Mass dependence of the Baryon Acoustic Oscillations in N-body simulations

Nataly Mateus Londoño

FACom, Grupo de física y astrofísica computacional

Instituto de física

Facultad de ciencias exactas y naturales

Advisor: Juan Carlos Muñoz Cuartas

UNIVERSIDAD DE ANTIOQUIA

FACULTAD DE CIENCIAS EXACTAS Y NATURALES

INSTITUTO DE FÍSICA

UNIVERSIDAD
DE ANTIOQUIA
1 8 0 3

Mass dependence of the Baryon Acoustic Oscillations in N-body simulations

AUTHOR:

_____

(Nataly Mateus Londoño)

ADVISOR:

_____

(Professor Juan Carlos Muñoz Cuartas)

Contents

Introduction

In the standard model of cosmology the universe was born in a big bang, an explosion that produced an expanding, isotropic and homogeneous Universe. From observations it has been found that this expansion is currently accelerating with time (Hamuy et al.,1996).

There are several components of the matter-energy content of the universe, dark and baryonic matter, radiation and dark energy. According to recent estimations, the last one accounts for around 70% of this content and is responsible for the accelerated expansion of the universe. The baryonic acoustic oscillations allows to study the nature of this expansion as it will be explained.

In the early universe the dark matter (DM) formed density fluctuations, causing baryonic matter to be unstable against gravitational perturbations. At this stage in the evolution of the universe the temperature was very high, allowing a coupling between baryonic matter and radiation through Thomson scattering. So the increase of baryonic matter in the DM density fluctuations not only caused an increase in density, but also radiation pressure against collapse. Therefore, an expanding wave centered in the fluctuation is caused because of the radiation pressure. This wave is the baryonic acoustic oscillation (BAO) (Hu and Sugiyama, 1996; Eisenstein and Hu, 1998).

Nevertheless, it is necessary to consider that the universe is expanding and this results in a temperature decrease. Therefore, when temperature is low enough the baryonic matter and

radiation decoupled, making BAO to stop expanding and leaving an imprint in the matter distribution. The distance that a BAO could have travelled by the time of decoupling is called sound horizon. This scale has been measured in the Cosmic Microwave Background as 146.8 ± 1.8Mpc, ([1]).

Since BAO do not change in size after decoupling they can be used as a standard ruler. They allow to measure the Hubble parameter and angular diameter distance as a function of z, and this way to measure the rate of expansion at different times during the evolution of the universe. Hence, BAO is key to constraint dark energy parameters.

A way to observe the imprint let by BAO is through the 2D point correlation function or the power spectrum that is its fourier pair, ([2], [3]). A peak due to the BAO appears in the correlation function (see figure 2.5) but there are several issues to take into consideration. For example, the non-linear clustering smear out the BAO imprint causing a broadening of the peak (Crocce and Scoccimarro, 2008). These, among other problems, have to be taken into account when BAO are studied.

In the present work, we plan to study the BAO from numerical cosmological simulations. More precisely, the BAO will be studied trying to answer what are the changes of the BAO's properties with the change of the scale of the tracer halo population? is there any change in the position peak? is there any change in the width peak? In general, the question we want to answer is: Is there any dependence in the width and amplitude of the BAO signal with the tracer halo population? Answering this questions will lead not only to profound understanding of the physics of BAO but a better understanding of the accelerated expansion of the universe that still has so many questions to be answered.

CHAPTER 2

Cosmological Background

Cosmology is the branch of physics that studies the Universe as a whole, therefore, it attempts to explain its origin, evolution and structure at big scales. Hence, a coarse grained approximation is mandatory due to the scales considered, this is, several approximations are necessary in the endeavour of such a task.

In this search, two major points are considered. The first one is the cosmological principle, it assumes that on sufficiently large scales the Universe is homogeneous and isotropic. In this context, homogeneity can be understood like invariance under translation and isotropy like invariance under rotation. Then, this principle establishes that for certain observers the Universe should appear isotropic and homogeneous, the fundamental observers. Since the Universe is expanding, the distance among fundamental observers changes with time but in an uniform way. Using these observers it is possible to synchronize clocks with a light pulse, the time measured by them is named cosmic time.

The overall isotropy and homogeneity have been found in observations of the cosmic microwave background radiation (CMB) and the sponge like structure of the distribution of galaxies. Until few years ago, all observations have agreed with this asseveration, however, recent evidence from Planck data have shown that anisotropies can appear at big scales [4].

The sponge like structure refers to the fact that, nowadays at cosmological scales, big structures such as halos, sheets and voids forming a filamentary distribution are observed. They are explained through the growth of initial seeds, small density perturbations at the

3

early Universe that evolve due to gravitational instabilities. These processes are occurring in an expanding Universe. This latter affirmation is predicted by the theory of general relativity in which modern cosmology is based, a second important point to consider. Here, Einstein field equations (EFE) serve as a set of fundamental equations to study the evolution of the Universe at big scales. Fortunately, isotropy and homogeneity led to a simple form of these ones and hence a relatively simple mathematical treatment in cosmology may be developed. From EFE, Friedmman equations are obtained which provide a theoretical framework to study the Universe expansion.

A standard model in cosmology that takes into consideration the aspects exposed previously is ΛCDM, where additionally to an expanding universe, there is a dark energy component that accelerates its expansion. This is precisely the framework that is going to be used in this work.

In this chapter, several basic concepts in ΛCDM standard model are going to be introduced to finally lead to bayronic acoustic oscillations (BAO). Several textbooks and articles were used as reference for this chapters such as [5], [6], [7], [8].

2.1   Robertson Walker Metric

As was mentioned before, observations of the Universe at big scales show that it is homogeneous and isotropic, at least as a good approximation, i.e., inhomogeneities appear also at big scales in the CMB. Nevertheless, it is taken as a postulate for ΛCDM cosmology . Let's see this in more detail

- Cosmological principle: *The Universe is homogeneous and isotropic at big scales.*

  In this context, homogeneous refers to that the fact that independently of where the reference system is located we are going to observe the same global patterns, i.e., the structure of the observed Universe is the same no matter the reference system used. On the other hand, isotropy establishes that regardless of the direction chosen, the same structure is going to be observed. Then, we are dealing with translational and rotational symmetry.

  These characteristics are observed on mega parsec scales, i.e., big scales. However, this is only valid for the actual epoch, the scale changes with time due to the expansion of the Universe.

- Weyl postulate : *Establishes that the geodesics, world lines of galaxies, do not intersect except in a singular point in the past.*

  This one defines a set of observers that move along the geodesics. The interception point allows to synchronize watches among different observers, defining a cosmic time. Therefore, the distance between galaxies can be measured at the same cosmic time.

As already stated the Universe is expanding. It was due to a research on nearby galaxies performed by Edwin Hubble that a redshift was found in most of the galaxies, i.e., they are moving away from us. Considering this movement, one could conclude we are in the center of the expansion. But this conclusion is wrong since the expansion Hubble law is valid independently where the coordinate system is defined.

A metric that satisfies homogeneity and isotropy and additionally contains a term that accounts for the Universe's expansion is the Robertson Walker metric. It is defined in general terms as $ds^2 = g_{\mu\nu}dx^\mu dx^\nu$, where $g_{\mu\nu}$ is the metric tensor and uses coordinates $x^\alpha = \{ct, x, y, z\}$. The metric tensor takes the next form $g_{\mu\nu} = diag\{1, -\frac{a^2}{1-k^2} - a^2r^2, -a^2r^2\sin^2\theta\}$, and the metric is

$$ds^2 = c^2dt^2 - a(t)^2 \left[\frac{d^2r}{1 - Kr^2} + r^2(d^2\theta + \sin^2\theta d^2\phi) \right] \tag{2.1}$$

The term a is the scale factor, it describes how the relative distance between two fundamental observers changes with time. The term K is the curvature constant for the actual time and defines the Universe geometry. When $K = 0$ an euclidean metric is recovered leading to a flat universe expanding indefinitely. If $K = 1$ the Universe would be described by a spherical geometry and it would collapse because of its energy matter density content. And finally, $K = -1$ corresponds to a hyperbolic geometry where the Universe would be in accelerated expansion. In the figure 2.1 we show the behavior of different cosmologies depending on the K value and density content.

One important aspect to consider is that the geometry depends on the total energy-matter density content, Ω_o. This can be concluded from the definition of the curvature constant $K = H_o^2(\Omega_o - 1)/c^2$.
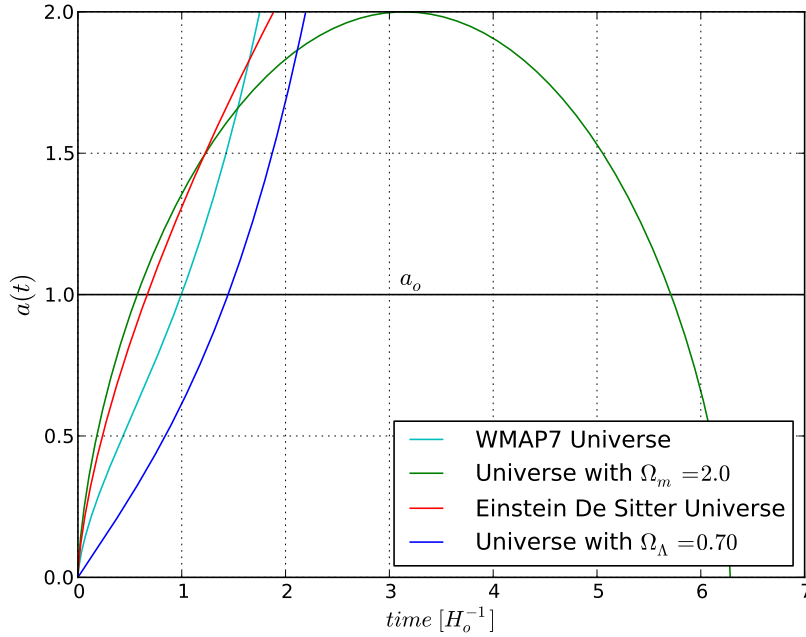
Figure 2.1:   Scale factor as a time function. The Universe expansion for different density contributions. A closed Universe is obtained when $\Omega_m = \Omega_o > 1$. Also, the PLANCK parameters show an accelerated expansion.

2.2   Hilbert-Einstein field equation

At big scales, the most important fundamental interaction of nature is the gravitational one. Hence, the theory of general relativity (TGR) is an essential tool in the study of the cosmos. At smaller scales, the Newtonian gravitational theory might be used, where the Poisson equation offers a relation between the second derivative of the gravitational potential Φ and the source of the field ρ

$$\nabla^2 \Phi = 4\pi G \rho$$

this equation is obtained from TGR for low velocities and a weak gravitational field $(\Phi/c^2 << 1)$. A key equation of TGR is the Hilbert-Einstein field equation, a 6 independent component tensorial equation

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R - g_{\mu\nu} \Lambda = \frac{8\pi G}{c^4} T_{\mu\nu} \tag{2.2}$$

in equation 2.2 the first term of the left side is Ricci's tensor (second derivatives of the metric tensor). The second one contains the scalar of curvature R that defines space-time

geometry. In the third term, Λ is the cosmological constant that could be associated with the vacuum density term and is responsible for the accelerated expansion of the Universe.

In the right side of the equation, the tensor energy-momentum $T_{\mu\nu}$ is present. It includes, as its name suggest, all the contributions to energy and momentum.

Hence, the left side of the equation has associated geometry terms, while the right one, the ones associated with the matter and energy distribution. Then, it could be interpreted as if geometry is determined by the matter-energy content of the Universe, though, strictly speaking, the energy-momentum tensor depends in the metric tensor too.

There is an interesting case of this tensor when we are dealing with a perfect fluid, i.e., without viscosity. It shows that not only density causes curvature of space-time but also pressure. The Universe can be modeled with this particular shape of the energy-momentum tensor. It can be expressed as

$$T_\sigma^\mu = diag\{c^2\rho, -P, -P, -P\}$$

where ρ is the density and P is the fluid pressure.

There are several solutions to the Einstein field equation but not many in an analytical form. An analytical solution is Schwarzchild's solution that represents the metric of a static spherical mass. Other possible solution is the Kerr metric that corresponds to a rotating uncharged mass. The Robertson Walker metric satisfies these equations too.

2.3   Friedmann equations

From HE field equations and the RW metric it is possible to propose cosmological models that give account for the observed dynamics in the Universe. In this direction, the components of the field equation can be taken, $\beta = \nu = 0$, time-time component, and $ii = 1, 2, 3$ (space-time components), from where one gets, for the scale factor

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + 3\frac{P}{c^2}\right) + \frac{\Lambda c^2}{3} \tag{2.3}$$

$$\frac{\ddot{a}}{a} + 2\frac{\dot{a}^2}{a^2} + 2\frac{c^2 K}{a^2} = 4\pi G\left(\rho - \frac{P}{c^2}\right) + \Lambda c^2 \tag{2.4}$$

where, it has been used the energy momentum tensor for an ideal fluid. The former expressions are the Friedmann equations and are written in a form that can account for the Universe expansion. In equations 2.3 and 2.4 $a(t)$ is the scale factor that is set to one for the actual epoch, $a(t_o) = 1$, ρ is the total density (radiation plus matter density), P is the total pressure.

The equation 2.3 has the form of force equation and it can be partially deduced from newtonian mechanics (without the pressure and cosmological constant terms). A most convenient and used form is obtained after algebraically manipulating them, a second form of FE that can be interpreted as an energy equation

$$H(t) = \frac{\dot{a}^2}{a^2} = \frac{8\pi G}{3}\left(\rho + \frac{\Lambda c^2}{8\pi G}\right) - \frac{Kc^2}{a^2} \tag{2.5}$$

where the first term in the right hand side is the potential energy. This equation also allows to define the Hubble parameter since $H(t) = \dot{a}^2/a^2$ and for the actual epoch its value is equal to $H(t_o) = H_o = 100h \text{ Km } s^{-1} \ Mpc^{-1}$ where $h = 0.6774$ according to Planck measures.

Additionally 2.5 can be expressed in terms of the critical density, i.e. the matter-energy density required for a flat Universe. Therefore, if the Universe has a bigger density, $\rho > \rho_{crit}$, it would collapse about itself. Conversely, the Universe would continue to expand indefinitely. The critical density is defined as

$$\rho_{crit}(t) = 3H(t)^2/8\pi G$$

Dividing 2.5 by the Hubble constant H_o and defining the density parameter $\Omega_{i,o} = \rho_{i,o}/\rho_{crit}(t_o)$ with $i = m, r, \Lambda$ is obtained the next expression, thus different contributions of the density to the Hubble parameter are observed, i.e., matter, radiation and vacuum density

$$\frac{H^2(z)}{H_o^2} = \Omega_{m,o}\left(1+z\right)^3 + \Omega_{r,o}\left(1+z\right)^4 + \Omega_{\Lambda,o} + \left(1 - \Omega_o\right)\left(1+z\right) \tag{2.6}$$

where $\Omega_o = \Omega_{m,o} + \Omega_{r,o} + \Omega_{\Lambda,o}$ and it has been introduced the relation between redshift and scale factor $1 + z = 1/a$. Every term of matter-energy density is a different function of the Universe expansion, although the vacuum energy does not depend on the redshift, this is, is constant through time.

Initially the Universe was dominated by the radiation, during this epoch matter and radiation were coupled, i.e., the De Broglie electrons wavelenght were comparable to the

wavelength of photons. Because of this, collisions between photons and electrons were very frequent causing that the mean free path of the photons be negligible and that the Universe would be opaque. During this coupling, radiation and matter had the same temperature and its behavior is explained as a black body. As can be seen in the figure 2.2, from $z = 3230$ matter becomes the major contribution to the Universe density. When $z = 1100$ the temperature drop is big enough and the recombination rate turns higher than the ionization one. The last radiation dispersion due to matter still can be observed, and it is called cosmic radiation background (CMB). Because of the Universe expansion, its temperature has been dropping, and it is nowadays around $T = 2.7K$.
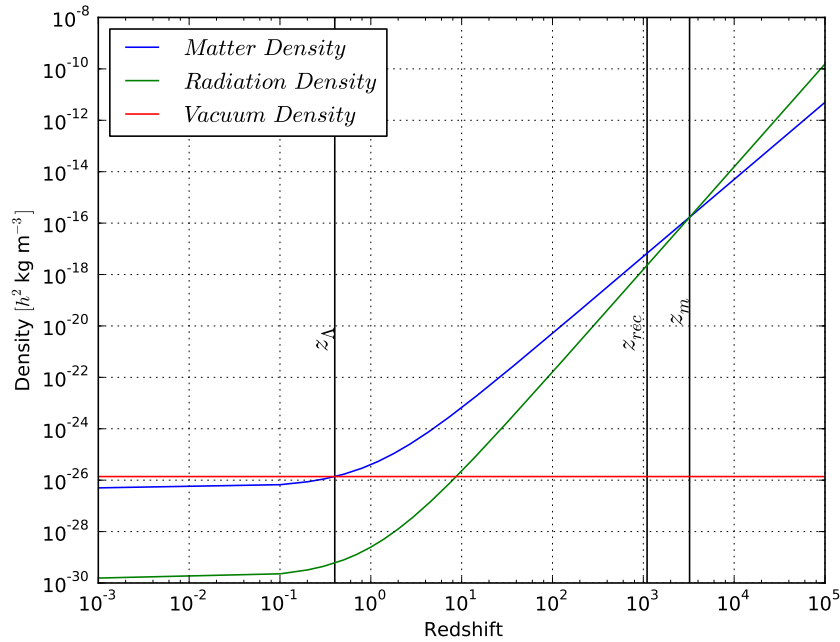

Figure 2.2: Dependence in redshift for Ω_Λ, Ω_m and Ω_r. The decoupling between matter and radiation is obtained when z_{rec}.

Nowadays, the dominant density component in the Universe is vacuum, ρ_Λ, though it is a constant since it does not depend on the scale factor, as already stated. In contrast, matter depends on the scale factor as a^{-3} and radiation as a^{-4} causing both components diminish in time.

The cosmological constant could be associated to vacuum energy that causes an opposed behavior in the Universe dynamics compared to mass density, i.e., it accounts for the accelerated universe expansion.

There are several solutions to 2.5, for instance in the Einstein de Sitter Universe, there

Parameter	Symbol	Best fit
Hubble constant (km/Mpc-s)	H_0	67.74±0.46
Baryon density	$\Omega_b h^2$	0.02230±0.00014
Cold dark matter density	$\Omega_c h^2$	0.1188±0.0010
Dark energy density	Ω_Λ	0.6911±0.0062
Scalar spectral index	n_s	0.9667±0.0040
Sigma 8	σ_8	0.8159±0.0086

Table 2.1:   Cosmological parameters from Planck results[4].

are no radiation or vacuum contributions to the density and the total density is $\Omega_o = 1.0$. In this particular case, the solution is

$$t = \frac{2}{3H_o}(1+z)^{-3/2}$$

Therefore, depending on the chosen density values, the equation 2.5 has different solutions and one can expect several Universe models, i.e., depending on the parameters chosen, the Universe's evolution changes. In the case of Planck, the parameters used are shown in table 2.1.

Other possible Universe models are for example, one obtained when matter density parameter is the only contribution to total universe density but it is bigger than 1. In such a case the Universe obtained is closed. Other one, it is one obtained when the Universe is dominated for the vacuum contribution. In this case, the Universe is always open. When all the contributions are present, the Universe can be open or closed depending on the total density parameter.

2.4   Equation of state

As mentioned before, scale factor characterizes the Universe expansion, then finding relations between each density component of the Universe with this factor is an important task. Below is shown only a relation of proportionality between them. For further details [7] is a good reference.

Matter density: Assuming that all matter content in the Universe is an isolated system, the first law of thermodynamics is expressed as $dU = -pdV$, where relativistic terms are

included in the internal energy term. Using the equipartition theorem and deriving the internal energy with respect to the scale factor is obtained

$$T \propto a^{-2}$$

but from the equation of state $P = NkT$ and taking into account that $N = N_o a^{-3}$

$$P \propto a^{-5} \tag{2.7}$$

Pressure exerted by matter decreases strongly with Universe's expansion, while density and temperature changes are more slow. The latter is other cause in order that vacuum dominates the Universe expansion.

Radiation density: Radiation energy density expressed in terms of the photon density, $N(\nu)$, is

$$\xi = \sum_{\nu} N(\nu) h\nu$$

where $N(\nu)$ satifies the relation $N \propto (1 + z)^3$ and $\nu \propto (1 + z)$ such that $\xi \propto a^{-4}$. Comparing the last result with Stefan-Boltzmann law one can concluded that

$$T \propto a^{-1}$$

Finally, radiation pressure dependence on scale factor is found with the equation $P = \epsilon_{total}/3$, so

$$P \propto a^{-4} \tag{2.8}$$

Vacuum density: On the other hand, vacuum satisfies $\epsilon total = \rho c^2$ where ρ is an effective density. Replacing this result in the first law of thermodynamics and deriving with respect to scale factor

$$P = -\rho c^2 = -\frac{\Lambda c^4}{8\pi G} \tag{2.9}$$

the vacuum density constancy is used for the deduction of the latter expression.

2.5   Perturbation evolution in the newtonian regime

As already stated, there is no radiation comming toward us from a previous epoch to decoupling. Although, due to the last scattering between radiation and matter, highly homogeneous and isotropic distribution of matter is inferred from the patterns obtained from background cosmic radiation[1] as shown in Figure 2.3.

In the CMB radiation, small temperature perturbations are observed indicating the presence of small matter perturbations at this epoch. These are the initial seeds from where structures observed nowadays formed.

In this structure growth, density perturbations are increasing but it is not until they got a size of $\delta \sim 1$ that this growth is dominated by gravity. The perturbations have grown enough to start talking about galaxy formation when their density gets around 1×10^6 compared to the background density, this happens for a epoch around $z \sim 100$ for halos with gas that have cooled enough.

But, it is still important to study the initial stages of the perturbations. Because of this, a linear regime treatment for perturbations when $\delta \ll 1$ are key in such a study.
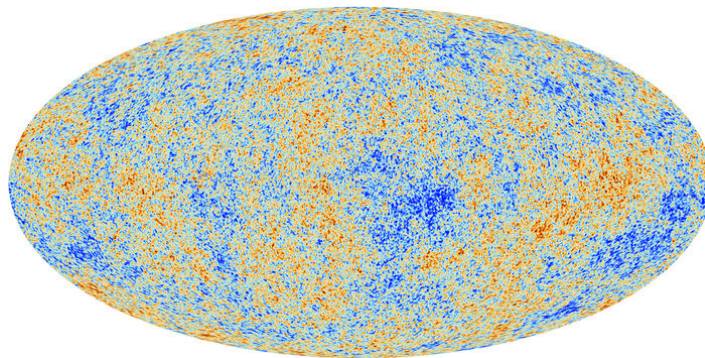

Figure 2.3: Cosmic background radiation image obtained by Planck satellite.

2.5.1   Newtonian description

Inflation is an exponential expansion of the space-time in the initial stages of the Universe. In this stage quantum perturbations were magnified to cosmic size, forming inhomogeneities of the density field, the seeds for the large scale structure of the Universe. Here, the initial density perturbations considered have a characteristic length much smaller such that the relativistic effects can be neglected, i.e., perturbations do not affect the local metric.

[1] Image taken from http://www.esa.int/spaceimages/Images/2013/03/Planck_CMB

In this Newtonian approximation the equations of gas dynamics for a fluid in a gravitational field can be considered. For a fluid in motion with a velocity distribution $\mathbf{u}$ and density ρ subject to a gravitational ϕ field that suffers changes in its pressure P satisfies

$$
\begin{aligned}
\frac{d\rho}{dt} &= -\rho \nabla_r \cdot \mathbf{u} \\
\frac{d\mathbf{u}}{dt} &= -\frac{\nabla_r P}{\rho} - \nabla_r \phi \\
\nabla_r^2 \phi &= 4\pi G \rho
\end{aligned}
\tag{2.10}
$$

where an Eulerian description is used, i.e., the partial derivatives in the expressions 2.10 describe the variations of the properties at a fixed point in space, r is the proper coordinate.

Since density perturbations are the ones that trigger potential wells, they are more of our interest than the density field itself. Then, it is useful to express the density as $\rho = \overline{\rho}(1 + \delta)$, where $\overline{\rho}$ is the background density and δ is the overdensity of interest.

Additionally, it is necessary to make clear another point. The velocity of the particles have two different contributions, the first one is because of the Universe expansion and the other one is the proper velocity of the particle, recessional and peculiar velocities respectively. From the latter mentioned, the coordinate system can be changed from 2.10 an Euler description to a Lagrangian one, i.e., moving with the Universe expansion. Let us see this in more detail, velocity in an Eulearian description is $\mathbf{u} = a\dot{\mathbf{x}} + \mathbf{x}\dot{a} = \mathbf{v} + \mathbf{x}\dot{a}$, where $\mathbf{v}$ is the peculiar velocity and $\mathbf{x}\dot{a}$ is the Universe expansion velocity. Then, transforming in comoving coordinates, coordinates that move with the Universe expansion, and changing the density ρ to density perturbations δ, the equations 2.10 can be written as

$$
\begin{aligned}
\frac{\partial \delta}{\partial t} &= -\frac{1}{a} \nabla \cdot [(1 + \delta)\mathbf{v}] \\
\frac{\partial \mathbf{v}}{\partial t} + \frac{\dot{a}}{a}\mathbf{v} + \frac{1}{a}(\mathbf{v} \cdot \nabla)\mathbf{v} &= -\frac{\nabla \Phi}{a} - \frac{\nabla P}{a\overline{\rho}(1 + \delta)} \\
\nabla^2 \Phi &= 4\pi G \overline{\rho} a^2 \delta
\end{aligned}
\tag{2.11}
$$

the first equation corresponds to the continuity equation, the second one is Euler's equation and the last one is Poissonian gravitational field equation with $\Phi = \phi + a\ddot{a}x^2/2$. Velocity components appear due to gravitational interactions and changes in pressure.

Additionally, equation of state relating the thermodynamic quantities pressure P, density ρ and entropy s for this cosmological fluid is

$$P(\rho, s) = \left[\frac{h^2}{2\pi(\mu m_p)^{5/3}} e^{-5/3}\right] \rho^{5/3} \exp\left(\frac{2}{3}\frac{\mu m_p s}{k_B}\right) \tag{2.12}$$

Manipulating algebraically the continuity equation, Poisson equation and state equation, a wave equation for density perturbations can be obtained

$$\frac{\partial^2 \delta}{\partial t^2} + 2\frac{\dot{a}}{a}\frac{\partial \delta}{\partial t} = 4\pi G\overline{\rho}\delta + \frac{C_s^2}{a^2}\nabla^2\delta + \frac{2}{3}\frac{\overline{T}}{a^2}\nabla^2 s \tag{2.13}$$

where $\overline{T}$ is the background temperature and C_s is the speed of sound. The Universe expansion is seen in the second term in the left hand side. Since for an expanding Universe the term $\dot{a}/a$ is positive, its effect opposes to the perturbation growth. This result was expected since the expansion is against collapse leading to a decrease in growth.

The right hand side shows the causes for the evolution of the density perturbations. Entropy can be considered as the heat interchange between the perturbation and its surroundings, causing the expansion or growth of the perturbation. As expected, gravitational field is a source for perturbation growth.

A solution to the perturbation equation in terms of Fourier series is proposed as

$$\delta(x,t) = \sum_k \delta_k(t) e^{ik\cdot x}$$
$$s(x,t) = \sum_k s(t) e^{ik\cdot x}$$

where **k** is the wave number and δ_k is a density mode that can be calculated using the discrete Fourier transform of the density field. Hence, every mode depends on all known values of the density perturbations.

An important aspect in the last expression is the independence of the functions $e^{ik\cdot x}$ allowing equation 2.13 to be expressed as

$$\frac{d^2\delta_k(t)}{dt^2} + 2\frac{\dot{a}}{a}\frac{d\delta_k(t)}{dt} = \left[4\pi G\overline{\rho} - \frac{C_s^2 k^2}{a^2}\right]\delta_k(t) - \frac{2}{3}\frac{\overline{T}}{a^2}k^2 s_k(t) \tag{2.14}$$

the solution of this equation provides expansion coefficients for the Fourier series, from where, the behavio r of density perturbations, their growth or dissipation, is obtained.

2.5.2   Jeans Inestability

Before solving equation 2.14, it is important to develop some intuition about the physical phenomena. This can be achieved making some simplifications, for example, taking an isentropic static Universe ($\dot{a} = 0$) the expression becomes

$$\frac{d^2\delta_k(t)}{dt^2} + \omega^2\delta_k(t) = 0$$

with $\omega^2 = C_s^2 k^2/a^2 - 4\pi G\bar{\rho}$. Now, let us analyze the different solutions of the previous expression. When the relation $C_s^2 k^2/a^2 > 4\pi G\bar{\rho}$ is satisfied the frequency ω is positive. The solution obtained represents a sound wave, an oscillatory solution where gravity instabilities are balanced by radiation pressure. Thus, this particular solution is not of our interest, gravity is not strong enough to agglomerate matter. On the other hand, if $4\pi G\bar{\rho} > C_s^2 k^2/a^2$ the solution takes the form $\delta_k(t) \propto e^{\Gamma_k t}$, where $\Gamma_k = i\omega_k$ is the growth rate. Hence, the density mode can grow or dissipate depending on the growth rate, a negative rate causes dissipation but a positive one produces a gravitational collapse.

Therefore, density modes tend to collapse because of gravitational instabilities but in this process, pressure gradients appear due to atomic interactions causing dissipation. Nevertheless only the perturbations that collapse are the ones of our interest, they are the seeds of the large scale structure observed nowadays.

Furthermore, one can defined for a perturbation, a minimum length to obtain an unstable evolution, $\lambda_J = 2\pi/k_j = C_s(\pi/G\rho)^{1/2}$, i.e., a collapsing perturbation. This length is called the Jeans' length and it is used to rewrite the growth rate in equation 2.15 . Now, with Jeans' length we could know the evolution of a perturbation, if $\lambda_{pert} >> \lambda_J$ is satisfied the perturbation collapses, where λ_{pert} is the length of the perturbation. A similar analysis can be performed with the Jean's mass defined as $M_J = \pi\bar{\rho}_{m,o}\lambda_J^3$. The latter expressions can be expressed as

$$\lambda_J \approx 0.01(\Omega_{b,o}h^2)^{-1/2} Mpc$$

$$M_J \approx 1.5 \times 10^5(\Omega_{b,o}h^2)^{-1/2} M_\odot$$

but only for an epoch before decoupling. During this time, speed of sound was affected not only by matter but for radiation, the last one being the most important contribution to density as shown in the figure 2.2. Since there is a change in the behavior of the density components with decoupling and this way in the speed of sound, it appears a change in the Jean's length and Jean's mass. It is around 2.6×10^{-5} for Jean's length and 1.8×10^{-14} for Jean's mass. From the previous asseveration a possible conclusion is that decoupling increases the gravitational collapse since the minimum characteristic length required for collapsing becomes smaller.

2.5.3   Evolution of perturbations

Until now dark matter has not been mentioned, a component that contrary to baryonic matter does not interact with radiation. But it is around 23% of the overall Universe matter-energy density content, making it responsible for the large scale mass distribution. First, let us see some observational evidence of dark matter.

One of the first observational evidence was found in the Coma cluster due to a mass estimation from the virial theorem, let us see this in more detail: the specific kinetic energy of the system is $T = v^2/2 \sim 3\sigma^2/2$ where σ is the galaxy velocity dispertion and the potential energy is $U = 3GM_{vir}/(5R_{vir})$. From the mass-luminosity ratio and the mean luminosity of the cluster, a second estimative of the mass is found. There is such a big discrepancy between the two values that is reasonable to affirm that $\sim 90\%$ of the cluster's mass is not visible.

The rotation curve of the galaxies can be other prove for dark matter existence, for example, the velocity measures performed with respect to the radious of Andromeda galaxy (or another spiral galaxies) is approximately the same independent of the radial distance of the stars to the center of the galaxy. From this, it could be affirmed that density is uniform along the galaxy contrary to expected for the observed number of star in function of the radius.

Another example is the Bullet cluster, composed by two two clusters that are colliding, an event not commonly observed. The gas of them reaches velocities around $\sim 10 \; millions \; of \; miles/h$ during the violent collision while they interact among them because of their charge. This interaction disminishes the gas velocity but this does not happen with dark matter cause it does not interact electrically. Using gravitational lensing a distortion map is obtained. Using the X rays detected and the distortion map, four different groups of matter are found, 2 bigger ones that correspond to the dark matter component and two smaller ones that correspond to luminous matter formed from the intercluster gas. These presents a strong evidence of dark matter existence.

Dark matter particles interact among them and with baryonic matter through gravitational interaction. This leads to dark matter potential wells where baryonic matter can also fall. It occurs around $z \sim 3500$. But, dark matter initial seeds for potential wells are formed before decoupling, since they do not interact with radiation, they can gather forming early density perturbations. In this epoch, radiation and baryonic matter are modelled as a fluid that interacts gravitationally with dark matter, leading to no formation of baryonic seeds pre-

cisely because of the coupling between this two density components, i.e., radiation dissipates the baryonic seeds. Later a more detailed explanation will be provided.

In the linear regime an ansatz for perturbations is $\delta = \delta_o D(z)$, where δ_o is an initial seed from where a dark matter potential well formed and $D(z)$ is a growing function of an initial seed. Then, assuming isotropy and neglecting the velocity term since perturbations of our interest satisfy $\lambda_J << \lambda_{pert}$, equation 2.14 can be written as

$$\frac{d^2 D(z)}{dz^2} + \left[\frac{H'(z)}{H(z)} - \frac{1}{1+z}\right]\frac{dD}{dz} = \frac{1}{H^2(z)}\left[\frac{4\pi G \overline{\rho}(z)}{(1+z)^2}\right] D(z)$$
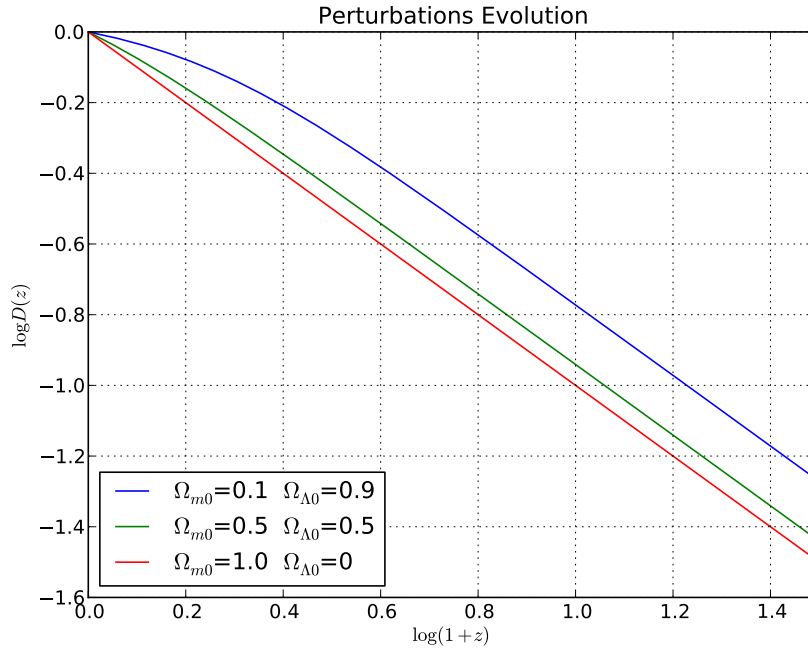

Figure 2.4: Perturbation evolution for a model mass-vacuum, the behavior for different contributions of each component. The largest redshift is $z = 0$.

In the last equation, different solutions are obtained by changing the density parameters. For example, an Einstein de Sitter Universe $\delta_k(z) = \delta_o(1+z)^{-1}$, an Universe dominated by radiation $\delta_k(z) = \delta_o(1+z)^{-1.22}$ and an Universe dominated by vacuum $\delta_k(z) = \delta_o(1+z)^{-0.58}$. In figure 2.4 we show the evolution of perturbations for models with different matter and vacuum contributions. As expected, for bigger matter density perturbations, the growth is faster while for universes with bigger vacuum contribution, perturbations grow slower because of the accelerated expansion induced by Λ. In the latter, a bigger initial mass is required to start the gravitational collapse.

2.6   Statistical properties of cosmological perturbations

To study the evolution of the Universe the equation 2.13 can be used to know the density field, its evolution with time. But using the concept of density contrast $\delta = (\rho - \overline{\rho})/\overline{\rho}$, leads to a density perturbation field, which is a clearer way to analyse the evolution of the density field.

In linear regime, there are a infinite amount of perturbations described by different fourier modes. Since they evolve independently their amplitudes change with time and can be modelled using a transfer function $T(k)$ and a linear growth rate $D(t)$. To characterize such amount of density field values, statistical properties must be defined. This is, not considering individual positions or properties but instead moments defined from some distribution function. This idea is supported by the fact that there is no access to the primordial perturbations that originated the large scale structure observed nowadays. Hence, our Universe could be considered as a realization of a random process where a statistical treatment results as a natural way to study it.

Consider that the Universe or certain region of this with volume V_u, can be divided into small cells, each of them with position x_i. These can be characterized statistically with a joint probability distribution and its moments that would describe the cosmic density perturbation field. Following this approach, the probability of having a mode between δ_k and $\delta_k + d\delta_k$ is,

$$\mathcal{P}(\delta_{\boldsymbol{\kappa}}) r_{\boldsymbol{\kappa}} dr_{\boldsymbol{\kappa}} d\phi_{\boldsymbol{\kappa}} = \exp\left[-\frac{r_{\boldsymbol{\kappa}}^2}{2V_u^{-1}P(\kappa)}\right] \frac{r_{\boldsymbol{\kappa}}}{V_u^{-1}} \frac{dr_{\boldsymbol{\kappa}}}{P(\kappa)} \frac{d\phi_{\boldsymbol{\kappa}}}{2\pi} \tag{2.15}$$

where $r_{\boldsymbol{\kappa}}$ corresponds to perturbations amplitude, $\phi_{\boldsymbol{\kappa}}$ is the phase and varies between $[0, 2\pi)$, i.e., $\delta_{\boldsymbol{\kappa}} = |\delta_{\boldsymbol{\kappa}}| \exp^{-i\phi_{\boldsymbol{\kappa}}} = r_{\boldsymbol{\kappa}} \exp^{-i\phi_{\boldsymbol{\kappa}}}$. The joint probability distribution function is useful because it allows the independence of the terms $\delta_{\boldsymbol{\kappa}}$, or in other words it is the product of every mode's probability

$$\mathcal{P}_{\boldsymbol{\kappa}}(\delta_{\boldsymbol{\kappa}1}, ..., \delta_{\boldsymbol{\kappa}N}) = \prod_{\boldsymbol{\kappa}} \mathcal{P}_{\boldsymbol{\kappa}}(\delta_{\boldsymbol{\kappa}})$$

This expression is not satisfied when the inverse fourier transform is done, since the probability density is not separable in the initial coordinate space. In the Fourier space, the term $P(\kappa)$ can be defined as the power spectrum and is related to the 2 point correlation function as

$$P(k) = \frac{4\pi}{V_u} \int_0^\infty \xi(r) \frac{sin(kr)}{kr} r^2 dr = \langle |\delta(\boldsymbol{\kappa})|^2 \rangle \tag{2.16}$$

from the latter expression can be seen that the isotropy of the Universe is taken into account since during the power spectrum calculation, an average is done over all possible orientations of the vector $\boldsymbol{\kappa}$.

The correlation function, $\xi(r)$, describes the distribution of points, the excess probability of finding a particle at a distance r from another particle selected at random compared against a random distribution. The two point correlation is defined as

$$\xi(r) = \langle \delta(\mathbf{x})\delta(\mathbf{x} + \mathbf{r}) \rangle \tag{2.17}$$

hence ξ only depends on the amplitude of r, it is due to the assumption of homogeneity and isotropy, i.e., depends on relative distances. Furthermore, from equations 2.17 and 2.16 it can be seen that the density field leads to a direct way to find the power spectrum using the Fourier transform.

In the correlation function, no clustering would imply that $\xi(r)$ is zero. A natural way to see this is through a conditional probability, given that there is a particle in a volume element dV_1 the probability there is other one in a volume element dV_2 at a distance r

$$dP(2|1) = n[1 + \xi(x_{12})]dV_2$$

if $\xi(x_{12}) > 0$ the probability of finding such pair of particles increases, i.e., there is clustering of structures. But if $\xi(x_{12}) < 0$ such probability diminishes leading to an anticorrelation. In the case where $\xi(x_{12}) = 0$ there would be no clustering, the distribution of particles would be the one of a random catalogue.

Also, the function $\xi(r)$ can be expressed as a power law of the form

$$\xi(r) = \left(\frac{r}{r_0}\right)^{-\Gamma} \tag{2.18}$$

being valid in the range $100h^{-1}$ kpc to $10h^{-1}$ Mpc. The preferred scale $r_0 = 5h^{-1}$ Mpc is the one where the galaxy density is greater than twice of the background. The exponent value is $\Gamma = 1.8$. This fit overestimates the correlation function for distances bigger than $20h^{-1}$ Mpc.

So far, it has been shown two statistical measures, a fourier pair, in real space the correlation function and in fourier space the power spectrum. But other moments can be specified as previously mentioned, in general an l point correlation function can be defined through the next expression $\xi^l(\vec{x}_1, \vec{x}_2, \dots \vec{x}_l) \equiv \langle \delta_1 \delta_2 \dots \delta_l \rangle$ where the connected terms are the ones

that contributed to the calculation. For example, the first moment of the distribution is $\langle \delta(x) \rangle = 0$ because of the definition of density perturbation field.

An important remark is that if initial density pertubations follow a gaussian distribution all moments higher than two (2 point correlation function) are zero, i.e., the density perturbation field is completely described by the two first moments of the distribution.
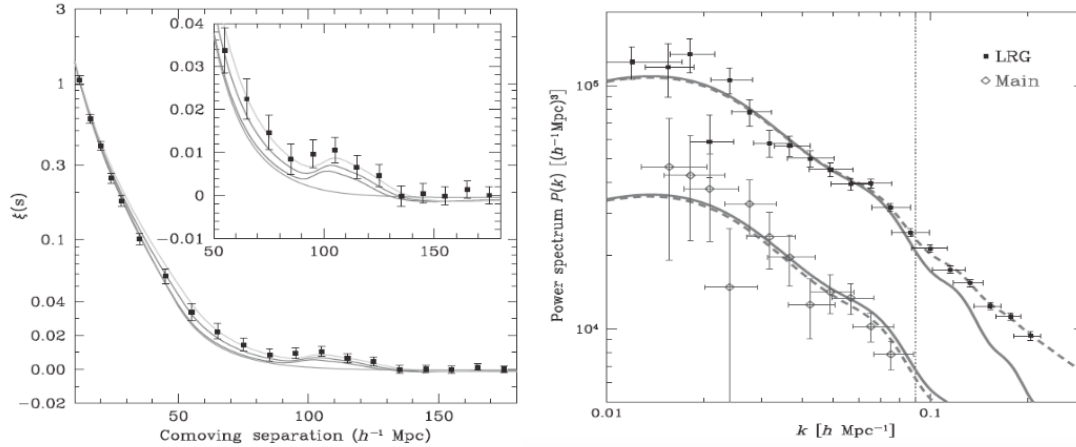

Figure 2.5: BAO peak in the correlation function in the left and the oscillations of BAO in the power spectrum in the right, ([2]). The lower curve is the main SDSS sample and the upper one is the LGR sample.

It is usually considered for the initial density field, that density contrasts follow a normal distribution centered in $\langle \delta \rangle = 0$. This idea is supported by inflationary scenarios where a random gaussian perturbation field arises naturally from quantum perturbations during inflation, i.e., statistical behaviour lies on quantum perturbations. Since there are a large number of modes, the central limit theory would support this idea if the mode phases are independent. One of the advantages of this model is that the perturbation field remains gaussian during linear evolution.

Additionally it has been found that the initial power spectrum expected from inflation theories has the form $P(\kappa) = k^n$. If $n = 1$ the power spectrum is called Harrison-Zeldovich which is commonly used.

With the initial power spectrum it can be done an inverse fourier transform and this way recreating the initial density field.

Furthermore, using inflationary models the shape of the linear power spectrum is well

determined but there are not amplitude predictions, i.e., there is not a defined normalization of the power spectrum. One commonly way to do such thing is through the variance of the galaxy distribution when sampled with randomly placed spheres at radius R. The relation between the variance of the density field and the power spectrum is

$$\sigma^2(R) = \frac{1}{2\pi^2} \int P(k)\hat{\omega}_R(k)^2 k^2 dk$$

where $\hat{\omega}_R(k)$ is the Fourier transform of the spherical top hot model.

$$\hat{\omega}_R(k) = \frac{3}{(kR)^2}[\sin(kR) - kR\cos(kR)]$$

In this approach $\sigma(R)$ is taken around one when $R = 8h^{-1}Mpc$ because of measures performed observationally. Then, normalizing the power spectrum would imply to force $\sigma(R)$ to be one for the mentioned distance.

But several problems arise, one is that this normalization is not precisely valid for linear regime since $\sigma(R) \approx 1$ when $\delta(R) \ll 1$. Other one is baryonic matter is probably a bias tracer of the mass distribution.

It is necessary to consider that after recombination epoch, density perturbations started growing in size causing a non linear growth to appear. This implies a change in the density field and likewise the power spectrum.

2.7 Baryonic acoustic oscillations

Let us consider the epoch before recombination. The baryonic plasma (ionized protons and electrons) was coupled with radiation via Thomson scattering, i.e., the electric field of photons accelerate charged particles making small density perturbations to disperse. Nevertheless, considering that dark matter do not interact with radiation, small dark matter density contrasts can form. Hence, the baryons are subject to two competing forces, radiation pressure and gravitation. Consider a particular dark matter density contrast that attract nearby baryons, they start clustering around the dark matter forming a bigger density contrast. But, due to the pressure caused by coupling, the outward force becomes bigger than gravity, making baryons to move outward as a sound wave. This oscillation of the baryonic plasma is known as baryonic acoustic oscillation.

When decoupling occurs and temperature drops, the force responsible for the expansion of the shell disappear, this is, the pressure caused by the coupling between baryons and radiation, leading baryons in the last position they were located. The scale of the baryonic acoustic oscillation is usually called the sound horizon and it can be computed as

$$s = \int_{z_{rec}}^{\infty} \frac{c_s dz}{H(z)} \tag{2.19}$$

where c_s is the velocity of the propagation and $H(z)$ is the Hubble param, ([8]).

Therefore, there is a spherical shell formed around the dark matter density perturbation. Now, not only dark matter density contrast seed gravitational instability but the baryons in the shell as well.

The structures continue to grow reaching non linear growth and wipe out the imprint lead by the baryonic acoustic oscillations except for the bigger ones. The estimated size of the remaining BAO is 150 Mpc, causing that scale to be more likely to have galaxy formation activity. The reason this distribution is not observed at cosmological scales is because of the amount of imprints, they smear out the preferred scale. Though, it is expected an enhancement in the two point correlation at scales of the baryonic oscillations. As was seen before, there is a direct correspondence between the two point correlation and the power spectrum. Hence, a characteristic oscillation in the power spectrum caused by the imprint of BAO is naturally found.

Since baryonic acoustic oscillations are primarily a linear phenomenon, they are preserved in the power spectrum despite of the temporal evolution. Then, BAO are used as an standard ruler, specifically for high redshift where other rulers tend to fail. This is commonly used for constraining dark matter models.

Although, the nonlinear collapse change the shape and position of baryonic acoustic oscillations, broaden and shift the peak. This is clearly seen in the figure (2.7), the broadening of the peak initially shown as a very sharp causes a damping in the frequency in the power spectrum. It is expected that this effect that is going to be studied in this work, affects baryonic acoustic oscillations on scales around ~ 10Mpc.

The diffusion damping (silk damping) also causes a reduction in size of density inequalities by the diffusion of photons from hot regions to colder ones during the epoch of recombination.
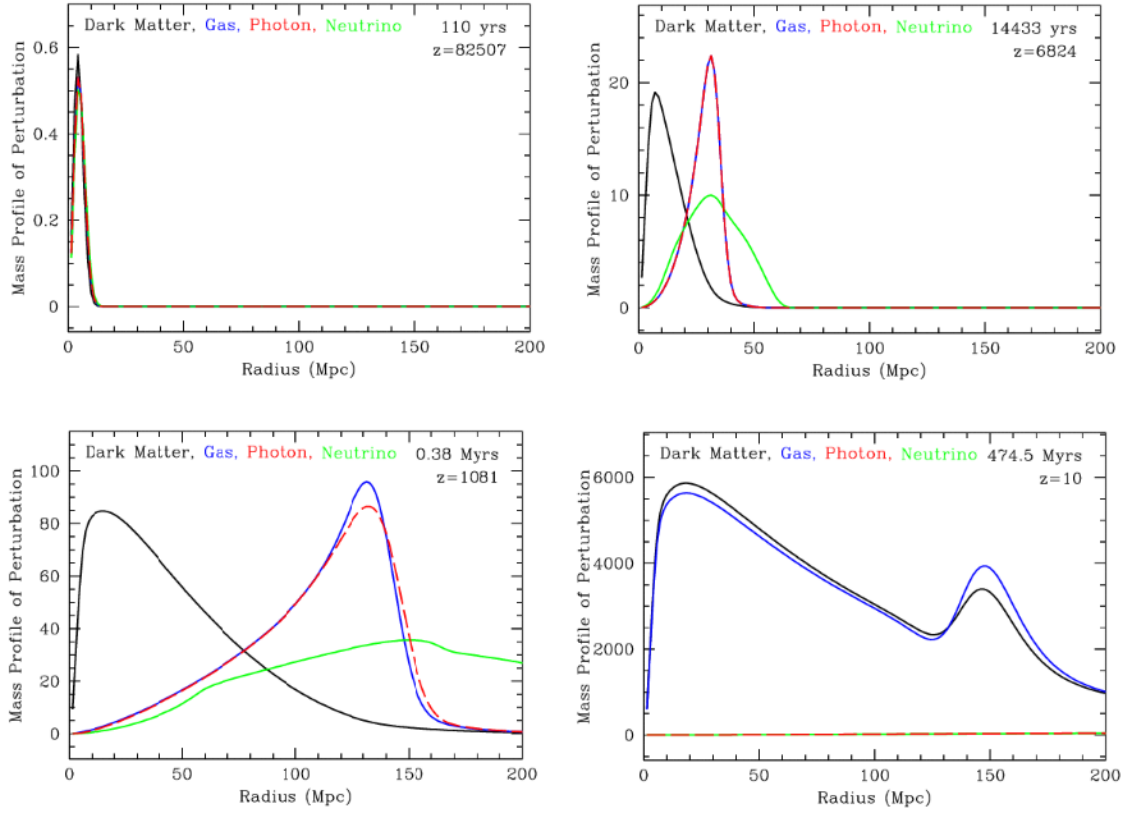
Figure 2.6: All the figures show the evolution of the radial mass profile of dark matter, baryons, photons and neutrinos. First one: Initial perturbations of the four species. Second one: the neutrinos do not interact and move away, the plasma of baryons and radiation overdensity expands because of radiation pressure, the dark matter continues to fall in the perturbation. Third one: The temperature drops enought to lead to decoupling, the baryons slows down until stopped, the radiation and neutrinos continue moving away. Fourth one: The dark matter and baryons eventually get the same distribution because of the gravitational interaction.
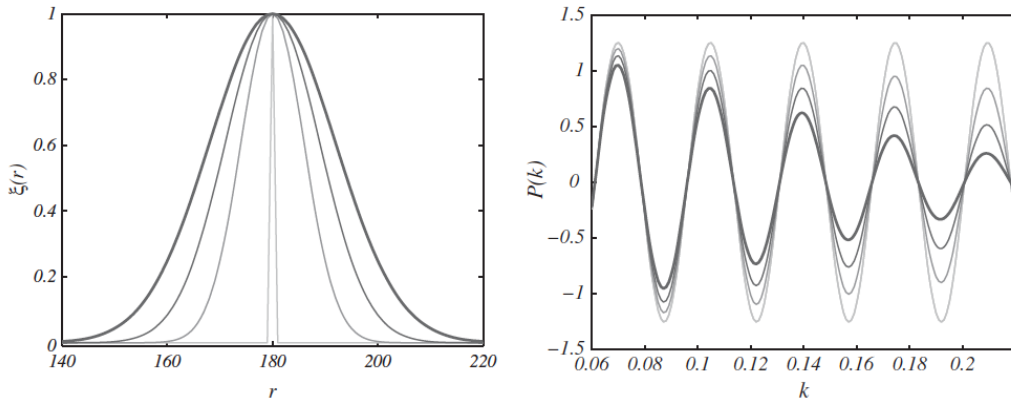

Figure 2.7: In the left the correlation function and the right the power spectrum. When the width of the peak is increased the acoustic oscillation obtained in the power spectrum is damped.

2.7.1   Shift for BAO

During the linear regime the BAO properties are not affected, i.e., the BAO shape does not change during this regime. But, once the non linear regime starts, it causes deviations in the BAO properties. But, let us see in more detail how these deviations affect the BAO properties. In order to do so, it can be studied the velocity field. This gives account for the changes in the mass distribution. Hence, we are going to focus in one model proposed by [9] that models the shift suffered for the BAO signal during the non linear regime.

As mentioned, the correlation function allows to study the clustering of the mass distribution. But a specific mass distribution is affected by the existing velocity field at the same time. Hence, it is necessary to find a way to relate the correlation function with the velocity field. In this scenario, two particles are separated by a distance $\vec{r} = \vec{r_2} - \vec{r_1}$ and thus the pairwise infall velocity, $u_{12}(\vec{r}, \eta)$, is directed along the separation unit vector $\hat{r}$, where η is proportional to the conformal time. However, this distance diminishes with time due to clustering. In this direction, an important equation is the pair conservation equation

$$\frac{\partial \ln[1 + \xi(r, \eta)]}{\partial \eta} - u_{12}(r, \eta)\frac{\partial \ln[1 + \xi(r, \eta)]}{\partial r} = \Theta(r, \eta) \tag{2.20}$$

where $\Theta(r, \eta) = \nabla \cdot [u_{12}(r, \eta)\hat{r}]$, is the divergence of the pairwise infall velocity. It allows to study the evolution of any feature, its movement, in the correlation function. Hence, the relation between the correlation function and the pairwise infall velocity allows to study the shift that BAO would suffer due to gravitational clustering. To solve the equation 2.20 is used the method of characteristics, the characteristics are the equations of motion of the pairs

$$\frac{dr}{d\eta} = -u_{12}(r, \eta) \tag{2.21}$$

whose solutions allow to obtain an ordinary equation from 2.20

$$\frac{d \ln[1 + \xi(r, \eta)]}{d\eta} = \Theta(r, \eta) \tag{2.22}$$

Besides, the solution of the last equation 2.22 solution is given by

$$1 + \xi(r, \eta) = (1 + \xi_0[r_0(r, \eta)]) \times \exp\left[\int_0^\eta \Theta[r_{\eta'}(r, \eta), \eta']d\eta'\right] \tag{2.23}$$

where $r_0(r, \eta)$ is the initial separation that corresponds to r at time η and likewise for $r_{\eta'}$. To find this general solution, the shift in the correlation function due to gravitational

clustering, it is necessary to know the divergence of the velocity field. A first approach that can tell us how the BAO peak evolves is through linear theory of velocities for $\Theta(r, \eta)$. This is shown in the expression below.

$$\Theta(r, \eta) = 2e^{2\eta}\xi_0(r) \tag{2.24}$$

where ξ_0 is the linear correlation function at $\eta_0 = 0$. In such case, the correlation function grows due to the infall velocities. It becomes the biggest contribution when η is large.

Now, using this expression and clearing the initial correlation function is obtained the plot 2.8, the two curves for $z = 0$ and $z = 1$ are represented with solid lines. Also, two curves, solid squares and solid triangles, for $z = 0$ and $z = 1$ respectively, show the behavior of the correlation function in cosmological simulations. From this, it can be said that in non linear regime the BAO has been wash out and it does not suffice the linear theory to produce such effect. This is in accordance with several studies that show a significant differenc between numerical simulations and the predictions of the linear theory. Hence, a next step is to include in 2.23 a divergence of the velocity that includes the nonlinear effects.
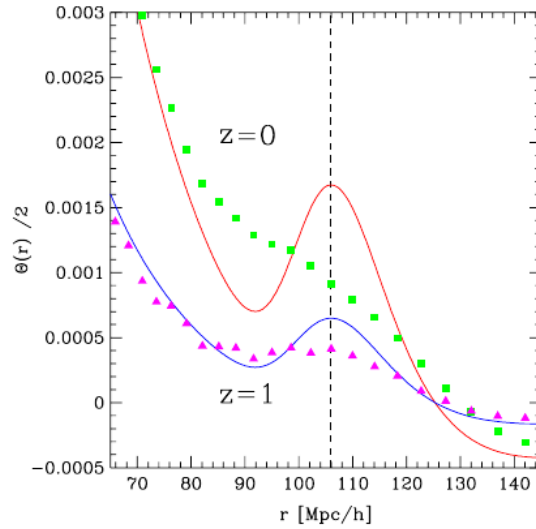

Figure 2.8: Using expression 2.24 for $z = 0$ and $z = 1$ is obtained the solid lines. From cosmological simulations the correlation function for $z = 0$ and $z = 1$ is shown, solid squares and solid triangles. Image taken from [9].

For nonlinear regime, the perturbation theory (PT) can be used to find the pairwise infall velocity. This one has two nonlinear contributions and its divergence can be expressed as

$$\Theta(r, \eta) = \Theta_2(r, \eta) + \Theta_3(r, \eta)$$

where $\Theta_2(r, \eta) = 2\nabla \cdot \langle \delta_1 \vec{u}_2 \rangle$ and $\Theta_3(r, \eta) = 2\nabla \cdot \langle \delta_1 \delta_2 \vec{u}_1 \rangle$. It can be shown that PT is used to find the last two terms as shown in [9]. Finally, the correlation function is expressed as

$$
\begin{aligned}
1 + \xi(r, \eta) \;\; = \;\; & (1 + b_i^2 \xi_0[r_0(r, \eta)]) \times \\
& \exp\left[\int_0^\eta d\eta'(b(\eta')\Theta_2[r_{\eta'}(r, \eta), \eta'] + b(\eta')^2 \Theta_3[r_{\eta'}(r, \eta), \eta']) \right] \quad (2.25)
\end{aligned}
$$

where b_i is the initial bias. This expression predicts the damping of the BAO that causes an apparent shifting to smaller scales. Hence, it can be concluded that indeed nonlinear regime induce changes on BAO properties. This idea is supported by [9], [10], [11], [12], [13], [14], where it is shown that the position of the BAO peak is changed around 1% to 3%. Despite of the fact it is not a big difference, it induce a significant bias in the parameter of the dark energy equation. Then, a model for BAO must include the nonlinear contributions.

Although this model gives account for the shift due to the transport of matter by the velocity field, there is other term that can contribute to the shifting, the impact of tidal gravitational fields in the growth of structure as explained in [14].

Computational methods

Currently in cosmology, one issue is to find the way the large scale structure formation evolves from its first stages to the actual epoch. This is, it is necessary to study the evolution of the matter content of the Universe under the influence of gravity. To solve this system, the Boltzmann equation (BE) could be used since it allows to predict the statistical behavior of a system that is not in equilibrium. Hence, a probability density distribution for the system would be obtained from solving the BE. But this equation cannot be solved for this system so complex, making necessary some other approximation. One possible solution is to assume an initial density field as one specific realization of the probability density distribution, i.e. a specific particle distribution that would trace the initial density field. Now, using this initial particle distribution, the evolution of the system would obtained finding the interaction among particles. Thus, due to the big amount of interacting particles necessary to simulate the system, computational resources appear as a necessary tool to tackle such problems. Hence, numerical approximations must be developed to find for every time step all the properties needed to describe the system, even when particles studied have masses with magnitude order of several stellar masses.

In cosmological simulations of structure formation, a key component to consider is dark matter, because dark matter dominates the gravitational interaction, not only because of its amount compared to baryonic matter but also because it only interacts in this way. Then, it is mostly because of this one that the Universe has a sponge like structure at large scales.

Defining as total density the sum of baryonic and dark matter, a key argument to ignore

baryonic matter in simulations can be given, dark matter would contribute with around 80% of all this density content. Hence, the assumption that the large scale structure formation is determined by dark matter is plausible.

This chapter is divided in several sections, the first one corresponds to the methods used in cosmological simulations to calculate the gravitational evolution of the system. The second one contains different criteria selection to detect a dark matter halo in simulations. The third and fourth sections are dedicated to explain how to build different statistical measures of clustering, the correlation function in real space and the power spectrum in fourier space. Detailed description can be found in [15], [5], [16], [17], [18],[19], [20], [21].

3.1   Numerical methods

To study the large scale structure formation at big scales, simulations of dark matter interacting particles inside a cosmological box are performed. In such cases it is necessary to suppose initial conditions, an initial configuration of the Universe, i.e., an initial density field or an initial shape for the power spectrum. Furthermore, two important parameters for this cosmological simulations are the box size L and the number of dark matter particles N^3 that would trace the initial density field, among others.

The gravitational interaction calculation of such a big number of particles could in principle be calculated through direct sum of forces. This first attempt is not very efficient or even it is not possible to perform since the computing time or the computational resources would be very big to be viable. The latter is the reason for approximate methods to appear as a possible solution that would require more reasonable computing times.

A main objective in a simulation is the study the formation process and further interactions that are produced among halos and vacuum regions that conform the sponge-like structure of the Universe. Next, two numerical methods used for cosmological simulations are going to be briefly exposed.

1. **Particle Mesh (PM)**[16]

   In this method a grid is created over the particle array, i.e., the cosmological box that contains the particles is divided in cells of the same size. But, to show the basics of this method, a 2D example is explained, as shown in the figure 3.1. In this case, the particles closer to a vertex, each corresponding to a specific cell, are assigned to this one. Then, an approximated density field is calculated using the mass per cell divided

by its volume. Other way to calculate the density field is cloud in cell, later explained in more detail, where particles are considered constant density cubes causing that a single particle contributes to different cells. Now, using the density field and the Poisson equation can be calculated the gravitational potential in every grid vertex.

This method reduces considerably the computing time since its of the order of $O(N + M\log M)$ with N being the number of particles and M the number of vertex. Although, the lack of resolution in the regions that are more dense makes this method insufficient to respond for the physical situation. Furthermore, it does not give account for a complex geometry or systems higly correlated. A step forward in this direction is P^3M that uses for smaller scales finer calculations of the potential performing a particle particle calculation.
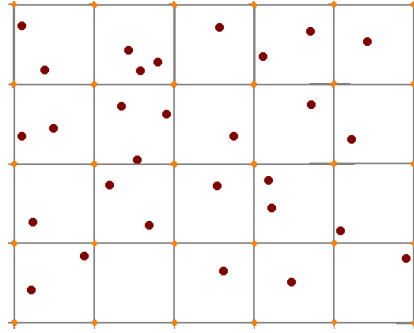

Figure 3.1: Particle mesh method. Every vertex of the grid gets properties calculated from the closer particles.

2. **Tree method**[16]

To illustrate this method let us consider a 2D particle array. In this case, the cosmological box is represented by a square. It is divided in cells of the same area where each particle is assigned to the cell where it falls in. If the number of particles is superior to one, subdivisions of the cell are performed. Again, if the number of particles per subcells is superior to one, subdivisions are made. This process is repeated until for every cell there is at most one particle. This subdivision is used to create a tree structure, this consists in a root, i.e., all the square area and the branches that are created with each subdivision performed. This serves as a map of the disposition of the particles in the square array. The particles are numbered from the upper left of the square until all particles in the first cell are numbered. Then, it is repeated the same process with the second cell and and so forth until the lower left of the square is reached.

When the gravitational calculation is performed, the contribution to the force exerted over a particle due to the more distant ones is much lower than with the nearer ones. Thus, the far ones can be approximated as a pseudoparticle with mass M and with a position $r_{CM} = \sum_i m_i r_i / M$. As a selection criteria the next expression is taken

$$s/q \leq \theta \tag{3.1}$$

where s is the cell size with wich the particle of interest is interacting with, d the distance cell particle and θ is a tolerance value to define. When the condition is satisfied the gravitational interaction is calculated directly with the pseudoparticle. In the opposite case the relation 3.1 for the subcells that contain the studied cell and that way successively until the condition is satisfied or only one particle is present per cell. In this way the direct calculation is avoided for far objects without avoiding the calculation for the nearer ones. Therefore, the computing time is reduced from $O(N^2)$ to $O(N \log(N))$.
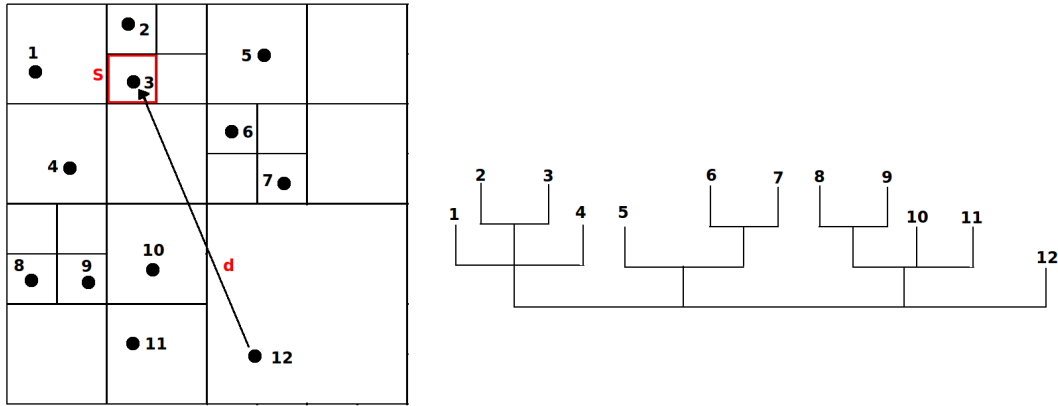

Figure 3.2: In the left panel is shown the array of the particles and the subdivisions performed until at most one particle is found per cell. In the right one, the tree found for such distribution is shown.

There are many methods that are hibrid among the 2 exposed, for example the already mentioned $P^3 M$. All of them have both advantages and disadvantages that need to be evaluated according to the needs of the simulation.

3.2   Halo selection

As a result of the dark matter particle interactions, perturbations grow enough to form bound objects that we will assume are in virial equilibrium, these are known as dark matter halos and satisfy the relation $E_k = -V/2$ where E_k is the kinetic energy and V is the potential energy. They are responsible for the potential wells that causes baryonic matter to fall in, forming finally the galaxies we observe today, i.e., dark matter halos host galaxies.

A main result of a cosmological simulation are the dark matter halos catalogue, which we are going to work with, that contains halo properties such position, velocity, mass and radius. Therefore, a key step is to identify halos from a cosmological simulation, the methods generally use to acomplish such task are FOF and BDM[1]

1. **Friend of Friends (FOF)**

   To identify if a particle group lies in a dark matter halo, i.e., particles are bound, a length is defined such that all the particles with their distances lower than this length are part of the same group. This treshold is called linking length. A condition is imposed, groups can not intersect among them, hence a particle can only belong to a specific group.
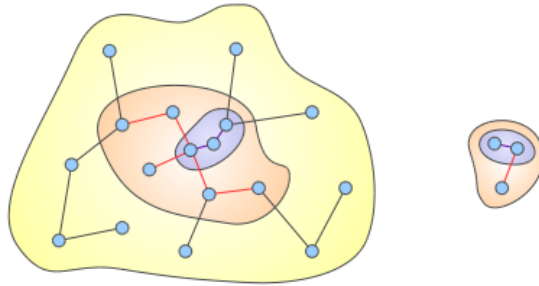

   Figure 3.3: Image taken from Cosmosim database.

   But there is a problem with this approach, even when there is a little amount of particles in common between two groups, some sort of small "bridge" that unites both of them, they are selected as one group not two as would be expected. This method also allows to define substructures, therefore using different linking lengths, groups inside groups would be obtained, the bigger ones would host the smaller ones.

2. **Bound Density Maximun(BDM)**

---

[1]`https://www.cosmosim.org/cms/simulations/halo-finders/`

For this method the local maximum densities in the particle array of the simulation are detected. From them, a spherical cut is defined and the particles inside form the dark matter halo. Particles with bigger or equal velocity than the scape one are not included in the halo. Contrary to FOF method, halos can overlap while the center of mass of one halo does not fall into the other one. Nevertheless, if the center of mass of one halo falls in the virial radius of other one, the first one is considered a subhalo of the last one. The standard overdensity limit of the halos is $360\rho_{back}$ where ρ_{back} is the background density.

3.3   Density field in a cosmological simulation

To construct a good approximation of the real density field from a cosmological simulation, a sampling of the continuous density field in a regular grid of size N^3 is performed, the subdivisions created are called cells. Hence, an assignment of the particle charge, i.e. particle mass, to the grid must be done. To obtain a more realistic density field approximation the grid points can be increased also disminishing problems due to numerical effects but it is more expensive computationally. Furthermore, the number of particles in a simulation is a restriction to the maximum value that N can have, it can not exceed $\sqrt[3]{N_p}$, there would not be enough particles to map correctly the density field per cell. A size grid around the value mentioned would optimal in the sense that the particle mean per cell would be one, hence a Poisson distribution would be followed. But the sampling made from the particle distribution is not a mere sampling but a sampling convolved with a window function (the way a particle mass is distributed in the grid), i.e., the window function W that is used affects the density field calculated.

Since the particles are located in a specific position it can be assured that the particle number density is

$$n_0(\mathbf{x}) = \sum_{i=1}^{N_p} \delta^D(\mathbf{x} - \mathbf{x}_i)$$

where $\mathbf{x}_i$ the position of the i-th particle. The window function quantifies how much of the particle number density is distributed to a grid point separated by $\mathbf{x}$, hence the sample particle number density can expressed as

$$n(\mathbf{x}_p) = \int_V d^3x' n_0(\mathbf{x'}) W(\mathbf{x}_p - \mathbf{x'})$$

Similarly, the sampled density contrast defined as $\delta^s(\mathbf{x}) = n(\mathbf{x}_p)/\bar{n} - 1$ can be found using the convolution of the real density contrast and the window function

$$\delta^s(\mathbf{x}) = [\delta * W](\mathbf{x}) \tag{3.2}$$

its fourier transformation is simply the product of the fourier transformation of the real density contrast and the window function

$$\delta^s(\mathbf{k}) = \delta(\mathbf{k})W(\mathbf{k}) \tag{3.3}$$

thus, the real density contrast can be obtained dividing the sampled density contrast with the window function used.

The procedure of convolving with a window function can be seen in a different way, if a point spreading or cloud shape function $S(x')$, being x' the distance from the particle position x_i, is carried by each particle then the charge assigned to the grid point x_p is given by the overlap of the shape function within the cubic cell p

$$W(x) = \int \Pi \left(\frac{x'}{H} \right) S(x' - x) dx'$$

where $\Pi(x)$ is the top hot function and $H = L/N$ is the size of a cell.

There are 3 commonly used schemes for the mass assignment, nearest grid point, cloud in cell and triangular shaped cloud. For each case we are going to consider a one dimensional window function. The sencond and third one are first and second order distribution schemes respectively, hence each of them is a better approximation than the previous one.

1. **Nearest grid point:** The first scheme considers that the particle charge is assigned to the cell where the particle falls, each cell is centered in a grid point, therefore the particle is assigned to the nearest grid point. Let us see this in more detail. If the cloud shape interpretation is used, the particle shape would be a Dirac delta function that would be assigned to the specific cell where particle falls in as shown in the figure 3.4 (a). If the other interpretation is considered, the window function would be a top hat function centered in the particle, the value assigned to grid point would the one that top hat function would get when is evaluated in that grid point as shown in figure 3.4 (b).

$$W_{NGP}(x) = \Pi\left(\frac{x}{H}\right) \quad \equiv \quad \frac{1}{H}\Pi\left(\frac{x}{H}\right) * \delta\left(\frac{x}{H}\right)$$
$$= \quad \frac{1}{H}\Pi\left(\frac{x}{H}\right) * S\left(\frac{x}{H}\right)$$

This window function in the fourier space is

$$W_{NGP}(k) = sinc\left(\frac{\pi k}{2k_N}\right)$$

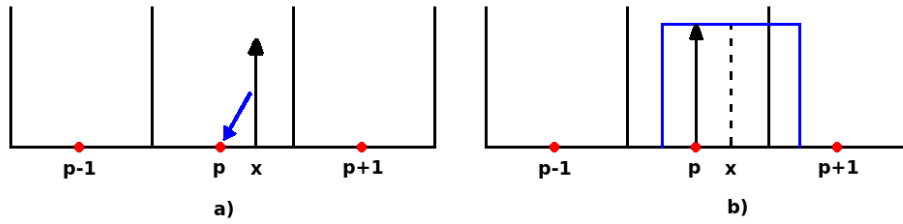where k_N is the Nyquist frequency that later will be defined.

Figure 3.4: The left figure shows the cloud shape interpretation where the Dirac function is assigned to the particle grid. The right one shows the window function interpretation where the top hat function evaluated in the grid point would give the charge assigned to it.

2. **Cloud in cell:** This scheme assumes that the charge of a specific particle assigned to a grid point is given by the overlap of a cell with a size H centered in the particle with the cell centered any the grid point. Then, the particle not only contributes to the cell where it falls in but also to some of the 26 neighbour cells. This explanation is shown in the left figure of 3.5 but according to the window function explanation, a "triangle" function $\Lambda(x)$ centered in the particle and length H is evaluated in the corresponding grid points of the cells, the one where particle falls in and the neighbour ones, finding the contribution of the charge to every one of them as shown in figure 3.5 (b).

$$W_{CIC}(x) = \Lambda\left(\frac{x}{H}\right) \quad \equiv \quad \frac{1}{H}\Pi\left(\frac{x}{H}\right) * \Pi\left(\frac{x}{H}\right)$$
$$= \quad \frac{1}{H}\Pi\left(\frac{x}{H}\right) * S\left(\frac{x}{H}\right)$$

This window function in the Fourier space is

$$W_{CIC}(k) = sinc^2\left(\frac{\pi k}{2k_N}\right)$$

hence, the Fourier transform of the CIC window function is the square of the Fourier transform of the NGP window function.
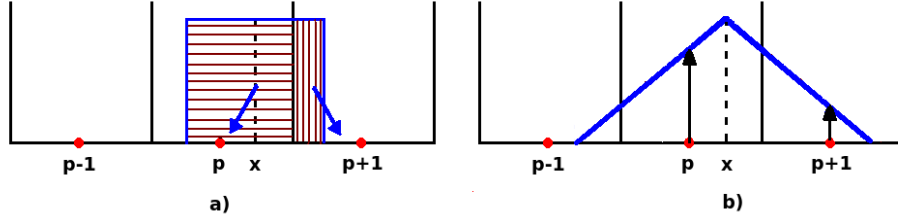

Figure 3.5: The left panel shows the CIC cloud shape function and the intersection between the cell centered in the particle with the cells provides the contribution of the charge to every cell. The right one shows a "triangle" function that is evaluated in every grid point to find the charge contribution to the cell.

3. **Triangular shaped cloud:** This scheme is as the two previously presented but the cloud shape and window function change. As it happens with CIC, TSC contributes to different cells, not only the one where it falls in. Both interpretations are shown in the figure 3.6. Next the expression to calculate the charge contribution to a specific cell is given by

$$
\begin{aligned}
W_{TSC}(x) &= \frac{1}{H}\Lambda\left(\frac{x}{H}\right) * \Pi\left(\frac{x}{H}\right) \\
&= \frac{1}{H}\Pi\left(\frac{x}{H}\right) * S\left(\frac{x}{H}\right)
\end{aligned}
$$

The Fourier transform of the window function is

$$
W_{TSC}(k) = sinc^3\left(\frac{\pi k}{2k_N}\right)
$$

hence, the Fourier transform of the TSC window function is the cubic of the Fourier transform of the NGP window function.

Hence, each successively higher order assignment function is obtained by convolving the previous assignment function with $\frac{1}{H}\Pi\left(\frac{x}{H}\right)$.

From a one dimensional window function can be obtained the three dimensional one, simply as the multiplication of the three one dimensional ones. This last asseveration is valid due to the grid used is regular.
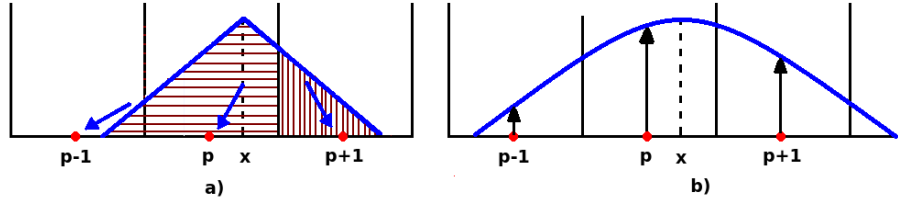
Figure 3.6: The left figure shows the cloud shape function, a "triangle" function and the overlap with the cells provides the value of the charge assigned to every cell. But using the window function interpretation the right figure is obtained, where the particle carries with a function that evaluated in every grid point gives the contribution to the specific cell.

Thus, for every cell contained in the cosmological box a value of the convolved density field is calculated using a specific mass assignment scheme.

3.4   Power spectrum in cosmological simulations

The density perturbations of the ensemble, convolved density field for every cell of the box, allow to calculate the power spectrum as shown in equation 2.16, an esemble average for every mode κ. Since this a statistical measure in the Fourier space let us see the Fourier transform in more detail.

3.4.1   Fourier transform

The fourier transform is defined for this work with the next convention

$$F(\boldsymbol{\kappa}) = \int_{-\infty}^{\infty} d^3x f(\mathbf{x}) e^{-i\boldsymbol{\kappa}\cdot\boldsymbol{x}} \tag{3.4}$$

with κ being the wave number vector. The inverse fourier transform as

$$f(\boldsymbol{x}) = \int_{-\infty}^{\infty} \frac{d^3\kappa}{(2\pi)^3} F(\boldsymbol{\kappa}) e^{i\boldsymbol{\kappa}\cdot\mathbf{x}} \tag{3.5}$$

The convolution of the functions $g(\mathbf{x})$ and $f(\mathbf{x})$ is defined as follows

$$h(\mathbf{x}) = |\mathbf{g} * \mathbf{f}|(\mathbf{x}) \equiv \int_{-\infty}^{\infty} \mathbf{g}(\mathbf{x}')\mathbf{f}(\mathbf{x} - \mathbf{x}')\mathbf{d^3x}'$$

but the Fourier transform of $h(\mathbf{x})$ is

$$H(\boldsymbol{\kappa}) = G(\boldsymbol{\kappa})F(\boldsymbol{\kappa})$$

this is known as the convolution theorem. It is used in our work since the fourier transform of the window function convolved with the density field (equation 3.2) allow us to obtain the real density field. From equation 3.3

$$\delta(\mathbf{k}) = \delta^s(\mathbf{k})/W(\mathbf{k})$$

In the situation we are dealing with, the function $F(\boldsymbol{\kappa})$ is only sampled at evenly spaced intervals (N^3 frequencies totally) since we only know $f(\mathbf{x})$ in N^3 points

$$F(\boldsymbol{\kappa}) = \begin{cases} F(\kappa_F \boldsymbol{n}_\kappa) & \boldsymbol{n}_\kappa = (i,j,k) \in Z^3 \\ 0 & \text{otherwise} \end{cases}$$

where $\kappa_F = 2\pi/L$ is the fundamental frequency.

Due to the functions are only sampled in specific points, the integrals defined in equations 3.4 and 3.5 can be approximated to the discrete Fourier transform. Let us express them in terms of the density fluctuations in real and Fourier space because they are the ones of interest for our work

$$\delta(\boldsymbol{\kappa}_p) = H^3 \sum_{n_p} \delta(\boldsymbol{r}_p) e^{-i\boldsymbol{\kappa}_p \cdot \boldsymbol{x}_p}$$

$$\delta(\boldsymbol{r}_p) = \frac{1}{L^3} \sum_{\boldsymbol{k}_p} \delta(\boldsymbol{\kappa}_p) e^{i\boldsymbol{\kappa}_p \cdot \boldsymbol{x_p}}$$

where $H = L/N$ is the separation of the grid in the real space, $\boldsymbol{\kappa}_p = k_F \mathbf{n}_p$ and $\boldsymbol{n}_p = (i,j,k)$ with each index varying from $-N/2 \leq i,j,k \leq N/2$. The function $\delta(\boldsymbol{r}_p)$ is sampled in the points $\boldsymbol{r}_p$ and $\delta(\boldsymbol{\kappa}_p)$ in the points $\boldsymbol{\kappa}_p$. Therefore, the Fourier space is divided into small cells, N cubes of size $\kappa_g = 2\pi/H$ per dimension as it was done for the simulation.

Furthermore, the extreme values for $\boldsymbol{n}_p$ correspond to Nyquist critical frequency

$$\kappa_N = \pi \frac{N}{L} = \frac{\pi}{H}$$

so $-\kappa_N < k < \kappa_N$. A phenomenon called aliasing appears when a continuous function is sampled and is not bandwidth limited to a frequency smaller than κ_N. It consists in a folding over or aliasing of the frequencies that fall outside the range as shown in figure 3.7.

To perform the discrete Fourier transform of the sampled density field was used the free library FFTW, where a fast Fourier discrete transformation (FFT) is implemented. Because of algorithmic details of the FFT the fourier coefficients are ordered in the following manner
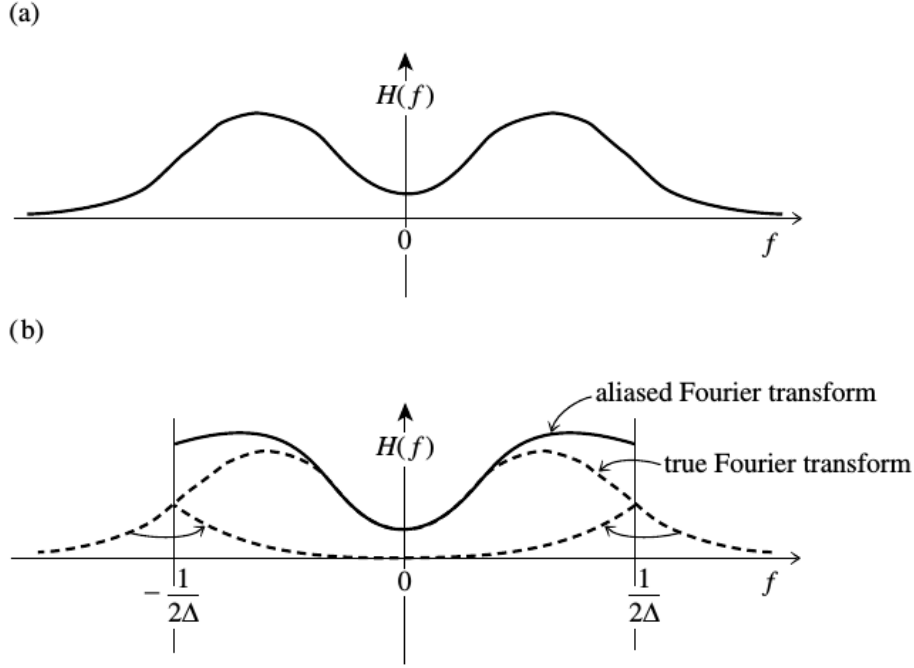
(a)

(b)

Figure 3.7: Aliasing effect for $H(f)$ sampled with a space interval Δ. Frequencies outside the frequency range are included into the range because of the discrete sampling of the function. Figure taken from [22].

$$\kappa_l(i) = \begin{cases} \frac{2\pi}{L}i & \text{if} \quad i = 0, \ldots, \frac{N}{2} \\ \frac{2\pi}{L}(-N+i) & \text{if} \quad i = \frac{N}{2}+1, \ldots, N-1 \end{cases}$$

where the subindex l stands for x, y or z coordinate. The library has different routines, one is a complex to complex function, this is, perfoms a Fourier transformation of a sampled complex function. Other one, named real to complex routine, takes the real samples of a function to find the Fourier transformation. The last one uses the Hermitian condition that allows to improve the calculation in speed and memory usage

$$\delta_\kappa(-\mathbf{n}_\kappa) = \delta_\kappa^*(\mathbf{n}_\kappa)$$

where the superscript $*$ denotes complex conjugate. In both cases a normalization must be taken into consideration, this can be noticed in the relation between the transformed density field obtained with FFTW and the sampled space density field

$$\delta^{FFTW}(\mathbf{n}_k) = \sum_{r_p} \delta(\mathbf{r}_p)e^{-i\boldsymbol{\kappa}_p \cdot \boldsymbol{r}_p} = \frac{\delta(\boldsymbol{\kappa}_p)}{H^3}$$

with the last expression and the definition of PS given in equation 2.16, the power spectrum from FFTW is given by [15]

$$P(\kappa_F n_1) = \frac{H^6 k_F^3}{(2\pi)^3} \langle \delta^{FFTW}(\boldsymbol{n}_1 \delta^{FFTW}(-\boldsymbol{n}_1)) \rangle = \frac{V}{N^6} \langle |\delta^{FFTW}(\boldsymbol{n}_1)|^2 \rangle \qquad (3.6)$$

this is the power spectrum estimator that is used throughout this work.

3.4.2   PS calculation

To calculate the power spectrum, the next steps are followed

1) From a cosmological box of size L, a grid with N^3 subdivisions is performed creating cells of volume H^3.

2) The sampled space density field is created using a specific window function, mass of particles are assigned to the grid.

3) With FFTW software, the FT of the sampled space density field is calculated.

4) To deconvolve and eliminate the aliasing effect $P(\kappa) \equiv |\delta(\kappa)|^2$ is divided by the next window function, at each grid point or equivalently each cell

$$W(\boldsymbol{\kappa}) = \prod_{i=1}^{3} \left[1 - \frac{2}{3} \sin^2 \left(\frac{\pi \kappa_i}{2 \kappa_N} \right) \right]$$

where $\boldsymbol{\kappa} = (\kappa_x, \kappa_y, \kappa_z)$ as proposed in [23].

5) The amount $P(\kappa)$ is calculated taking the spherical average of $P(\boldsymbol{\kappa})$ corrected inside the shell $\kappa - \Delta\kappa/2 < |\boldsymbol{\kappa}| < \kappa + \Delta\kappa/2$.

3.5   Correlation functions in cosmological simulations

In practice, to calculate in a cosmological simulation the correlation function a at distance r, it has to be performed an average of the number of neighbours per particle at a given scale or the binned comoving separation. In this direction correlation function estimators can be used, one of the most basic ones is shown below. Two catalogues are considered for this estimator, one is the properties of the particles of the box, data-data catalogue (DD) and the second one is generated randomly with at least the same number of particles and the same size of the box, random-random catalogue (RR). The DD catalogue should have

regions with more or less clustering than a homogeneous distribution, this is precisely the RR catalogue role, a way to measure how much the DD catalogue deviates from the homogenous distribution.

From this estimator is easier to notice that $\xi(r)$ is a measure of excess or deficiency of clustering at r making it more intuitive

$$1 + \xi(r) = \frac{\eta_{DD}(r)}{\eta_{RR}(r)}$$

here $\eta_{RR}(r)$ is the number of pairs of particles at a distance r in the catalogue DD and $\eta_{RR}(r)$ is the number of pairs of particles at a distance r in the catalogue RR.

Other common estimators also need an additional catalogue, the data-random, where the pair of particles would not only include the DD or RR but a mixed catalogue containing both arrays, making more robust the estimator proposed [24]

$$\text{Landy-Szalay Estimator} \quad \xi_{LS}(r) \;=\; 1 + \frac{DD(r)}{RR(r)}\left(\frac{N_R}{N}\right)^2 - 2\frac{DR(r)}{RR(r)}\left(\frac{N_R}{N}\right)$$

$$\text{Hamilton Estimator} \quad \xi_{HAM}(r) \;=\; \frac{DD(r)RR(r)}{DR(r)^2} - 1$$

where N_R is the number of points of the RR catalogue and N of the DD catalogue.

3.5.1 Correlation function calculation

To calculate the correlation function the Landy-Szalay estimator is used, the next rough steps are followed to calculate it

1) A catalogue of N_R particles is generated randomly in a cubic box of size L.

2) For a bin around r the magnitude of the distances between DD, RR and DR particles are found, that way, pairs of galaxies that are at a distance r for every catalogue are obtained. Finally, using Landy-Szalay estimator the correlation function is calculated for each radial bin.

Results

During this chapter the questions initially stated are going to be studied in detail, looking forward to answer them. Though there are several questions, the central one consists in finding if there is any difference in BAO properties when the scale of the tracer halo population is changed. A possible way to account for this, it is to study the clustering at BAO scales for such halo populations. That is precisely what is proposed in this work. One statistical tool that will be used to study clustering were already explained in chapter 3, the correlation function will provide our main results.

Now, using this tool, the simulation Multidark Planck (MDPL) will be studied. It belongs to the MultiDark database in which Planck parameters were used (table 2.1). The characteristics of MDPL simulation are a box length of $L = 1\text{Gpc/h}$, a number of particles equal to 3840^3 and a mass resolution of $1.51 \times 10^9 M_\odot/h$. Furthermore, the MDPL simulation has available a halo catalogue constructed with Friend of Friends algorithm and a linking length of 0.2 [1]. From this halo catalog are constructed several populations using the mass of the halos to classify them. This is more clear from the table 4.1 where each mass range for the 4 different populations constructed are shown. They are going to be called the thick bins.

But, the populations have a range mass that is arbitrary, so the results obtained can have some sort of bias. To avoid this, it will be studied in further detail the effect that the size of the mass bins have in the results, i.e., the mass range considered. The bin size could be

---

[1] Data taken from `https://www.cosmosim.org/cms/simulations/MDPL/`

MDPL population	Mass range $M_\odot/h$	Number of halos
4	$M \geqslant 1 \times 10^{14}$	32436
3	$1 \times 10^{13} \leq M < 1 \times 10^{14}$	443356
2	$1 \times 10^{12} \leq M < 1 \times 10^{13}$	3687677
1	$1 \times 10^{11} \leq M < 1 \times 10^{12}$	32868688

Table 4.1:   Populations constructed from the mass range that appears in column two (z=0).

"masking" information contained for smaller mass ranges.

Hence, there are other subsamples created from every thick bin. Four populations are constructed for the population 1, each of them has the same number of halos. They are labelled as q_i with $i = 1, \ldots, 4$, q_1 is the extreme nearer to population 3 and q_4 is the extreme nearer to population 1. Likewise, for the population 2 other four populations are built where the same labelling is used. They will be called quarter populations or thin populations. Their mean masses in logarithmic scale are: quartiles for population 1: $10^{11.77}$, $10^{11.39}$, $10^{11.19}$ and $10^{11.06}$ M_{odot}/h, quartiles for population 2: $10^{12.75}$, $10^{12.39}$, $10^{12.19}$ and $10^{12.06}$ M_{odot}/h. There was another reason to construct such populations, the correlation function found for the population 2 has an irregular shape around values of BAO peak.

Concluding, all of these subdivisions are created to study the effect of the scale of the populations on the BAO signal. This is found to be related with the nonlinear gravitational effects, i.e. for smaller mass halos there is a coupling among different density modes, that is not so strong for larger scales.

This lead us to other experiment, to revise if there is a difference in the BAO properties found for the same cosmological box but for $z = 1$, where nonlinear gravitational effects should be less prominent. In the table 4.2, we show the halo populations used for $z = 1$. The same procedure that will be exposed for the populations with $z = 0$ is followed for $z = 1$, but we are going to concentrate in the first redshift where a deeper study was performed.

4.1   Correlation functions for MDPL populations

To compute the correlation function was implemented a parallel C code that uses MPI. There are several important quantities in order to run the code and obtain the correlation function for a specific population, the random sample factor N_r, the minimum radial value R_{min}, the maximum radial value R_{max}, the radial number of bins N_{bins}, and the number of particles of
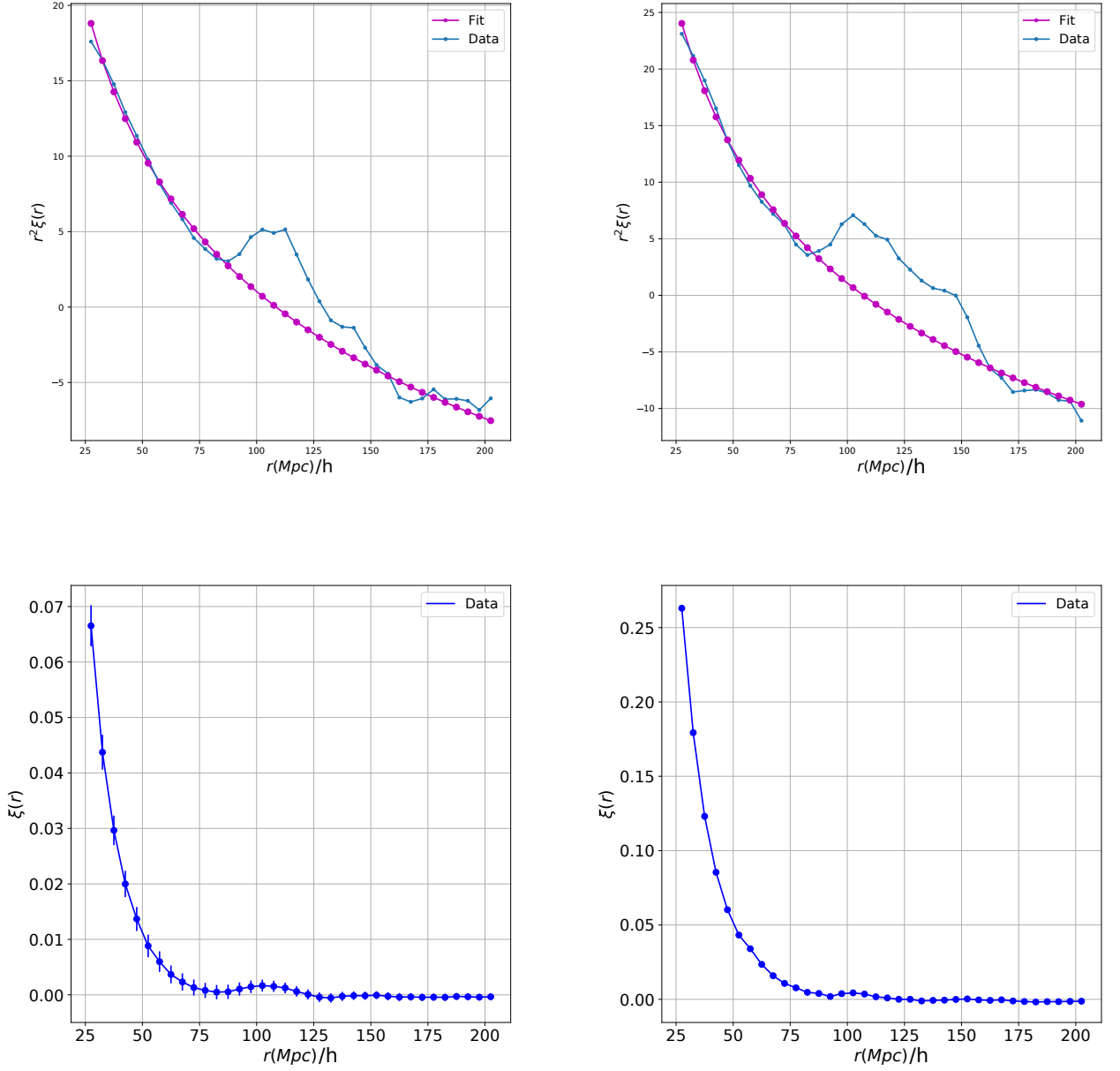
Figure 4.1: The correlation functions for populations 1 with $N_{random} = 10$, 2 with $N_{random} = 10$, 3 with with $N_{random} = 10$ and 4 with with $N_{random} = 40$ are displayed from the upper left to the lower right. For each population 3 runs with 12 bins were performed.

MDPL population	Mean Mass $M_\odot/h$	Number of halos
9	$1 \times 10^{13.5}$	8462
8	1×10^{13}	66473
7	$1 \times 10^{12.5}$	326163
6	1×10^{12}	1258759
5	$1 \times 10^{11.5}$	4333890

Table 4.2:   Populations constructed for $z = 1$.

the population N_{part}. The random sample factor accounts for the size of the random-random catalogue, the total population for the RR catalogue is $N_r N_{part}$. The quanties R_{min} and R_{max} define a range where correlation function is found. With this range and the radial number of bins a distance is estimated, points where the correlation function is computed. The correlation function was estimated using the LS estimator shown in the section 3.5.

Next, in the figure 4.1, the correlation function for every population is shown. The random factor used is not the same for each case, it diminish with the bigger populations since it would be too expensive computationally to use a big enough value for all of them. For each population 3 different runs were performed with the same N_r and $N_b = 12$. The first one used $R_{min} = 20$ and $R_{max} = 200$ Mpc/h, the second one $R_{min} = 25$ and $R_{max} = 205$ Mpc/h and the last one $R_{min} = 30$ and $R_{max} = 210$ Mpc/h. In this way the noise in the final correlation function plotted per population is reduced. This is, if only one run from $R_{min} = 20$ to $R_{max} = 210$ Mpc/h with bins width of 5 Mpc/h would have been performed, less data would have been used to calculate the correlation function per bin compared with the 3 runs performed, making more robust the results in the last scenario.

In the different figures the correlation function obtained has a similar shape and there is a bump around ~ 105 Mpc/h that corresponds to the BAO peak. But there is a difference in the amplitude and the particular population. The more massive halo populations have a larger amplitude compared with the lower ones. This is more clear from the maximum value obtained in each case.

Further, an error bar was found for each point where the correlation function was found. For this, 10 realizations with the same characteristics for each population were calculated. The standard deviation found using the realizations are the values used for the error bars. In the figures there are a decrease in the error bars for each population as they become more

massive. Hence, the correlation function for the realizations are more alike for the smaller populations.

4.1.1 Random Sampling and Number of particles

Since the random sampling factor is intended to reduce the shot noise, it is important to study its impact on the calculation of the correlation function (CF). A first exercise in this direction is displayed in the figure 4.2 for the population 4, the correlation function was calculated for 3 different random sampling factors. There is something important to highlight, there were performed ten realizations per N_r and the CFs shown are the mean of these realizations. From the figure what can be seen is that there is no significant change in the CF obtanied due to the random sampling number, not at least for radial values smaller than 150 Mpc/h. For bigger values, the CF with a smaller random sampling factor behavior becomes a little bit noisier but this scales are not of our interest.

If the same exercise is repeated with specific realizations instead of the mean CFs, more fluctuatios are introduced producing a more notorious difference for scales larger than 150 Mpc/h.
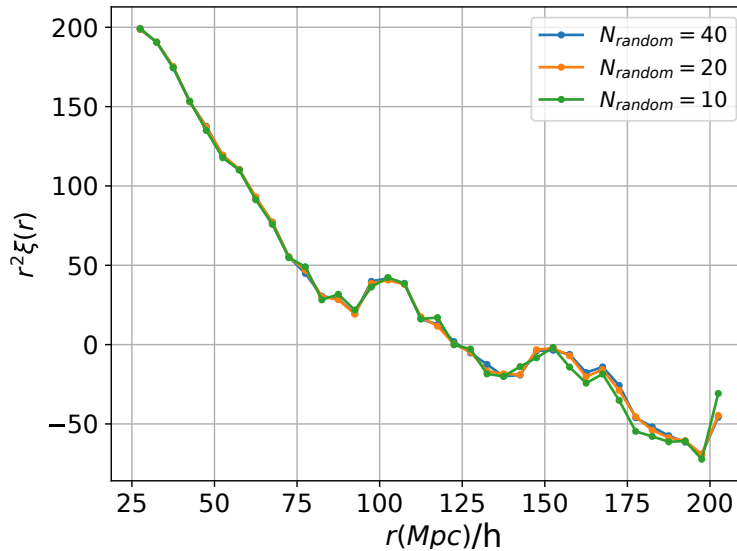

Figure 4.2: Correlation function for population 4 with different random factor numbers.

Other exercise performed to find the effect of the N_r on the CF is shown in the figure 4.3. In this case, for a thick population, $10^{14.5}$ $M_\odot/h$ several realizations with the same characteristics were performed except by the N_r used. The two values for N_r are 50 and 100

with equal number of realizations per calculation. The plotted line in each CF corresponds
to the mean of the realizations. The xrange was reduced to see in more detail the bump of
the BAO. The dispersion of the CFs values for the two N_r values is very similar. Hence, there
is no change in the CF estimation because of N_r value. Though the left figure appears with
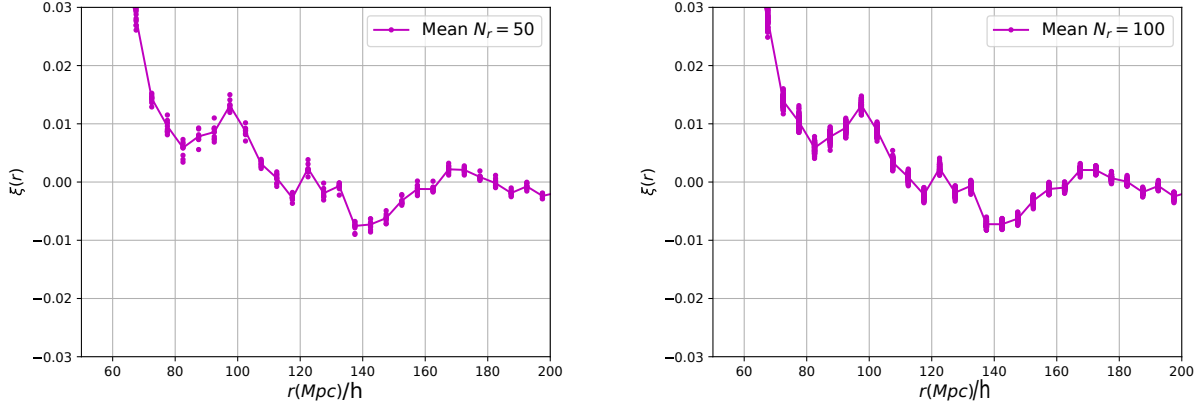more points per bin, this is caused because of the difference in the number of realizations
done.

Figure 4.3: Correlation function for population 14.5 with two N_r values. The left panel
displays 50 different CFs and the right one 100 different CFs.

There was another factor that has to be considered, the number of particles taken for the
correlation function calculation. Since there is a large number of particles for the population
1 and 2, there is not enough computational resources to run the program, hence it is necessary
to use a subsample representative of all of population. Now, the idea is to study the effect it
has on the estimation of the correlation function. In the figure 4.4 is shown for the population
1 the correlation function obtained for two different subsamples, one is around 7.6% of the
total population and the other one is around 15.2% percentage of the total population. Both
correlation functions coincide for ranges lower to 80 Mpc/h, in the region where BAO.

4.2   Correlation function fit

A first step torward obtaining the BAO signal and its properties from the CF, it is to make a
fit precisely of the CF. In the figure 4.5 the CF for different populations are shown, the label
used for them is *data*. As it can be seen the y axis corresponds to $r^2\xi(r)$ since it allows to
visualize better the BAO bump.

The CF function fits that we are finding do not try to reproduce the bumps observed
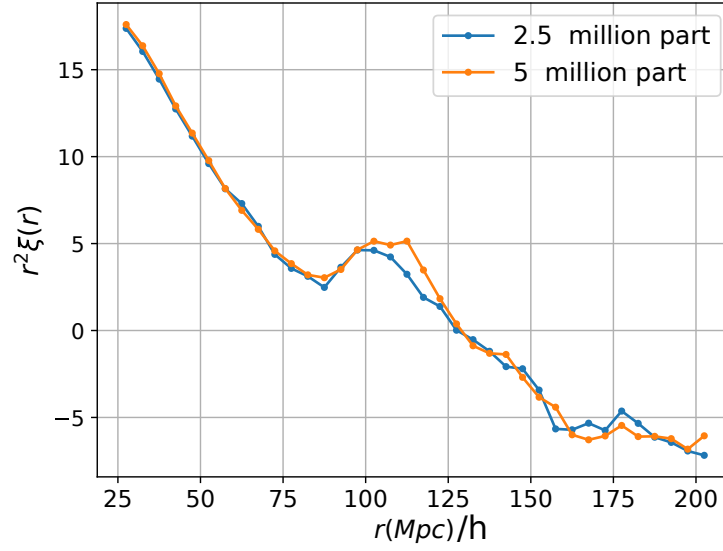
Figure 4.4: Correlation function for two different subsamples of population population 1 .

in the correlation function. Because of this, not all the points obtained for a CF estimation were used to make the fit, only those ones that do not fall in the bumps.

The fit of the points were adjusted to the function

$$f(r) = r^2\xi(r) = \frac{r^\lambda}{a_0} \tag{4.1}$$

where λ and a_0 are the parameters found with the fit. The theoretical function used to adjust $\xi(r)$ was already shown in 2.18. Since it is is a power law function, an easier way to carry out the fit is in logarithmic scale. In this way, it is performed a linear fit: $ar + b$. But, the values of the CF can be negative. Then, to avoid any trouble a Δ value is summed to the CF to get only positive values before perfoming the fit.

When the coefficients a and b are obtained, we can recover the initial parameters λ and a_0 by considering the fact that $\gamma = a$ and $a_0 = \exp(-b)$. Now, replacing the parameters in the expression 4.1 and subtracting the amount Δ, the CF can be plotted.

This procedure is repeated to obtain the fit for every population and all of its realizations. In the figure 4.5 some of the fits obtained for the samples are displayed. Here, one important thing is to get a measure of the robustness of the fit. Hence, error bars are calculated using the different realizations of the CF, i.e., the standard deviation is obtained. The error bars are shown for every figure of 4.5 but because of the similarities of the fits they are almost no visible. So the CF fits are considered robust enough.
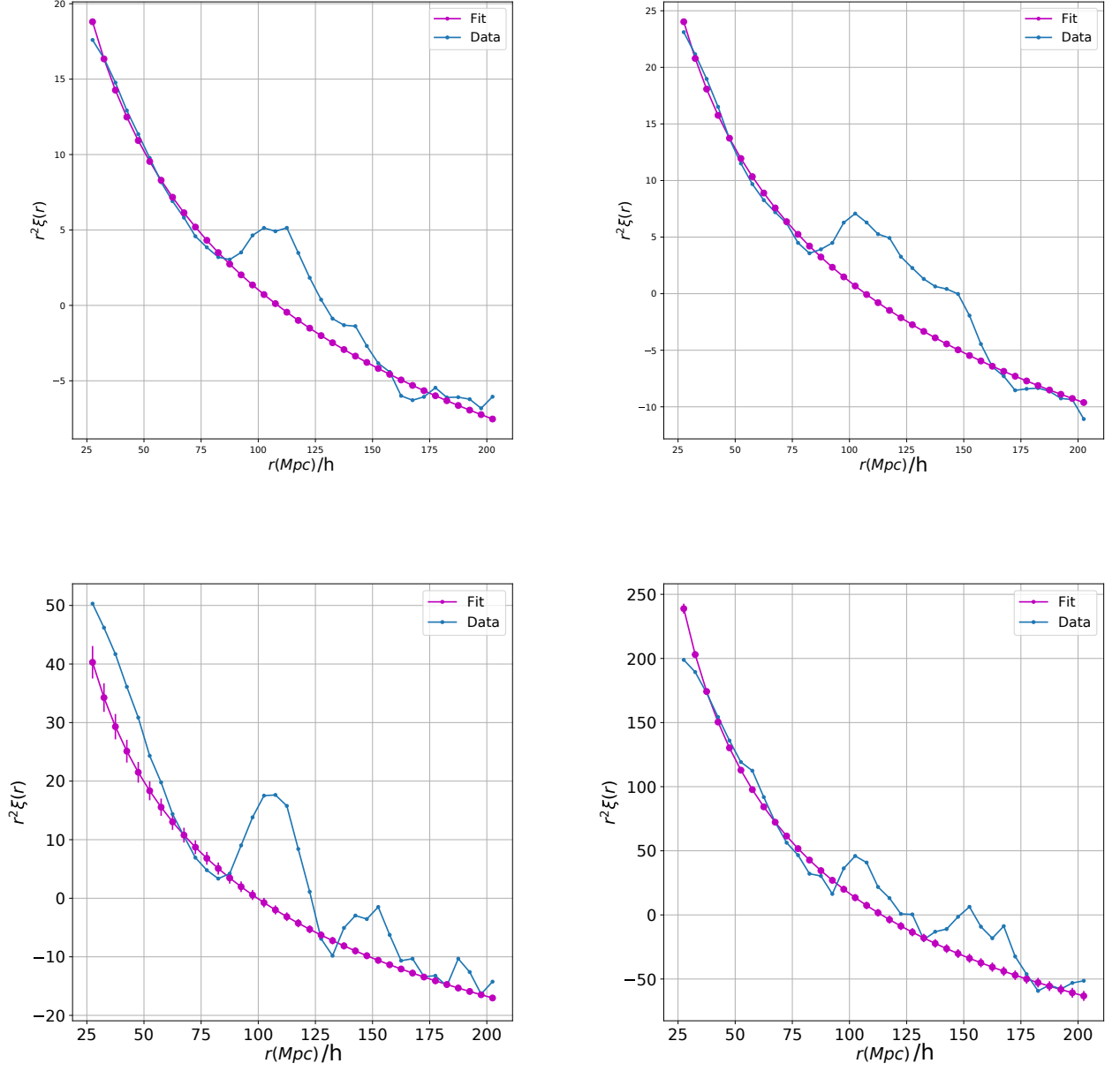
Figure 4.5: The correlation functions for populations thick 1, 2, 3 and 4 are displayed from the upper left to the lower right. For each population 3 runs with 12 bins were carried out.

In all of the CFs in 4.5 there are visible two bumps, the first one corresponds to the BAO since it agrees with the value observed of BAO peak as it was shown in figure 2.5. Also the position coincides with the BAO position measured for galaxy clusters, a mass range we are considering in our populations.

Something to highlight is that in the population q_2 de $1 \times 10^{12} \, M_\odot/h$ there is no separation of this two bumps, making more difficult to recognize the BAO signal. It also happens with the population 2 not shown in this figure. This was precisely the reason to study in more detail the effect of the mass bins in the CF estimation and thus the BAO bump.

4.2.1 BAO fit

Since the measure of our interest is the BAO bump, it becomes necessary to extract it from the CF and thus to be able to obtain the properties of BAO we are looking for to analyse. In this direction, a correlation function model can be useful

$$CF(r) = \xi(r) - \Delta + GF(r, A, \mu, \sigma) \tag{4.2}$$

the term $\xi(r)$ corresponds to the theoretical form shown in 2.18, for us it is recovered from the CF fit divided by r^2. The second term is the one that it is summed to the CF as was explained in the previous section, so it must be subtracted here. The function GF is a gaussian fit that reproduces the BAO shape recovered from the correlation function.

$$GF(r, A, \mu, \sigma) = Ae^{-(r-\mu)^2/(2\sigma^2)}$$

This model where BAO is fitted with a gaussian function is proposed in [9].

Then, the next steps are followed to recover the BAO bump for every population. The median of the realizations is taken as the principal CF and the standard deviation is obtained through the realizations.

1) The term Δ is substracted from the CF fit. After the function is divided by r^2 obtaining $\xi(r)$.

2) The function $\xi(r)$ is substracted from the estimated CF leaving the signal of the two bumps.

3) Only the points corresponding to the first bump located around 105 Mpc/h are selected.

4) A gaussian fit of the first bump is performed. Only the more central points of the signal are considered since the outer ones are the noiser parts of the signal. This noise could be diminished using more realizations per population.

5) The fit is also performed for all the realizations, the same points considered for the BAO signal fit of the mean CF are taken for the remaining realizations.

6) The parameters amplitude A, mean μ and standard deviation σ are obtained for every population and the realizations.

The parameters that characterizes the BAO bump are the amplitude A, the position μ and the width 2σ. In the figure 4.6 some fits of the BAO signal are shown. It can be noticed that the points that correspond to the BAO bump have a gaussian like distribution. The Gaussian fits performed are also shown with the parameters found. It can be noticed a good coincidence between the data and the fit obtained. The error bars displayed were obtained through the BAO fits performed for the different realizations, with them the standard deviation per population was found.

For every plot displayed in the figure 4.6, there is a curve labeled as B spline. This fit was obtained using basis splines. It has a different behavior than the one observed for the Gaussian fit. At least in general terms, it fits better the BAO signal recovered. Furthermore, there is a difference in the main peaks of the two fits performed, Gaussian and basis spline fit. Despite the difference between the peaks, the functional form obtained for the properties is similar. This will be shown in the next section.

4.3   BAO properties in the populations of MDPL

As mentioned previously, the BAO peak is clearly detected for every population and the fit to the BAO signal was properly calculated using a Gaussian function. Let us see the situation in more detail. There are different populations, each of them have the ranges in mass shown in the beginning of this chapter.

For the thick bins, we are considering per population, more massive halos each time. That way, we are be able to analyze if there is an effect on the properties of BAO obtained for each population. It is important to take into account that more massive halos trace higher density peaks in the matter density field. This should lead, in principle, to a stronger correlation in the most massive populations compared with the less massive ones. Thus, a better detection
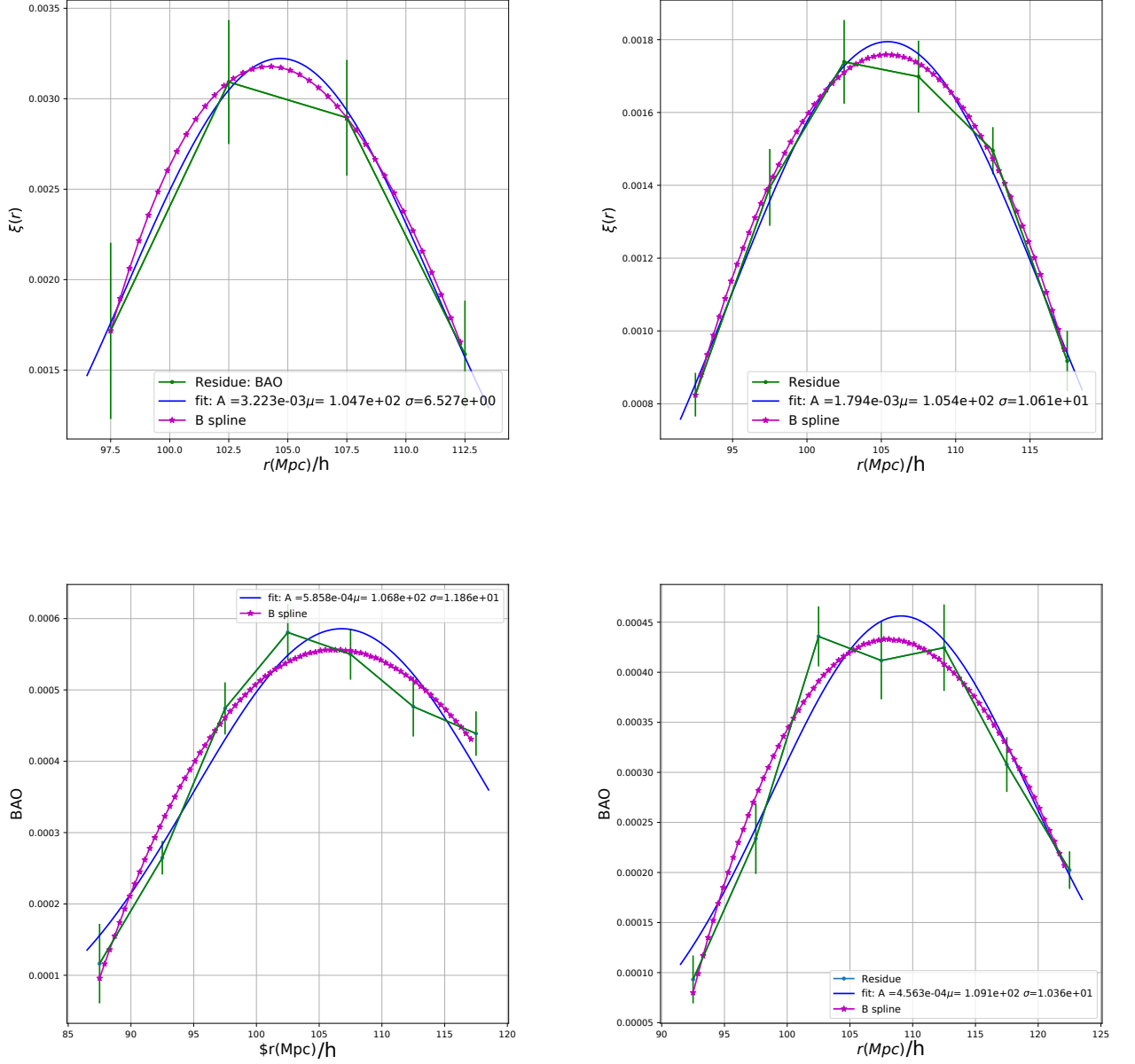
Figure 4.6: BAO signal for population 4, 3, q_4 of $1e12$ and q_1 of $1e11$ from upper left to lower right.

of the BAO signal. In the left figure 4.7 an increase of the amplitude of the BAO for more massive halos is obtained as precisely expected for a stronger correlation.
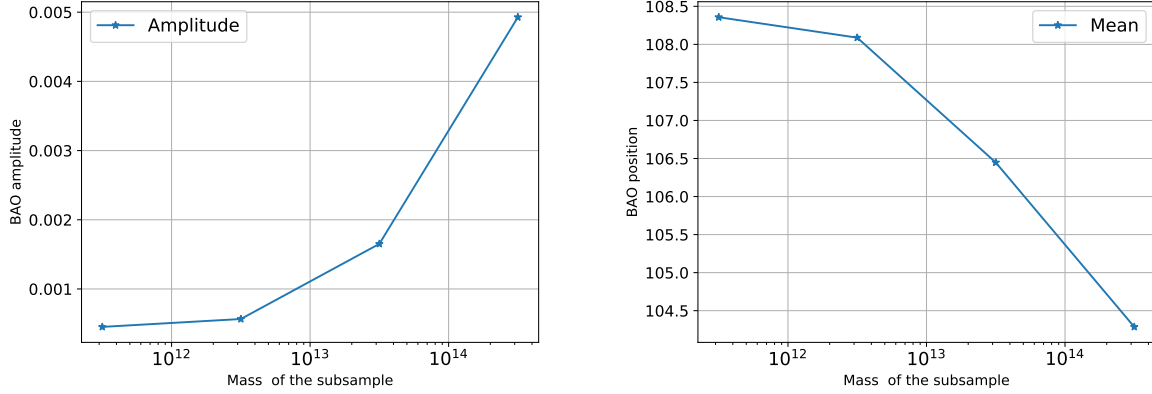

Figure 4.7: In the left panel the amplitude of BAO in function of the mass of the thick populations is shown. In the right one the position of BAO in function of the mass of the thick population is displayed.

The initial position of the BAO depends on the sound horizon scale as mentioned in 2.19, but as explained in the model exposed in 2.7.1, this position changes due to nonlinear effects. Now, the plot in the left of 4.7 shows the position of the BAO in function of the halo mass. It can be noticed a decrease in the value of the position as the halo mass increase. This behavior could be expected due to nonlinear gravitational collapse. The velocity field causes a movement of the BAO peak to bigger scales for the nonlinear regime. In this case, this corresponds to the smaller halo masses for which the density modes are coupled among them. The porcentual difference is around 3.8%. This is precisely of the same order of magnitude found in [9]. This result contributes in a different way as [9] since they consider cosmological boxes with a very small resolution 640^3 in contrast to the populations used in this study. Furthermore, we study the BAO behavior for a more extended mass range and this could let us see the effect of BAO properties in the nonlinear range.

In order to understand our results, lets propose a toy model that may help to see what is happening on the BAO. Consider a single density perturbation that is sourronded by a BAO as it is shown in the figure of 4.9 a. Here, we are going to analyse the behavior of the BAO properties due to nonlinear effects. As the system evolves, the velocity field causes changes in the BAO properties. To analyse what happens, we make a separation by halo mass and analyse the difference in the velocity field in each case. Now, consider the right figure of 4.9 where a toy density profile is shown and separated in two parts. The left curve is the one
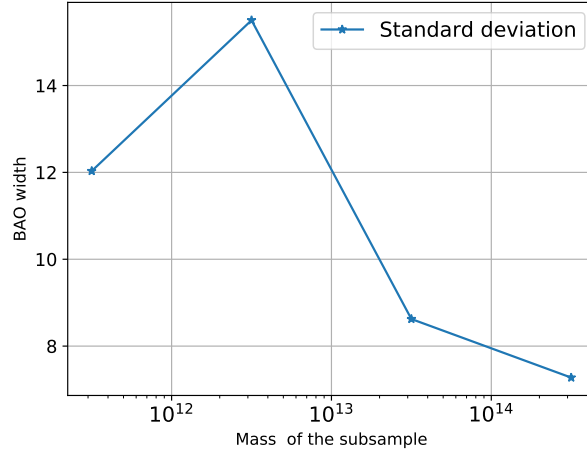
Figure 4.8: Width of BAO versus the masss of the thick populations.

halo term and the right curve is the BAO signal. The one halo term is esentially the density profile of a halo and it is affected by the dispersion of velocities. This dispersion causes a broadening of the profile. Hence, a similar behavior is expected for the BAO signal. This is, the halos in the BAO signal have a dispersion of velocities that causes a broadening. But, the halo mass considered changes the broadening. In this scenario, when we are considering halo populations with smaller masses, it is expected a bigger dispersion of the velocities compared with the populations with bigger masses. This occurs due to more massive halos are harder to move, i.e., they move closer to the ordered linear flows. Thus the deviation from the mean velocity is not so big. At least, compared with the smaller halos that can have a wider range of velocities. This idea is supported by the distribution of velocities observed in different populations as shown in the figure 4.10, where we see that for larger halo masses populations the velocity dispersion is smaller. This is in accordance with the toy model proposed. In the figure 4.10 is also seen a bigger mean velocity for larger masses populations.

The previous discussion is a possible explanation of what it is seen in the figure 4.8 where the width of the BAO decreases with more massive halos. But, the behavior of the population 2 does not follow this tendency. Because of this, the quarter and thin populations were created to study in more detail the effect of the mass bins in the properties measure of BAO.

The BAO properties obtained for all the populations, except 1×10^{11} and 1×10^{12}, are plotted in the figures 4.11 and 4.12. It is noticed that when it is used thiner masses bins, we do not recover the same tendency observed in the previous figures. Let us start with the
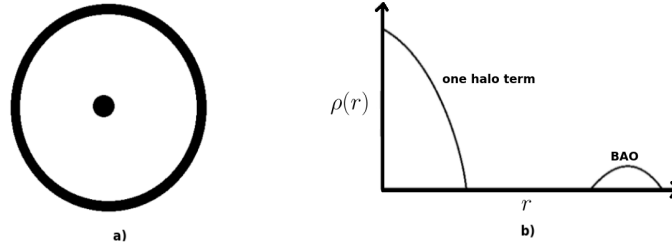
Figure 4.9: a) Scheme of a BAO. b)Profile that includes one halo term and the BAO signal.

left figure of 4.11. In this case, there is still a tendency of the amplitude to increase with an increase of the halo mass. Though, for masses smaller than $\sim 10^{12.4}$ $M_\odot/h$, there is a more irregular behavior compared with the one obtained for the thicker bins.

Now, in the right figure of 4.11, blue curve, it is not recovered the same behavior as 4.8. But, for halo masses bigger than $\sim 10^{12.4}$ $M_\odot/h$, the width of the BAO decreases with mass. This was expected as explained for 4.8. For masses smaller than $\sim 10^{12.4}$ $M_\odot/h$, the nonlinear effects causes that there is no tendency of the data. The red curve shown in the same figure corresponds to the width in function of the halo mass for $z = 1$. It has a better behavior since nonlinear effects are not so strong. Hence, it can be noticed that because of the structure evolution between $z = 1$ and $z = 0$, the coupling among different density modes becomes more notorius.

Figure 4.12, shows the position of the BAO vs the halo mass. In the case of the solid blue line, the populations considered are those ones that belong to $z = 0$, with the properties recovered using the Gaussian fit. To analyse it, let us again divide the figure in two regions. For halo masses bigger than $\sim 10^{12.4}$ $M_\odot/h$, a similar behavior compared with 4.7, i.e., a decrease in the BAO position with the increase of the halo mass. As was previously mentioned, this behavior is expected as explained in 2.7.1. But, for masses smaller than $\sim 10^{12.4}$ $M_\odot/h$, it is not observed any specific tendency. It is in this region where the coupling among density modes makes the nonlinear effects too big to extract any possible tendency. Hence, it is out of the purpose of this work to go deeper into the subject. A second curve is the dashed blue line, it corresponds the same populations as the solid blue line. But, in this case, the position of the BAO was recovered using a basis spline fit. This was perfomed to make the results obtanied with the Gaussian BAO fit more robust. Although, there is a difference in the properties recovered, this values are similar and also they behave in a similar way. Thus, the region for masses larger than $\sim 10^{12.4}$ $M_\odot/h$ shows a shift of the BAO position to smaller distances. Hence, this supports the result obtained with the Gaussian fit. The red solid curve
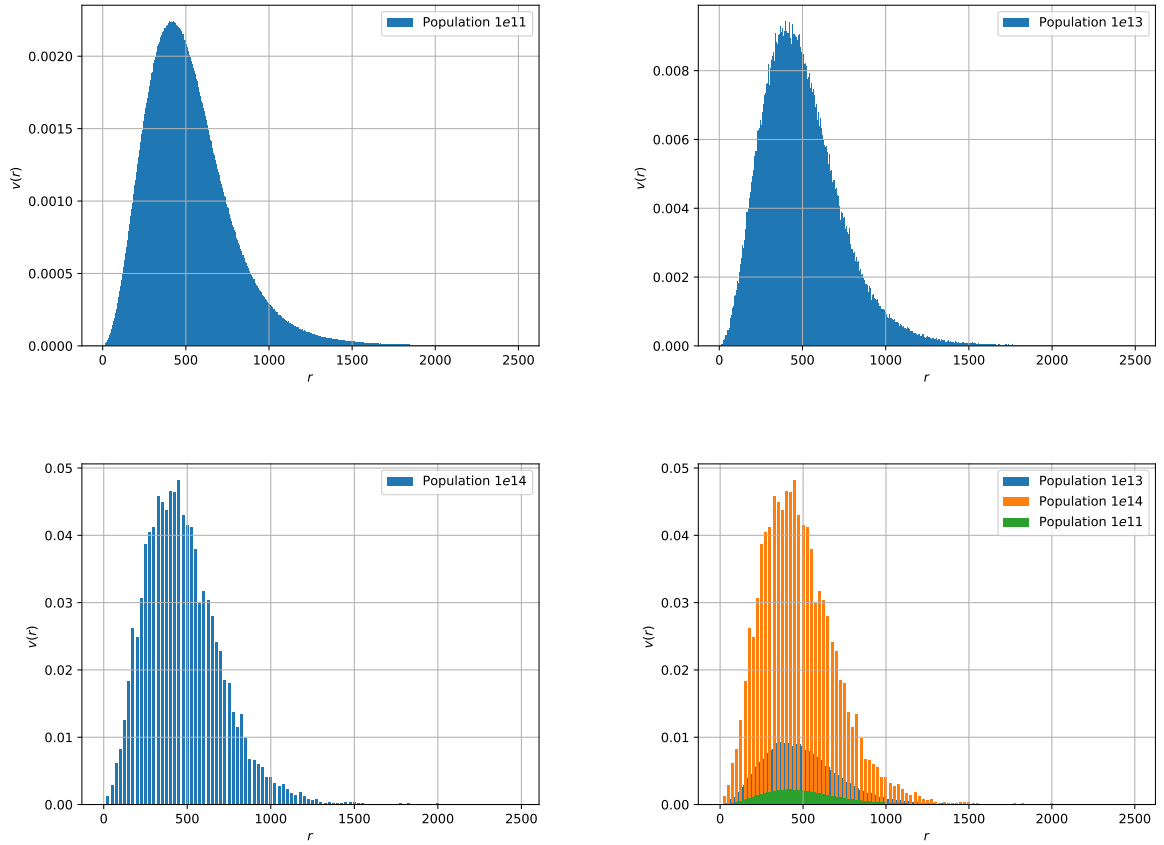
Figure 4.10: Figure with population 1×10^{11} has a mean of 519.51 km/sh and it's estimated variance is 67214 km/sh. Figure with population 1×10^{13} has a mean of 501.81 km/sh and it's estimated variance is 58720 km/sh. Figure with population 1×10^{14} has a mean of 490.21 km/sh and it's estimated variance is 53063 km/sh. It can be seen that the population $1e14$ has the smaller variance as it is predicted by our model. In the last figure are compared all the populations.

shows the BAO properties recovered for the populations at $z = 1$. It can be seen that BAO position does not change significantly because the nonlinear effects are less prominent in this epoch. The last dot of this curve was produced with a population were small, hence this specific result is not so reliable. Similarly, the dashed red line displays the properties found for the same populations as the solid red line, but it recovers the position using a basis spline fit. For both curves, there is not a notorious change in the position of the BAO. Hence, these two curves support the idea that nonlinear effects are causing a shift in the BAO position, not so evident for $z = 1$.

Concluding, the properties recovered for the BAO using the thicker bins are behaving as expected but they are possibly "masking" information of smaller mass scales. For this, the thiner bins provided us with a more accurate properties. This lead us to see that for masses
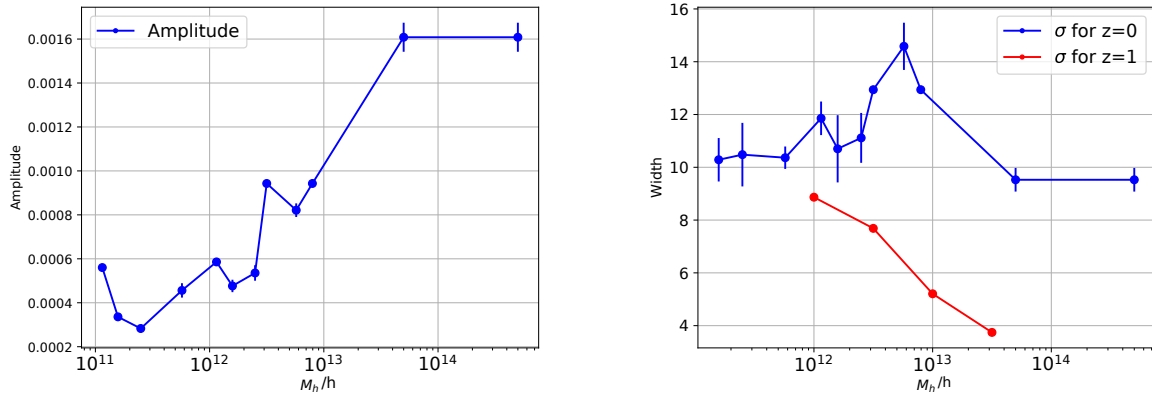
Figure 4.11: In the left panel the amplitude of BAO in function of the mass of all populations is shown. In the right one width of BAO versus the masss of the thick populations.
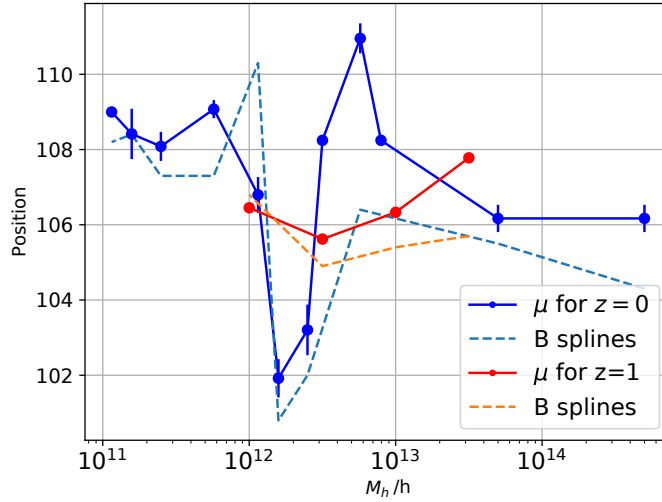

Figure 4.12: The position of BAO in function of the mass of the thick population is displayed.

larger than $\sim 10^{12.4}\ M_\odot/h$, we recovered the expected tendency. But for the smaller masses, the nonlinear effects do not let us extract any specific behavior of the properties. A nonlinear behavior is dominating the BAO properties at these mass scales.

During this chapter the results for two different numerical experiments were performed using a halo catologue from MDPL simulation. The first case used thick bins for the populations built. It was noticed that BAO properties changed due to the scale of the tracer halo population used. The amplitude increases with larger halo masses due to the bias. The position and width decreases for larger halo masses. The last two results were explained us-

ing a toy model that only considered the one halo term and a single BAO contribution for a density profile as shown in the figure 4.9. The second experiment was performed in the same way that the previous one but we built populations using thin bins. A similar behavior was recovered for BAO properties but only for masses larger than $\sim 10^{12.4}$ $M_\odot/h$. This lead us to study the BAO properties in two different regions. For larger masses than $\sim 10^{12.4}$ $M_\odot/h$ were recovered a similar behavior compared with the first experiment. But for smaller masses than $\sim 10^{12.4}$ $M_\odot/h$, it was not obtained a specific tendency. We suppose this behavior is caused by stronger nonlinear effects.

CHAPTER 5

Summary and conclusions

The bayronic acoustic oscillations is an imprint that appears in the distribution of matter for large scales. And it is mainly studied because its size scarcely changes with the evolution of structure formation. Particularly, we want to find if there is some change in BAO properties with the tracer halo population used. Using a halo catalogue from the MDPL simulation, different populations were constructed.

A first experiment was perfomed for the thick populations shown in 4.1. The CF was found for every population and the BAO signal was extracted . Later, using a Gaussian fit the BAO properties were obtained, i.e., amplitude, position and width. From this first results, the figures 4.7 and 4.8 were obtained where the it is shown how every propertie changes with the halo mass considered. For the amplitude it can be concluded that it suffers an increase with halo mass increase. This is due to the bias that is bigger as the halo masses considered are bigger. For the BAO position, it was found among every halo mass considered a difference of around 3%. Though it is a small diference, the result is of the same order of what was found in [9]. For the width, there is a change of the width with the halo mass, except for the second population. The general trend is that BAO width is samller for more massive halos. As there is an increase in the halo mass of the population studied, it is obtanied more clustered populations. This could be explained because the dispersion of velocities of the halos becomes smaller as it is increased the halo masses considered. This relation is not satisfied for the population 2 hence it appears as a motivation to perform a smaller binned, specially in this region to study in more detail how the amplitude behaves.

In a second experiment that is esentially equal to the previous one but using the thin populations. Once again the CF is calculated for every population and the BAO signal is extracted. And using a BAO fit the properties of interest are found. Thus, using this results the plots 4.11 and 4.12 are constructed.

To analyse them, two regions for the plots are used. A linear region, where it is observed a similar behavior to the ones recovered in the first experiment for each property. This is, from the population $10^{12.75}$ $M_\odot/h$ to the most massive population, it is observed a similar tendency than the one observed in the previous figures. The nonlinear region is from masses smaller than $10^{12.75}$ $M_\odot/h$ to the less massive ones where the density modes are coupled among them.

Now, the amplitude in the left figure of 4.11 recovers a similar shape but for smaller mass scales, we argue that the nonlinear gravitational effects causes that the tendency expected is not properly recovered.

The width of the BAO plotted in the right figure 4.11 for linear region has the same tendency once again. But for nonlinear region, it can not be explained the behavior observed and it is out of the scope of this work.

Similarly, the position of the BAO observed in the figure 4.12 can be analyzed for the two mentioned regions and in the nonlinear region, we could not recovered the same expected tendency.

Concluding, for populations with smaller masses than $10^{12.75}$ $M_\odot/h$ the nonlinear gravitational effects lead to a non expected behavior that can be not easily explained and that is further of the interest of this work. But in bigger mass scales, a tendency is recovered for each property. The main behavior is explained with the velocity field of the structures "within" the BAO. For nonlinear gravitational effects the velocity field causes a shift to bigger scales of the BAO peak. This is what is observed in the figure 4.12 where as the halo mass increases the BAO position diminishes. Furthermore, the velocity dispersion of the different populations depend on the mass scale. Thus, the bigger the halo mass considered the smaller the velocity dispersion is. Hence, the width of the BAO must decrease as the halo mass taken increases. This is precisely what is observed in the right figure 4.11 for the linear region. Also, the amplitude diminishes while the halo mass increases as shown in the left figure 4.11 due to the bias. According to the last results, it is concluded that the BAO properties depend on the halo mass tracer for the linear region defined.

The previous results led us conclude that the BAO position was shift to smaller distances

for larger halo masses, around $\sim 4\%$. This has a direct impact on the parameters of the equation of state of the dark energy. As it was mentioned in [9], [10], [11], [12], [13], [14], the importance of the change in the position of the BAO peak of $\sim 1\%$ lead to a change of the dark energy parameters of $\sim 4\%$. Thus, with the results that we found should imply a notorious change in the dark matter parameters.

References

[1] E. Komatsu, J. Dunkley, M. R. Nolta, and C. L. Bennett, "Five-year wilkinson microwave anisotropy probe (wmap) observations: Cosmological interpretation," 2008.

[2] D. J. Eisenstein, I. Zehavi, and et al., "Detection of the baryon acoustic peak in the large-scale correlation function of sdss luminous red galaxies," 2005.

[3] M. Tegmark, D. J. Eisenstein, and et al., "Cosmological constraints from the sdss luminous red galaxies," 2006.

[4] P. "Ade, N. Aghanim, and e. a. Akrami, "Planck 2015 results. XVI. Isotropy and statistics of the CMB," *Astron. Astrophys.*, vol. 594, p. A16. 62 p, Jun 2015.

[5] M. S. Longair, *Galaxy Formation.* Springer, second ed.

[6] H. Mo, F. van den Bosch, and S. White, *Galaxy Formation and Evolution.* Cambridge University Press, 2010.

[7] T. Padmanabhan, *Cosmology and Astrophysics Through Problems.* Cambridge University Press, 1996.

[8] P. Ruiz-Lapuente, *Dark Energy: Observational and theoretical approaches.* 2010.

[9] R. E. Smith, R. Scoccimarro, and R. K. Sheth, "Motion of the acoustic peak in the correlation function," *Phys. Rev. D*, 2008.

[10] T. Baldauf and V. Desjacques, "Phenomenology of baryon acoustic oscillation evolution from lagrangian to eulerian space," *Phys. Rev. D*, 2017.

[11] S. Anselmi, G. D. Starkman, P.-S. Corasanitii, and et al., "The linear point: A cleaner cosmological standard ruler," *arXiv:1703.01275 astro-ph*, 2017.

[12] M. M. Ivanov, "Systematic treatment of non-linear effects in baryon acoustic oscillations," *EPJ Web of Conferences 125 , 03006*, 2016.

[13] R. E. Smith, R. Scoccimarro, and R. K. Sheth, "Scale dependence of halo and galaxy bias: Effects in real space," *Physical Review D*, 2007.

[14] C. Martín and R. Scoccimarro, "Nonlinear evolution of baryon acoustic oscillations," *Physical Review D*, 2008.

[15] D. Jeong, *Cosmology with high (z > 1) redshift galaxy surveys.* Doctoral dissertation, The University of Texas, 2010.

[16] S. Pfalzner and P. Gibbon, *Many-Body Tree Methods in Physics.* Cambridge University Press, 1996.

[17] A. Klypin, "Numerical simulations in cosmology i: Methods," 2000.

[18] A. Klypin, "Numerical simulations in cosmology ii: Spatial and velocity biases," 2000.

[19] A. Klypin, "Numerical simulations in cosmology iii: Dark matter halos," 2000.

[20] Y. P. Jing, "Correcting for the alias effect when measuring the power spectrum using a fast fourier transform," *The Astrophysical Journal*, vol. 620, no. 2, p. 559, 2005.

[21] F. Montesano, A. G. Sanchez, and S. Phleps, "A new model for the full shape of the large-scale power spectrum," *Monthly Notices of the Royal Astronomical Society*, vol. 408, no. 4, pp. 2397–2412, 2010.

[22] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes 3rd Edition: The Art of Scientific Computing.* New York, NY, USA: Cambridge University Press, 3 ed., 2007.

[23] D. Jeong and E. Komatsu, "Perturbation theory reloaded. ii. nonlinear bias, baryon acoustic oscillations, and millennium simulation in real space," 2009.

[24] M. J. P. Borderia, V. J. Martinez, D. Stoyan, H. Stoyan, and E. Saar, "Comparing estimators of the galaxy correlation function," *The Astrophysical Journal*, 1999.