here can be continued in several directions. First of all, in this thesis we do not attempt to study experimental constraints and forecasts. As far as dark energy is concerned, it is important to assess the robustness of our estimation of the mass function by cosmological simulations. For instance, these should clarify that our estimate of the effect of the accretion of dark energy is correct, and quantify the effect of clustering quintessence in the $f_{\rm gas}$ parameter. Before doing simulations, however, it is important to understand how cusps in the scalar field can be resolved. Then, from the observational side, one can use our results to analyze cluster data. Whether the effect be measurable or not will depend on the ability to break the degeneracies with other cosmological parameters and on the precise determination of the cluster mass.

Switching to non-Gaussianities, there are a few things that can be studied. From the purely theoretical side, our calculation can be improved by thoroughly considering the effects due to the choice of the filter function, which is the main source of error in our expression. A very interesting problem will be to apply the formalism to the void mass function, which has not received

much attention yet but it may become an important tool to constrain non-Gaussianities. Then, a comparison with N-body simulations should be performed, in order to quantitatively assess the possibility of constraining NG using our work.

The future of cosmology will heavily rely on large scale structure observations. The interplay between observational techniques and theoretical advances is of paramount importance to push the knowledge of our Universe forward. We hope to have contributed with some small steps.

Appendix A

Higher derivative operators in effective field theories

In this appendix we want to study the Ghost Condensate and its deformations from an effective field theory point of view. In particular we want to show that, although the operator $(\nabla^2\pi)^2$ dominates the dynamics, operators containing higher time derivatives such as $\ddot{\pi}^2$ must be treated perturbatively.

In the Ghost Condensate limit, the free π action is

$$S = \frac{M^4}{2}\int d^3x dt \left[\dot{\pi}^2 - \frac{(\nabla^2\pi)^2}{M^2}\right] , \tag{A.1}$$

where we neglected the mixing with gravity – as we are interested in the high energy behavior of the theory – and for simplicity we assumed that there is a single scale M ($M \simeq \bar{M}$). This action is manifestly invariant under the energy scaling [28]

$$E \to sE , \qquad t \to s^{-1}t , \qquad x \to s^{-1/2}x , \qquad \pi \to s^{1/4}\pi . \tag{A.2}$$

As the theory is not Lorentz invariant, time and space behave differently under rescaling, and π does not scale as s^1 as in a Lorentz invariant theory (see for instance [137] for an introduction to scaling in non-Lorentz invariant field theories).

What is the physical meaning of this scaling transformation? Assuming that the theory is weakly coupled, the free action gives the leading contribution to the correlation functions, so that these will be invariant under the scaling above. For instance, a relativistic massless scalar has

scaling dimension 1. Thus, the two-point function satisfies[1]

$$\langle\phi\phi\rangle(x-y) = s^{-2}\langle\phi\phi\rangle\left(\frac{x-y}{s}\right) \qquad \Rightarrow \qquad \langle\phi\phi\rangle \propto \frac{1}{|x-y|^2}\,. \tag{A.3}$$

In the case of the action (A.1) above, the scaling transformation (A.2) yields

$$\langle\pi\pi\rangle(\Delta t, \Delta\vec{x}) = s^{-1/2}\langle\pi\pi\rangle\left(\frac{\Delta t}{s}, \frac{\Delta\vec{x}}{s^{1/2}}\right)\,. \tag{A.4}$$

Not only the scaling transformation gives information on the free theory, but, more importantly, it allows one to estimate the effect of different operators added to the free action. In particular, in the Ghost Condensate case, one can check that all additional operators allowed by symmetries have positive scaling dimensions, so that their importance is suppressed by E/M elevated to a positive power [28]. This implies that at low energy the theory is perturbative. For instance, the leading irrelevant operator is $\dot\pi(\nabla\pi)^2$, which has scaling dimension 1/4.

Operators containing higher time derivatives have positive scaling dimensions so that they must be treated perturbatively. For instance $\ddot\pi^2$ has scaling dimension 2, so that at low energy it is negligible. Additional time derivatives naïvely suggest the existence of more and more solutions of the equations of motion. However, these solutions are non-perturbative in the expansion parameter E/M, and there is no reason to expect that they have any physical meaning. For example, taking seriously these solutions would imply that the Minkowski vacuum is unstable when considering higher order corrections to the Einstein-Hilbert action [138]. The correct way of treating these terms is perturbatively, i.e., evaluating them using the lower order equations of motion [139]. Following this logic, the additional solutions studied in the context of the Ghost Condensate theory in [140] are non-physical, as already pointed out in [141].

As we discussed in this paper, in certain regimes quintessence behaves as a deformation of the Ghost Condensate theory. The free action (A.1) is deformed by the addition of a $(\nabla\pi)^2$ term:

$$S = \frac{M^4}{2}\int d^3x dt\left[\dot\pi^2 - c_s^2(\nabla\pi)^2 - \frac{(\nabla^2\pi)^2}{M^2}\right]\,. \tag{A.5}$$

In these cases the dispersion relation is not exactly $\omega \sim k^2/M$ but it contains also a linear term $\omega \sim c_s k$, with $c_s \ll 1$ (for this discussion we assume that c_s^2 is positive).

The situation is now trickier than before because one cannot find a scaling transformation which leaves the full action (A.5) invariant. On the other hand, one can separate two regimes, depending on which of the gradient terms dominates, as illustrated in figure A.1.

[1] Note that the scaling dimension has nothing to do with the mass dimension of the field ϕ. Indeed, eq. (A.3) remains the same if we choose a non-conventional normalization of the action such that ϕ has not mass dimension 1.

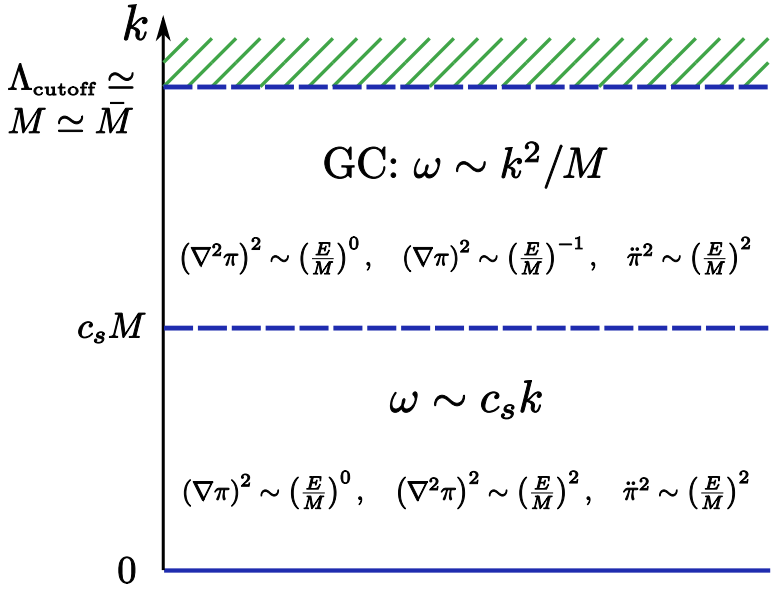


Figure A.1: *The two scaling regimes as a function of the momentum k together with the scaling dimensions of some of the operators. Above: the Ghost Condensate regime where $\omega \sim k^2/M$; below: the regime where $\omega \sim c_s k$.*

For $k \gg c_s M$, the dispersion relation is dominated by $(\nabla^2\pi)^2$ and the theory behaves as the Ghost Condensate. In this regime the scaling of all additional operators can be obtained from eq. (A.2). Notice that now there is a relevant operator, $(\nabla\pi)^2$, that becomes more and more important at low momenta (and energies). The coefficient of this operator is however suppressed by the small deformation parameter $c_s \ll 1$. Thus, it can be treated perturbatively as long as $k \gg c_s M$.

On the other hand, when $k \ll c_s M$ the $(\nabla\pi)^2$ operator dominates the free action. In this regime the scaling becomes the same as in the relativistic case. Now both $\ddot{\pi}^2$ and $(\nabla^2\pi)^2$ are irrelevant operators with the same scaling dimension 2. However, time and spatial derivatives are still on a different footing because the time derivatives are suppressed by c_s with respect to the spatial ones, $\omega \sim c_s k$. Thus $\ddot{\pi}^2 \sim c_s^4 (\nabla^2\pi)^2$ for $k \ll c_s M$. In the intermediate regime $k \sim c_s M$ this suppression can also be obtained from the Ghost Condensate limit. Indeed, at high momenta the operator $\ddot{\pi}^2$ scales like $(k/M)^4$, so that it is suppressed by c_s^4 for $k \sim c_s M$.

Even though these theories make perfect sense as effective field theories, it is not clear whether one can find a UV completion. In particular, the violation of the null energy condition may be problematic in the context of black hole thermodynamics [142].

Appendix B

Modification of gravity with $\hat{M}$

In the main text we studied the effects of the operator $\hat{M}$, focusing on stability and on the phenomenology at cosmological scales. In analogy to what happens for the Ghost Condensate, we expect that this operator will also be relevant at short scales, inducing a modification of gravity. In this appendix we perform a preliminary analysis, restricted to linear perturbations only, although the non-linear dynamics has been shown to be relevant and quite rich in the Ghost Condensate case (see for example [50]). To simplify the analysis we set $\rho_Q + p_Q = 0$ and $\bar{M} = 0$. Although the background quintessence stress-energy tensor is the one of the cosmological constant, there is still a propagating scalar degree of freedom. Its mixing with gravity induces a deviation from General Relativity; indeed the Ghost Condensate was originally proposed as a consistent modification of gravity in the infrared. The simplest setting to study this modification of gravity is in the Newtonian regime $\omega^2 \ll k^2$ around Minkowski spacetime, where the new scalar degree of freedom modifies the Newtonian potential Φ. For this purpose we will closely follow the discussion done in [28] for the Ghost Condensate case, i.e., for the operator $\bar{M}$.

Working in Newtonian gauge with $\Psi = \Phi$, the metric is $\mathrm{d}s^2 = -(1+2\Phi)\mathrm{d}t^2 + (1-2\Phi)\mathrm{d}\vec{x}^2$ and the quadratic Lagrangian for π and Φ reads

$$\mathcal{L} = -M_{\rm Pl}^2(\nabla\Phi)^2 + 2M^4(\Phi - \dot{\pi})^2 + \hat{M}^3(\Phi - \dot{\pi})(4\dot{\Phi} - \ddot{\pi} + \nabla^2\pi) \ . \tag{B.1}$$

Let us first assume that $\hat{M}$ is time independent. Dropping total derivatives and terms which are negligible in the limit $\omega^2 \ll k^2$, we are left with

$$\mathcal{L} = -M_{\rm Pl}^2(\nabla\Phi)^2 + 2M^4(\Phi - \dot{\pi})^2 + \hat{M}^3\pi\nabla^2\Phi \ . \tag{B.2}$$

In terms of the canonically normalized fields $\pi_c \equiv 2M^2\pi$ and $\Phi_c \equiv \sqrt{2}M_{\rm Pl}\Phi$, the Lagrangian in Fourier space can be written as

$$\mathcal{L} = \frac{1}{2}\left(\pi_c\ \Phi_c\right)\mathcal{M}\begin{pmatrix}\pi_c\\ \Phi_c\end{pmatrix}\ , \tag{B.3}$$

with

$$\mathcal{M} \equiv \begin{pmatrix}\omega^2 & -i\omega\sqrt{2}M^2/M_{\rm Pl} - k^2\hat{M}^3/(2\sqrt{2}M^2M_{\rm Pl})\\ i\omega\sqrt{2}M^2/M_{\rm Pl} - k^2\hat{M}^3/(2\sqrt{2}M^2M_{\rm Pl}) & -k^2 + 2M^4/M_{\rm Pl}^2\end{pmatrix}\ . \tag{B.4}$$

Setting to zero the determinant of this matrix gives the dispersion relation

$$\omega^2 = -\frac{\hat{M}^6}{8M^4M_{\rm Pl}^2}k^2\ , \tag{B.5}$$

which reproduces the Jeans instability already shown in eq. (3.76). The Jeans instability arises from the non-diagonal (mixing) term and it is thus proportional to $\hat{M}^6$ instead of $\hat{M}^3$.

To study the corrections to the Newtonian theory, one can look at the propagator of Φ that is the $\langle\Phi,\Phi\rangle$ entry of $\mathcal{M}^{-1}$. This can be written as

$$-\frac{1}{k^2}\cdot\left[1 - \frac{k^2\hat{M}^6}{8M^4M_{\rm Pl}^2}\cdot\frac{1}{\omega^2 + k^2\hat{M}^6/(8M^4M_{\rm Pl}^2)}\right]\ , \tag{B.6}$$

where the term $-1/k^2$ is simply the standard Newtonian propagator. As expected, a substantial deviation requires, for a given distance k^{-1}, a sufficient time ω^{-1} for the Jeans instability to develop, i.e.,

$$\omega^2 \lesssim \frac{\hat{M}^6}{8M^4M_{\rm Pl}^2}k^2\ . \tag{B.7}$$

We can now consider the case of a time dependent $\hat{M}$. In this way we introduce new terms in the action that were previously dropped because total derivatives. The same happens if we had considered a time dependent spatial metric, but here we stick to Minkowski for simplicity. The matrix $\mathcal{M}$ becomes

$$\mathcal{M} \equiv \begin{pmatrix}\omega^2 - k^2H\hat{M}^3/4M^4 & -i\omega\sqrt{2}M^2/M_{\rm Pl} - k^2\hat{M}^3/(2\sqrt{2}M^2M_{\rm Pl})\\ i\omega\sqrt{2}M^2/M_{\rm Pl} - k^2\hat{M}^3/(2\sqrt{2}M^2M_{\rm Pl}) & -k^2 + 2M^4/M_{\rm Pl}^2 - 2H\hat{M}^3/M_{\rm Pl}^2\end{pmatrix}\ , \tag{B.8}$$

where H is the typical rate of variation of $\hat{M}^3$, $\dot{\hat{M}}^3 = H\hat{M}^3$ (if the time dependence is induced by the metric this becomes the Hubble rate). Computing the determinant and restricting to frequencies much larger than H we get the dispersion relation

$$\omega^2 = \frac{H\hat{M}^3}{4M^4}k^2 - \frac{\hat{M}^6}{8M^4M_{\rm Pl}^2}k^2\ , \tag{B.9}$$

which correctly matches eq. (3.76). The value of $\hat{M}$ can be chosen to avoid the Jeans instability and have a healthy dispersion relation. The propagator becomes

$$-\frac{1}{k^2}\cdot\left[1-\frac{k^2\hat{M}^6}{8M^4M_{\rm Pl}^2}\cdot\frac{1}{\omega^2-k^2\hat{M}^3H/(4M^4)+k^2\hat{M}^6/(8M^4M_{\rm Pl}^2)}\right]\,. \tag{B.10}$$

The scalar degree of freedom induces a $1/r$ force which adds to the Newton law: this force, however, propagates at a very small speed

$$c_s^2\approx\frac{\hat{M}^3H}{4M^4}. \tag{B.11}$$

Given the absence of a Jeans instability, the modification of gravity induced by $\hat{M}$ is very different at linear and non-linear level with respect to the Ghost Condensate case [28, 50]. More work is needed to understand the constraints on $\hat{M}$ coming from the modifications of gravity that it produces.

Appendix C

Hierarchy of terms in the non-Gaussian mass function

Here we argue why the hierarchy of terms ordered by powers of ν^{-2} emerges on expanding the exponentiated derivative operators in (7.24). 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

$$\sim (\varepsilon_1 S^{3/2})^m \sum_{j_1,..,j_{3m}} \int_{-\infty}^{\delta_c} \mathrm{d}\delta_1\ldots\mathrm{d}\delta_n\,[(1-S_{j_1}/S)\ldots(1-S_{j_m}/S)]^p$$
$$\big[\big(1-S_{j_{m+1}}/S\big)\ldots(1-S_{j_{2m}}/S)\big]^q \times \big[\big(1-S_{j_{2m+1}}/S\big)\ldots(1-S_{j_{3m}}/S)\big]^r\,\partial_{j_1}\ldots\partial_{j_{3m}}\,W^{\rm gm}\,, \quad \text{(C.1)}$$

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 (7.25c) has the same structure as the $(1,2,0,0)$ term in (7.25b) – 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

$$\sum_j (1 - S_j/S) \int \mathrm{d}\delta_1 \dots \mathrm{d}\delta_n \partial_j W^{\mathrm{gm}} ,$$

which in the continuum limit can be shown to reduce to the integral

$$g_{(0)}\left(\frac{\nu^2}{2}\right) \equiv \int_0^1 \frac{\mathrm{d}y}{y^{3/2}} (1-y)^{1/2} e^{-\nu^2/2y} = \frac{\sqrt{\pi}}{2} \Gamma\left(-\frac{1}{2}, \frac{\nu^2}{2}\right) . \tag{C.2}$$

Notice the similarity with the function $h(\nu)$ in (7.28), 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

$$g_{(p)}(A) \equiv \int_0^1 \frac{dy}{y^{3/2}} (1-y)^{1/2+p} e^{-A/y} . \tag{C.3}$$

Starting with $p = 1$ and manipulating the integrals, it is straightforward to establish the recurrence

$$g_{(p+1)}(A) = g_{(p)}(A) - \int_A^\infty \mathrm{d}\tilde{A}\, g_{(p)}(\tilde{A}) . \tag{C.4}$$

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 (7.24).

Appendix D

The saddle point approximation

In this appendix we discuss the saddle point approximation of the integrals of the type appearing in section 7.2, 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 [143]. 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 (7.31) 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 (7.42) as described in section 7.2. We begin with expression (7.31):

$$f(\nu) = \left(\frac{2}{\pi}\right)^{1/2} \nu \int_{-\infty}^{\infty} \frac{\mathrm{d}\lambda}{\sqrt{2\pi}} e^{g(\lambda)} \,, \tag{D.1}$$

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

$$\lambda_* = i\nu\big(1 - \varepsilon_1\nu/2 + \mathcal{O}(\epsilon^2\nu^2)\big) \,, \tag{D.2}$$

$$g(\lambda_*) = -\frac{\nu^2}{2}(1 - \frac{1}{3}\varepsilon_1\nu + \mathcal{O}(\epsilon^2\nu^2)) \,, \tag{D.3}$$

$$g''(\lambda_*) = -1 - \varepsilon_1\nu + \mathcal{O}(\epsilon^2\nu^2) \,. \tag{D.4}$$

The saddle point approximation consists roughly of performing a Taylor expansion of $g(\lambda)$ to second order around λ_* in the integrand of (D.1) 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 (D.1) 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 (D.1), which we discuss below,

$$\begin{aligned}\int_{-\infty}^{\infty} \frac{\mathrm{d}\lambda}{\sqrt{2\pi}} e^{g(\lambda)} &\approx \int_{-\infty}^{\infty} \frac{\mathrm{d}\lambda}{\sqrt{2\pi}}\, e^{g(\lambda_*)+g''(\lambda_*)(\lambda-\lambda_*)^2/2} \\ &= e^{g(\lambda_*)} \left(-g''(\lambda_*)\right)^{-1/2} \\ &= e^{-\frac{1}{2}\nu^2\left(1-\varepsilon_1\nu/3+\mathcal{O}(\epsilon^2\nu^2)\right)} \left(1+\varepsilon_1\nu+\mathcal{O}(\epsilon^2\nu^2)\right)^{-1/2} . \end{aligned} \tag{D.5}$$

Here the integrations are performed along the deformed contour.

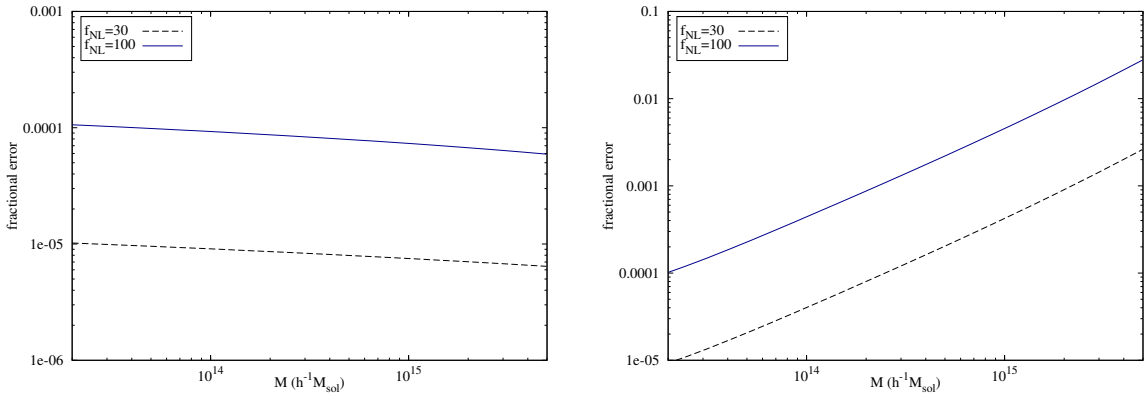
In order to estimate the errors induced by the approximation done in equation (D.5), one can keep higher orders in the Taylor expansion of the function in the exponential:

$$\begin{aligned}\int_{-\infty}^{\infty} \frac{\mathrm{d}\lambda}{\sqrt{2\pi}} &\approx \int_{-\infty}^{\infty} \frac{d\lambda}{\sqrt{2\pi}}\, e^{g(\lambda_*)+g''(\lambda_*)(\lambda-\lambda_*)^2/2+g^{(3)}(\lambda_*)(\lambda-\lambda_*)^3/6+g^{(4)}(\lambda_*)(\lambda-\lambda_*)^4/24+\ldots} \\ &\approx e^{g(\lambda_*)} \left(-g''(\lambda_*)\right)^{-1/2} \\ &\quad + \int_{-\infty}^{\infty} \frac{\mathrm{d}z}{\sqrt{2\pi}} \left\{ \frac{1}{6} g^{(3)}(\lambda_*) z^3 + \frac{1}{72}\left[g^{(3)}(\lambda_*)\right]^2 z^6 + \frac{1}{24} g^{(4)}(\lambda_*) z^4 + \ldots \right\} e^{g(\lambda_*)+g''(\lambda_*)z^2/2} \\ &= e^{g(\lambda_*)} \left(-g''(\lambda_*)\right)^{-1/2} \left(1+\mathcal{O}(\epsilon^2)\right) . \end{aligned} \tag{D.6}$$

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 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

[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 (D.6).



(a) (b)

Figure D.1: Panel (a): Fractional difference between the saddle point approximation on the r.h.s. of (D.7) 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 (D.7) 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 (D.7) and the approximation (D.5). Both panels show the results for a local NG with two values of $f_{\rm NL}$.

and all higher order cumulants ε_n for $n \geq 2$ to be zero[2]. For such a model the integral is

$$\int_{-\infty}^{\infty} \mathrm{d}\lambda \; e^{i\nu\lambda - \lambda^2/2 + (-i\lambda)^3\varepsilon_1/6} \approx \left(\frac{2\pi}{\sqrt{1+2\varepsilon_1\nu}}\right)^{1/2} \exp\left(\frac{1-\sqrt{1+2\varepsilon_1\nu}+\varepsilon_1\nu\left(3-2\sqrt{1+2\varepsilon_1\nu}\right)}{3\varepsilon_1^2}\right). \tag{D.7}$$

In 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. D.1), one can see that the errors introduced by the saddle point approximation are indeed of order ϵ^2 as indicated by (D.6). On the other hand, one can use the numerical integration of the left hand side of this equation and compare it with the approximation (D.5) (panel (b) of fig. D.1), 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).

[2]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 $\epsilon\nu$.

Acknowledgements

Resting for a moment and looking back over my shoulders, my thoughts and deep feelings of gratitude go to many people I met along the path of my life: people who came along and shared my walk, people who pushed me through the hard times, people who helped overcome every kind of obstacle, people who said goodbye.

I want to thank my mom and Marta for their love and support, together with my grandparents, my aunt, my uncle and my cousin.

I am greatly indebted to my advisor, Paolo, for the guidance he provided in my formation as a scientist during these years, and for treating me as a collaborator and friend, rather than a student. Of course, special thanks go to all the members of our cosmology group, Filippo, Lotfi, Marcello and Aseem. Thank you, Jorge, for our work together and for having put up with me as your officemate.

How to forget my "Triestinian" friends? Before I go far away, I would like to hug everyone: il Buffone, Lucone, Bond, sig. Fuffi, the Tiger, Robert, Giulia, Pratika, Carmelo, Barbara, Alessandro, Daniel, Luca, Giorgio, Jarah, Tom, Fabio, il Dottore, Giovanni and many others. I know I will miss you all.

Thank you, Elisa, for all the moments we spent together, and for our future adventure. My life would simply not be the same without you.

Finally, thank you and goodbye, dad. Wherever you are.

Bibliography

[1] E. Komatsu *et. al.*, *Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation*, arXiv:1001.4538. 1, 2, 6, 90, 95, 98

[2] **Boomerang** Collaboration, A. E. Lange *et. al.*, *Cosmological parameters from the first results of BOOMERANG*, *Phys. Rev.* **D63** (2001) 042001, [astro-ph/0005004]. 1

[3] **ACBAR** Collaboration, C.-l. Kuo *et. al.*, *High Resolution Observations of the CMB Power Spectrum with ACBAR*, *Astrophys. J.* **600** (2004) 32–51, [astro-ph/0212289]. 1

[4] C.-P. Ma and E. Bertschinger, *Cosmological perturbation theory in the synchronous and conformal Newtonian gauges*, *Astrophys. J.* **455** (1995) 7–25, [astro-ph/9506072]. 1, 20, 23, 28, 29, 32

[5] W. Hu and N. Sugiyama, *Anisotropies in the Cosmic Microwave Background: An Analytic Approach*, *Astrophys. J.* **444** (1995) 489–506, [astro-ph/9407093]. 1

[6] W. Hu and N. Sugiyama, *Toward understanding CMB anisotropies and their implications*, *Phys. Rev.* **D51** (1995) 2599–2630, [astro-ph/9411008]. 1

[7] M. Zaldarriaga, D. N. Spergel, and U. Seljak, *Microwave Background Constraints on Cosmological Parameters*, *Astrophys. J.* **488** (1997) 1–13, [astro-ph/9702157]. 1

[8] W. Hu, U. Seljak, M. J. White, and M. Zaldarriaga, *A Complete Treatment of CMB Anisotropies in a FRW Universe*, *Phys. Rev.* **D57** (1998) 3290–3301, [astro-ph/9709066]. 1

[9] G. Jungman, M. Kamionkowski, A. Kosowsky, and D. N. Spergel, *Cosmological parameter determination with microwave background maps*, *Phys. Rev.* **D54** (1996) 1332–1344, [astro-ph/9512139]. 1

[10] U. Seljak and M. Zaldarriaga, *A Line of Sight Approach to Cosmic Microwave Background Anisotropies*, *Astrophys. J.* **469** (1996) 437–444, [astro-ph/9603033]. 1, 98

[11] A. Lewis, A. Challinor, and A. Lasenby, *Efficient Computation of CMB anisotropies in closed FRW models*, *Astrophys. J.* **538** (2000) 473–476, [astro-ph/9911177]. 1, 70, 98

[12] **Planck** Collaboration, T. P. Collaboration, *Planck: The scientific programme*, astro-ph/0604069. 1

[13] **CMBPol Study Team** Collaboration, D. Baumann *et. al.*, *CMBPol Mission Concept Study: Probing Inflation with CMB Polarization*, *AIP Conf. Proc.* **1141** (2009) 10–120, [arXiv:0811.3919]. 1

[14] N. Dalal, O. Dore, D. Huterer, and A. Shirokov, *The imprints of primordial non-gaussianities on large- scale structure: scale dependent bias and abundance of virialized objects*, *Phys. Rev.* **D77** (2008) 123514, [arXiv:0710.4560]. 2, 95

[15] S. Matarrese and L. Verde, *The effect of primordial non-Gaussianity on halo bias*, *Astrophys. J.* **677** (2008) L77, [arXiv:0801.4826]. 2, 95

[16] A. Slosar, C. Hirata, U. Seljak, S. Ho, and N. Padmanabhan, *Constraints on local primordial non-Gaussianity from large scale structure*, *JCAP* **0808** (2008) 031, [arXiv:0805.3580]. 2, 95

[17] **Supernova Cosmology Project** Collaboration, S. Perlmutter *et. al.*, *Measurements of Omega and Lambda from 42 High-Redshift Supernovae*, *Astrophys. J.* **517** (1999) 565–586, [astro-ph/9812133]. 4

[18] **Supernova Search Team** Collaboration, A. G. Riess *et. al.*, *Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant*, *Astron. J.* **116** (1998) 1009–1038, [astro-ph/9805201]. 4

[19] **Supernova Cosmology Project** Collaboration, R. A. Knop *et. al.*, *New Constraints on Ω_M, Ω_Λ, and w from an Independent Set of Eleven High-Redshift Supernovae Observed with HST*, *Astrophys. J.* **598** (2003) 102, [astro-ph/0309368]. 5

[20] R. G. Crittenden and N. Turok, *Looking for Λ with the Rees-Sciama Effect*, *Phys. Rev. Lett.* **76** (1996) 575, [astro-ph/9510072]. 6

[21] S. Ho, C. Hirata, N. Padmanabhan, U. Seljak, and N. Bahcall, *Correlation of CMB with large-scale structure: I. ISW Tomography and Cosmological Implications*, *Phys. Rev.* **D78** (2008) 043519, [arXiv:0801.0642]. 6

[22] T. Giannantonio and C. Porciani, *Structure formation from non-Gaussian initial conditions: multivariate biasing, statistics, and comparison with N- body simulations*, *Phys. Rev.* **D81** (2010) 063530, [arXiv:0911.0017]. 6, 114

[23] **SDSS** Collaboration, D. J. Eisenstein *et. al.*, *Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies*, *Astrophys. J.* **633** (2005) 560–574, [astro-ph/0501171]. 7

[24] R. Amanullah *et. al.*, *Spectra and Light Curves of Six Type Ia Supernovae at 0.511 ¡ z ¡ 1.12 and the Union2 Compilation*, arXiv:1004.1711. 7

[25] S. Weinberg, *Anthropic Bound on the Cosmological Constant*, *Phys. Rev. Lett.* **59** (1987) 2607. 8

[26] R. Bousso and J. Polchinski, *Quantization of four-form fluxes and dynamical neutralization of the cosmological constant*, *JHEP* **06** (2000) 006, [hep-th/0004134]. 8

[27] N. Arkani-Hamed, H. Georgi, and M. D. Schwartz, *Effective field theory for massive gravitons and gravity in theory space*, *Ann. Phys.* **305** (2003) 96–118, [hep-th/0210184]. 9

[28] N. Arkani-Hamed, H.-C. Cheng, M. A. Luty, and S. Mukohyama, *Ghost condensation and a consistent infrared modification of gravity*, *JHEP* **05** (2004) 074, [hep-th/0312099]. 9, 12, 19, 25, 27, 28, 33, 38, 129, 130, 133, 135

[29] G. R. Dvali, G. Gabadadze, and M. Porrati, *4D gravity on a brane in 5D Minkowski space*, *Phys. Lett.* **B485** (2000) 208–214, [hep-th/0005016]. 9

[30] A. Nicolis, R. Rattazzi, and E. Trincherini, *The galileon as a local modification of gravity*, *Phys. Rev.* **D79** (2009) 064036, [arXiv:0811.2197]. 9, 54, 55

[31] A. Albrecht *et. al.*, *Report of the Dark Energy Task Force*, astro-ph/0609591. 9, 11, 47

[32] P. Creminelli, M. A. Luty, A. Nicolis, and L. Senatore, *Starting the universe: Stable violation of the null energy condition and non-standard cosmologies*, *JHEP* **12** (2006) 080, [hep-th/0606090]. 11, 12, 13, 19, 21, 33, 35, 37, 38, 50, 51, 61

[33] C. Cheung, P. Creminelli, A. L. Fitzpatrick, J. Kaplan, and L. Senatore, *The Effective Field Theory of Inflation*, *JHEP* **03** (2008) 014, [arXiv:0709.0293]. 11, 12, 13, 19, 21, 30, 33, 37, 51

[34] M.-z. Li, B. Feng, and X.-m. Zhang, *A single scalar field model of dark energy with equation of state crossing -1*, *JCAP* **0512** (2005) 002, [hep-ph/0503268]. 12

[35] B. Feng, X.-L. Wang, and X.-M. Zhang, *Dark Energy Constraints from the Cosmic Age and Supernova*, *Phys. Lett.* **B607** (2005) 35–41, [astro-ph/0404224]. 12

[36] W. Hu, *Crossing the phantom divide: Dark energy internal degrees of freedom*, *Phys. Rev.* **D71** (2005) 047301, [astro-ph/0410680]. 12, 30

[37] A. Vikman, *Can dark energy evolve to the phantom?*, *Phys. Rev.* **D71** (2005) 023515, [astro-ph/0407107]. 12, 17, 30

[38] R. R. Caldwell and M. Doran, *Dark-energy evolution across the cosmological-constant boundary*, *Phys. Rev.* **D72** (2005) 043527, [astro-ph/0501104]. 12, 30

[39] C. Armendariz-Picon, T. Damour, and V. F. Mukhanov, *k-Inflation*, *Phys. Lett.* **B458** (1999) 209–218, [hep-th/9904075]. 13, 49

[40] C. Armendariz-Picon, V. F. Mukhanov, and P. J. Steinhardt, *A dynamical solution to the problem of a small cosmological constant and late-time cosmic acceleration*, *Phys. Rev. Lett.* **85** (2000) 4438–4441, [astro-ph/0004134]. 13, 49

[41] J. M. Cline, S. Jeon, and G. D. Moore, *The phantom menaced: Constraints on low-energy effective ghosts*, *Phys. Rev.* **D70** (2004) 043543, [hep-ph/0311312]. 16

[42] B. Holdom, *Accelerated expansion and the Goldstone ghost*, *JHEP* **07** (2004) 063, [hep-th/0404109]. 16

[43] R. R. Caldwell, *A Phantom Menace?*, *Phys. Lett.* **B545** (2002) 23–29, [astro-ph/9908168]. 16

[44] J. Garriga and V. F. Mukhanov, *Perturbations in k-inflation*, *Phys. Lett.* **B458** (1999) 219–225, [hep-th/9904176]. 16, 50

[45] C. Bonvin, C. Caprini, and R. Durrer, *Superluminal motion and closed signal curves*, arXiv:0706.1538. 16

[46] E. Babichev, V. Mukhanov, and A. Vikman, *k-Essence, superluminal propagation, causality and emergent geometry*, *JHEP* **02** (2008) 101, [arXiv:0708.0561]. 16

[47] A. Adams, N. Arkani-Hamed, S. Dubovsky, A. Nicolis, and R. Rattazzi, *Causality, analyticity and an IR obstruction to UV completion*, *JHEP* **10** (2006) 014, [hep-th/0602178]. 16

[48] S. D. H. Hsu, A. Jenkins, and M. B. Wise, *Gradient instability for $w < -1$*, *Phys. Lett.* **B597** (2004) 270–274, [astro-ph/0406043]. 16, 51

[49] L. Senatore, *Tilted ghost inflation*, *Phys. Rev.* **D71** (2005) 043512, [astro-ph/0406187]. 19, 33, 37

[50] N. Arkani-Hamed, H.-C. Cheng, M. A. Luty, S. Mukohyama, and T. Wiseman, *Dynamics of Gravity in a Higgs Phase*, *JHEP* **01** (2007) 036, [hep-ph/0507120]. 21, 25, 38, 52, 75, 133, 135

[51] H. Kodama and M. Sasaki, *Cosmological Perturbation Theory*, *Prog. Theor. Phys. Suppl.* **78** (1984) 1–166. 21, 32

[52] S. Mukohyama, *Accelerating universe and cosmological perturbation in the ghost condensate*, *JCAP* **0610** (2006) 011, [hep-th/0607181]. 24

[53] R. Dave, R. R. Caldwell, and P. J. Steinhardt, *Sensitivity of the cosmic microwave background anisotropy to initial conditions in quintessence cosmology*, *Phys. Rev.* **D66** (2002) 023516, [astro-ph/0206372]. 28

[54] J. K. Erickson, R. R. Caldwell, P. J. Steinhardt, C. Armendariz-Picon, and V. F. Mukhanov, *Measuring the speed of sound of quintessence*, *Phys. Rev. Lett.* **88** (2002) 121301, [astro-ph/0112438]. 29

[55] S. DeDeo, R. R. Caldwell, and P. J. Steinhardt, *Effects of the sound speed of quintessence on the microwave background and large scale structure*, *Phys. Rev.* **D67** (2003) 103509, [astro-ph/0301284]. 29, 38, 48

[56] J. Weller and A. M. Lewis, *Large Scale Cosmic Microwave Background Anisotropies and Dark Energy*, *Mon. Not. Roy. Astron. Soc.* **346** (2003) 987–993, [astro-ph/0307104]. 29, 38, 48, 53

[57] R. Bean and O. Dore, *Probing dark energy perturbations: the dark energy equation of state and speed of sound as measured by WMAP*, *Phys. Rev.* **D69** (2004) 083503, [astro-ph/0307100]. 32, 38, 48

[58] W. Hu, *Structure Formation with Generalized Dark Matter*, *Astrophys. J.* **506** (1998) 485–494, [astro-ph/9801234]. 32

[59] S. Hannestad, *Constraints on the sound speed of dark energy*, *Phys. Rev.* **D71** (2005) 103519, [astro-ph/0504017]. 38, 48

[60] M. Takada, *Can A Galaxy Redshift Survey Measure Dark Energy Clustering?*, *Phys. Rev.* **D74** (2006) 043505, [astro-ph/0606533]. 38, 48

[61] A. Torres-Rodriguez and C. M. Cress, *Constraining the Nature of Dark Energy using the SKA*, *Mon. Not. Roy. Astron. Soc.* **376** (2007) 1831–1837, [astro-ph/0702113]. 38, 48

[62] W. Hu and R. Scranton, *Measuring Dark Energy Clustering with CMB-Galaxy Correlations*, *Phys. Rev.* **D70** (2004) 123002, [astro-ph/0408456]. 38, 48

[63] P.-S. Corasaniti, T. Giannantonio, and A. Melchiorri, *Constraining dark energy with cross-correlated CMB and Large Scale Structure data*, *Phys. Rev.* **D71** (2005) 123521, [astro-ph/0504115]. 38, 48

[64] A. R. Zentner, *The Excursion Set Theory of Halo Mass Functions, Halo Clustering, and Halo Growth*, *Int. J. Mod. Phys.* **D16** (2007) 763–816, [astro-ph/0611454]. 39, 45

[65] R. K. Sheth, *Large Scale Structure and Galaxies*, in *American Institute of Physics Conference Series* (M. Novello & S. Perez, ed.), vol. 1132 of *American Institute of Physics Conference Series*, pp. 158–198, May, 2009. 39

[66] J. E. Gunn and J. R. Gott, III, *On the infall of matter into cluster of galaxies and some effects on their evolution*, *Astrophys. J.* **176** (1972) 1–19. 39, 48, 57, 66

[67] W. H. Press and P. Schechter, *Formation of galaxies and clusters of galaxies by selfsimilar gravitational condensation*, *Astrophys. J.* **187** (1974) 425–438. 39, 41, 48, 69

[68] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time*. Cambridge University Press, 1975. 39

[69] J. R. Bond, S. Cole, G. Efstathiou, and N. Kaiser, *Excursion set mass functions for hierarchical Gaussian fluctuations*, *Astrophys. J.* **379** (1991) 440. 43, 48, 69

[70] S. Chandrasekhar, *Stochastic problems in physics and astronomy*, *Rev. Mod. Phys.* **15** (1943) 1–89. 43

[71] M. Maggiore and A. 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*, arXiv:0903.1249. 43, 45, 70, 85, 86, 96, 97, 101, 115, 117, 122

[72] M. Maggiore and A. Riotto, *The halo mass function from the excursion set method. II. The diffusing barrier*, arXiv:0903.1250. 43, 85, 86, 96, 97, 103, 116, 118, 121

[73] M. Maggiore and A. Riotto, *The halo mass function from the excursion set method. III. First principle derivation for non-Gaussian theories*, arXiv:0903.1251. 43, 85, 86, 96, 97, 101, 106, 114, 115, 117, 118, 123, 124

[74] R. K. Sheth and G. Tormen, *Large scale bias and the peak background split*, *Mon. Not. Roy. Astron. Soc.* **308** (1999) 119, [astro-ph/9901122]. 45, 72

[75] A. Jenkins *et. al.*, *Mass function of dark matter halos*, *Mon. Not. Roy. Astron. Soc.* **321** (2001) 372, [astro-ph/0005260]. 45

[76] I. Zlatev, L.-M. Wang, and P. J. Steinhardt, *Quintessence, Cosmic Coincidence, and the Cosmological Constant*, *Phys. Rev. Lett.* **82** (1999) 896–899, [astro-ph/9807002]. 47, 66

[77] P. G. Ferreira and M. Joyce, *Structure formation with a self-tuning scalar field*, *Phys. Rev. Lett.* **79** (1997) 4740–4743, [astro-ph/9707286]. 47

[78] J. M. Alimi *et. al.*, *Imprints of Dark Energy on Cosmic Structure Formation I) Realistic Quintessence Models*, *Mon. Not. Roy. Astron. Soc.* **401** (2010) 775, [arXiv:0903.5490]. 48

[79] O. Lahav, P. B. Lilje, J. R. Primack, and M. J. Rees, *Dynamical effects of the cosmological constant*, *Mon. Not. Roy. Astron. Soc.* **251** (1991) 128–136. 48, 58, 66, 77

[80] L.-M. Wang and P. J. Steinhardt, *Cluster Abundance Constraints on Quintessence Models*, *Astrophys. J.* **508** (1998) 483–490, [astro-ph/9804015]. 48, 58, 65

[81] N. N. Weinberg and M. Kamionkowski, *Constraining dark energy from the abundance of weak gravitational lenses*, *Mon. Not. Roy. Astron. Soc.* **341** (2003) 251, [astro-ph/0210134]. 48

[82] R. A. Battye and J. Weller, *Constraining cosmological parameters using Sunyaev- Zel'dovich cluster surveys*, *Phys. Rev.* **D68** (2003) 083506, [astro-ph/0305568]. 48

[83] D. F. Mota and C. van de Bruck, *On the spherical collapse model in dark energy cosmologies*, *Astron. Astrophys.* **421** (2004) 71–81, [astro-ph/0401504]. 48, 49, 53

[84] W. J. Percival, *Cosmological structure formation in a homogeneous dark energy background*, *Astron. Astrophys.* **443** (2005) 819, [astro-ph/0508156]. 48, 68

[85] P. Creminelli, G. D'Amico, J. Norena, L. Senatore, and F. Vernizzi, *Spherical collapse in quintessence models with zero speed of sound*, *JCAP* **1003** (2010) 027, [arXiv:0911.2701]. 48

[86] I. Maor and O. Lahav, *On virialization with dark energy*, *JCAP* **0507** (2005) 003, [astro-ph/0505308]. 49, 53

[87] M. Manera and D. F. Mota, *Cluster number counts dependence on dark energy inhomogeneities and coupling to dark matter*, *Mon. Not. Roy. Astron. Soc.* **371** (2006) 1373, [astro-ph/0504519]. 49

[88] N. J. Nunes and D. F. Mota, *Structure Formation in Inhomogeneous Dark Energy Models*, *Mon. Not. Roy. Astron. Soc.* **368** (2006) 751–758, [astro-ph/0409481]. 49

[89] P. Creminelli, G. D'Amico, J. Norena, and F. Vernizzi, *The Effective Theory of Quintessence: the $w < -1$ Side Unveiled*, *JCAP* **0902** (2009) 018, [arXiv:0811.0827]. 50, 51, 52, 64

[90] F. K. Manasse and C. W. Misner, *Fermi Normal Coordinates and Some Basic Concepts in Differential Geometry*, *Journal of Mathematical Physics* **4** (June, 1963) 735–745. 54

[91] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation*. San Francisco: W.H. Freeman and Co., 1973. 54

[92] S. Weinberg, *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*. Wiley-VCH, 1972. 56

[93] L. Boubekeur, P. Creminelli, J. Norena, and F. Vernizzi, *Action approach to cosmological perturbations: the 2nd order metric in matter dominance*, *JCAP* **0808** (2008) 028, [arXiv:0806.1016]. 60

[94] P. J. E. Peebles, *Principles of physical cosmology*. Princeton Series in Physics, Princeton, NJ: Princeton University Press, 1993. 63

[95] A. R. Liddle and D. H. Lyth, *Cosmological Inflation and Large-Scale Structure*. Cambridge, UK: Cambridge University Press, Apr., 2000. 65

[96] **WMAP** Collaboration, E. Komatsu *et. al.*, *Five-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations:Cosmological Interpretation*, *Astrophys. J. Suppl.* **180** (2009) 330–376, [arXiv:0803.0547]. 66, 70

[97] S. Mukohyama, *Dark matter as integration constant in Horava-Lifshitz gravity*, *Phys. Rev.* **D80** (2009) 064005, [arXiv:0905.3563]. 75

[98] D. Blas, O. Pujolas, and S. Sibiryakov, *On the Extra Mode and Inconsistency of Horava Gravity*, *JHEP* **10** (2009) 029, [arXiv:0906.3046]. 75

[99] A. Pillepich, C. Porciani, and O. Hahn, *Universal halo mass function and scale-dependent bias from N-body simulations with non-Gaussian initial conditions*, arXiv:0811.4176. 78

[100] S. Borgani, *Cosmology with clusters of galaxies*, astro-ph/0605575. 80

[101] S. D. M. White, J. F. Navarro, A. E. Evrard, and C. S. Frenk, *The Baryon content of galaxy clusters: A Challenge to cosmological orthodoxy*, *Nature* **366** (1993) 429–433. 80

[102] G. M. Voit, *Tracing cosmic evolution with clusters of galaxies*, *Rev. Mod. Phys.* **77** (2005) 207–258, [astro-ph/0410173]. 81

[103] A. Vikhlinin *et. al.*, *Chandra sample of nearby relaxed galaxy clusters: mass, gas fraction, and mass-temperature relation*, *Astrophys. J.* **640** (2006) 691, [astro-ph/0507092]. 81

[104] A. Mantz, S. W. Allen, H. Ebeling, D. Rapetti, and A. Drlica-Wagner, *The Observed Growth of Massive Galaxy Clusters II: X-ray Scaling Relations*, arXiv:0909.3099. 81

[105] M. LoVerde, A. Miller, S. Shandera, and L. Verde, *Effects of Scale-Dependent Non-Gaussianity on Cosmological Structures*, *JCAP* **0804** (2008) 014, [arXiv:0711.4126]. 82, 86, 96, 97, 114, 115, 123

[106] S. Matarrese, L. Verde, and R. Jimenez, *The abundance of high-redshift objects as a probe of non- Gaussian initial conditions*, *Astrophys. J.* **541** (2000) 10, [astro-ph/0001366]. 86, 96, 97, 114, 123, 124

[107] J. Zinn-Justin, *Quantum field theory and critical phenomena*, *Int. Ser. Monogr. Phys.* **113** (2002) 1–1054. 87

[108] N. Sugiyama, *Cosmic background anistropies in CDM cosmology*, *Astrophys. J. Suppl.* **100** (1995) 281, [astro-ph/9412025]. 90, 98

[109] J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay, *The Statistics of Peaks of Gaussian Random Fields*, *Astrophys. J.* **304** (1986) 15–61. 91, 97, 98

[110] J. R. Bond and S. T. Myers, *The Hierarchical peak patch picture of cosmic catalogs. 1. Algorithms*, *Astrophys. J. Suppl.* **103** (1996) 1. 91

[111] E. Audit, R. Teyssier, and J.-M. Alimi, *Non-linear Dynamics and Mass Function of Cosmic Structures: I Analytical Results*, *Astron. Astrophys.* **325** (1997) 439–449, [astro-ph/9704023]. 91

[112] R. K. Sheth, H. J. Mo, and G. Tormen, *Ellipsoidal collapse and an improved model for the number and spatial distribution of dark matter haloes*, *Mon. Not. Roy. Astron. Soc.* **323** (2001) 1, [astro-ph/9907024]. 91, 97, 117

[113] V. Desjacques, *Environmental dependence in the ellipsoidal collapse model*, *Mon. Not. Roy. Astron. Soc.* **388** (2008) 638, [arXiv:0707.4670]. 91

[114] S. Redner, *A guide to first-passage processes*. Cambridge University Press, 2001. 92, 93

[115] R. K. Sheth and G. Tormen, *An Excursion set model of hierarchical clustering : Ellipsoidal collapse and the moving barrier*, *Mon. Not. Roy. Astron. Soc.* **329** (2002) 61, [astro-ph/0105113]. 92, 96, 114

[116] B. Sartoris *et. al.*, *The potential of X-ray cluster surveys to constrain primordial non-Gaussianity*, arXiv:1003.0841. 96

[117] C. Cunha, D. Huterer, and O. Dore, *Primordial non-Gaussianity from the covariance of galaxy cluster counts*, arXiv:1003.2416. 96

[118] E. Sefusatti, *1-loop Perturbative Corrections to the Matter and Galaxy Bispectrum with non-Gaussian Initial Conditions*, *Phys. Rev.* **D80** (2009) 123002, [arXiv:0905.0717]. 96

[119] R. Jimenez and L. Verde, *Implications for Primordial Non-Gaussianity ($f_{\rm NL}$) from weak lensing masses of high-z galaxy clusters*, *Phys. Rev.* **D80** (2009) 127302, [arXiv:0909.0403]. 96

[120] L. Verde, *Non-Gaussianity from Large-Scale Structure Surveys*, arXiv:1001.5217. 96

[121] V. Desjacques and U. Seljak, *Primordial non-Gaussianity from the large scale structure*, arXiv:1003.5020. 96

[122] T. Y. Lam and R. K. Sheth, *Halo abundances in the f_{nl} model*, arXiv:0905.1702. 97, 115

[123] D. Babich, P. Creminelli, and M. Zaldarriaga, *The shape of non-Gaussianities*, *JCAP* **0408** (2004) 009, [astro-ph/0405356]. 99

[124] D. H. Lyth, C. Ungarelli, and D. Wands, *The primordial density perturbation in the curvaton scenario*, *Phys. Rev.* **D67** (2003) 023503, [astro-ph/0208055]. 100

[125] N. Bartolo, S. Matarrese, and A. Riotto, *On non-Gaussianity in the curvaton scenario*, *Phys. Rev.* **D69** (2004) 043503, [hep-ph/0309033]. 100

[126] G. Dvali, A. Gruzinov, and M. Zaldarriaga, *A new mechanism for generating density perturbations from inflation*, *Phys. Rev.* **D69** (2004) 023505, [astro-ph/0303591]. 100

[127] M. Alishahiha, E. Silverstein, and D. Tong, *DBI in the sky*, *Phys. Rev.* **D70** (2004) 123505, [hep-th/0404084]. 100

[128] N. Arkani-Hamed, P. Creminelli, S. Mukohyama, and M. Zaldarriaga, *Ghost Inflation*, *JCAP* **0404** (2004) 001, [hep-th/0312100]. 100

[129] P. Creminelli, *On non-gaussianities in single-field inflation*, *JCAP* **0310** (2003) 003, [astro-ph/0306122]. 100

[130] P. Creminelli, L. Senatore, M. Zaldarriaga, and M. Tegmark, *Limits on $f_{\rm NL}$ parameters from WMAP 3yr data*, *JCAP* **0703** (2007) 005, [astro-ph/0610600]. 100

[131] P. Creminelli, A. Nicolis, L. Senatore, M. Tegmark, and M. Zaldarriaga, *Limits on non-Gaussianities from WMAP data*, *JCAP* **0605** (2006) 004, [astro-ph/0509029]. 100

[132] C. Carbone *et. al.*, *The properties of the dark matter halo distribution in non-Gaussian scenarios*, *Nucl. Phys. Proc. Suppl.* **194** (2009) 22–27. 114

[133] F. Bernardeau, S. Colombi, E. Gaztanaga, and R. Scoccimarro, *Large-scale structure of the universe and cosmological perturbation theory*, *Phys. Rept.* **367** (2002) 1–248, [astro-ph/0112551]. 115

[134] B. Robertson, A. Kravtsov, J. Tinker, and A. Zentner, *Collapse Barriers and Halo Abundance: Testing the Excursion Set Ansatz*, *Astrophys. J.* **696** (2009) 636–652, [arXiv:0812.3148]. 116, 117, 118, 121

[135] R. K. Sheth and R. van de Weygaert, *A hierarchy of voids: Much ado about nothing*, *Mon. Not. Roy. Astron. Soc.* **350** (2004) 517, [astro-ph/0311260]. 124

[136] M. Kamionkowski, L. Verde, and R. Jimenez, *The Void Abundance with Non-Gaussian Primordial Perturbations*, *JCAP* **0901** (2009) 010, [arXiv:0809.0506]. 124

[137] J. Polchinski, *Effective field theory and the Fermi surface*, hep-th/9210046. 129

[138] J. Z. Simon, *The Stability of flat space, semiclassical gravity, and higher derivatives*, *Phys. Rev.* **D43** (1991) 3308–3316. 130

[139] J. Z. Simon, *Higher Derivative Lagrangians, Non-locality, Problems and Solutions*, *Phys. Rev.* **D41** (1990) 3720. 130

[140] R. Kallosh, J. U. Kang, A. D. Linde, and V. Mukhanov, *The New Ekpyrotic Ghost*, *JCAP* **0804** (2008) 018, [arXiv:0712.2040]. 130

[141] S. Weinberg, *Effective Field Theory for Inflation*, *Phys. Rev.* **D77** (2008) 123541, [arXiv:0804.4291]. 130

[142] N. Arkani-Hamed, S. Dubovsky, A. Nicolis, E. Trincherini, and G. Villadoro, *A Measure of de Sitter Entropy and Eternal Inflation*, *JHEP* **05** (2007) 055, [arXiv:0704.1814]. 131

[143] A. Erdélyi, *Asymptotic Expansions*. New York, NY: Dover, 1956. 139