

# Large scale structure of the Universe: statistics from cosmological simulations

---

## Author

Martín Manuel Gómez Míguez

## Tutors

Claudio Dalla Vecchia  
Andrés Balaguera Antolínez  
*Instituto de Astrofísica de Canarias*  
*Departamento de Astrofísica ULL*



Grado en Física



San Cristóbal de la Laguna , 16 de Septiembre de 2019

# Agradecimientos

La historia detrás de este trabajo es un tanto compleja. Contrario a lo esperado, mi primera idea a cerca del proyecto de fin de grado estaba relacionada con dinámica de fluidos y fractales. En primer lugar, debo agradecer al profesor Teodoro Roca toda la ayuda ofrecida para intentar dar forma a la idea inicial, aunque finalmente no fructificase. Debo agradecer también todo el apoyo recibido por parte del investigador Ricardo Gil-Hutton, el cual desinteresadamente me tendió una mano y me sirvió como guía para orientarme en mi formación como físico. Fue gracias a él que me puse en contacto con el doctor Licandro del IAC para conocer si existía la posibilidad de poder realizar mi propuesta de proyecto. Dado que no había ningún investigador experto en el área en el IAC y mi interés por la cosmología, optó por presentarme a Claudio, quien me propuso el trabajo que aquí se expone.

Una vez situados en el punto de partida de este proyecto, debo mostrar mi más sincero agradecimiento a mis dos tutores, Claudio y Andrés, los cuales me propusieron un ambicioso proyecto con el cual han avivado mi vocación por la física y la cosmología. A parte de por compartir esas horas de desesperación en las que nada parecía claro, les estoy especialmente agradecido por ofrecerme la posibilidad de incluirme en la dinámica de sus grupos de investigación, lo cual ha supuesto todo un estímulo para mí, haciéndome sentir en todo momento integrado, ser uno más del grupo.

Deseo reflejar también un caluroso agradecimiento a mi familia, quien no comprendiendo bien el camino en el que embarcaba cuando abandoné mi hogar en mi querida Galicia para adentrarme más profundamente en el mundo de la física y saciar mis inquietudes, siempre me ha apoyado en todo momento.

Quisiera cerrar los agradecimientos haciendo una especial mención a mi padre, que no pudo acompañarme hasta el final de este viaje. Su esfuerzo a la hora de trabajar siempre ha sido todo un ejemplo para mí y espero poder hacer que sienta orgulloso del trabajo aquí presentado.

# Extended Abstract

La mejora de las técnicas observacionales ha permitido aumentar considerablemente tanto el catálogo de observaciones como el conocimiento sobre las anisotropías del fondo cósmico de microondas, favoreciendo el desarrollo de la cosmología. Con el objetivo de reproducir la evolución de los fenómenos cosmológicos y predecir las observaciones de futuras misiones científicas se realizan simulaciones de N-cuerpos en supercomputadores, debido al alto costo tanto en memoria como en tiempo computacional al considerar un número elevado de partículas. En el caso práctico del presente trabajo, se realiza una breve introducción al modelo cosmológico estándar así como a los parámetros que lo caracterizan y se realizan diferentes simulaciones cosmológicas empleando el código PKDGRAV3 para un volumen cúbico de lado 100 Mpc/h, partiendo de un redshift inicial  $z = 49$  hasta tiempos actuales almacenando diez pasos temporales de la distribución de partículas. Los objetos a estudiar serán partículas de materia oscura cuya dinámica puede ser descrita por las ecuaciones de un fluido ideal de presión nula. Para analizar el comportamiento estadístico de las distribuciones de materia obtenidas, se desarrollará el formalismo teórico de la estadística a un punto y a dos puntos con el fin de aplicarlo sobre dichos resultados.

En primer lugar, se implementarán dos de los criterios de asignación de masa más comunes para interpolar la distribución de materia, de naturaleza discreta, a una malla que divide el volumen total en celdas cúbicas. A continuación, se obtiene la función de distribución de las fluctuaciones de la densidad de materia sobre la densidad crítica usando como base los resultados de simulaciones de  $128^3$  y  $200^3$  partículas, comparando cómo la función filtro escogida afecta a la construcción de esta y las diferencias observadas para ambas simulaciones. El siguiente punto será ajustar las mediciones al modelo lognormal, el cual es el modelo analítico más sencillo para caracterizar comportamientos no lineales de una distribución estadística. Asimismo, se estudiará la evolución de la distribución con el redshift a través del análisis de los cuatro primeros momentos de esta, obteniéndose que, con el paso de tiempo, las zonas con sobredensidad de materia colapsan, concentrando la mayor parte de la materia y vaciando progresivamente las demás regiones del Universo, lo que da lugar a la formación de estructuras como halos de materia oscura.

En relación al análisis estadístico a dos puntos, se parte de la distribución discreta de materia para obtener la función de correlación, la cual mide la probabilidad de encontrar un par de objetos separados por una cierta distancia en comparación con lo que se observaría para una distribución aleatoria. Para ello, se miden los histogramas DD (Data-Data), DR (Data-Random) y RR (Random-Random), los cuales dan

pie a definir una serie de estimadores como los sugeridos por Peebles y Landy-Szalay, ofreciendo resultados muy similares. La medición también es realizada para diferentes instantes temporales, haciéndose un estudio de la evolución de la influencia de la gravedad en la formación de estructura a diferentes escalas en distintas épocas del universo. El espectro de potencias, definido como la transformada de Fourier de la función de correlación, se expresa en términos del módulo al cuadrado de la transformada de Fourier de la función de sobredensidad, lo que presenta una ventaja en tiempo computacional con respecto a la función de correlación, por lo que es la función más extendida para caracterizar la estadística a dos puntos. Debido a la naturaleza discreta de la muestra estadística, las mediciones presentan un error conocido como 'shot-noise' el cual se puede modelizar como Poissoniano, además de introducirse deformaciones en la forma del espectro al emplear la implementación FFT (Fast Fourier Transform) debidas al criterio de asignación de masa y al efecto de aliasing, introducido al considerar condiciones de frontera periódicas en virtud del Principio Cosmológico. La asignación de masa se traducirá en una pérdida de potencia a bajas escalas en referencia a las predicciones teóricas, más acentuada para CIC que para NGP y el aliasing producirá un exceso de potencia cerca de la frecuencia de Nyquist debido a la influencia de los modos de mayor frecuencia, aspecto que será tratado en detalle en el desarrollo teórico.

Para contrastar si los resultados de la simulación son los esperados y si la medición del espectro se ha realizado adecuadamente, se siguen dos procedimientos. en primer lugar, se mide el espectro de potencias empleando la suma directa, ofreciéndonos una medida de este libre del efecto de asignación de masa, lo que permite corroborar la bondad de las correcciones realizadas. A continuación, se emplea la herramienta online CAMB, la cual resuelve las ecuaciones de Boltzmann, ofreciendo un espectro de potencias teórico como solución, el cual podemos ajustar a los parámetros cosmológicos empleados por PKDGRAV3 y que nos permite comprobar que los resultados de la simulación y los teóricos coinciden. Finalmente, se cierra el proyecto con la discusión sobre la evolución temporal del espectro de potencias con el tiempo.

---

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Objectives</b>	<b>3</b>
<b>3</b>	<b>The Standard Cosmological Model</b>	<b>4</b>
<b>4</b>	<b>Statistical analysis of the large scale structure of the Universe</b>	<b>7</b>
4.1	Introduction of mass assignment schemes to measure the fluctuations of density field . . . . .	7
4.2	A model for the probability density function . . . . .	10
4.3	The two-points correlation function . . . . .	12
4.4	Power Spectrum . . . . .	14
4.5	Measurement of Power Spectrum with FFT, the aliasing effect . . . . .	15
<b>5</b>	<b>PKDGRAV3</b>	<b>20</b>
5.1	Simulations . . . . .	21
<b>6</b>	<b>Statistical analysis of the N-body simulation</b>	<b>23</b>
6.1	Difference between NGP and CIC mass assignment . . . . .	23
6.2	The log-normal model for the distribution of matter . . . . .	27
6.3	Measuring the correlation function . . . . .	29
6.4	Measuring the matter Power Spectrum . . . . .	31
<b>7</b>	<b>Conclusion</b>	<b>36</b>
	<b>Bibliography</b>	<b>38</b>
<b>8</b>	<b>Appendix I: Friedmann equations</b>	<b>40</b>
<b>9</b>	<b>Appendix II: Linear Power Spectrum</b>	<b>43</b>



# 1 Introduction

Cosmology is an observational science, i.e., the sample of events that we can observe is unique in the sense that it is not possible to reproduce the same event in the same conditions as many times as we want, we can only observe them in a particular place and at particular time. A galaxy survey has galaxies placed at different cosmological times and then try to reconstruct the complete process joining the different observations of the evolution. The development of supercomputers with high processing and storing capacity provides the possibility to run N-body simulations with large number of particles to follow the evolution of the matter distribution. The mass tracers we choose to develop the calculations are test particles whose properties are similar to dark matter particles, a theoretical type of matter postulated for the first time by Lord Kelvin and Henri Poincaré and inferred by the German astronomer Fritz Zwicky in 1933 from his work on the Coma Cluster. It was introduced to explain experimental observations in which the amount of baryonic matter is not enough to justify the gravitational effects observed, being its principal characteristic that this type of matter must not feel electromagnetic interaction. In the framework of Cosmology, the existence of this kind of matter is essential to be able to build a model that explains the formation of structures at different scales; in other words, without dark matter we cannot explain the formation of galaxies and clusters.

The particles will evolve under the rules of a dynamic scenario, according to the Standard Cosmological Model. The natural law that takes the central role of this study is the gravitational force, formulated by Isaac Newton in 1687, which establishes the interaction between two objects with mass as a central and attractive force whose module is directly proportional to the product of the masses and inversely proportional to the square of the distance between their centers. The Cosmological principle, which is based on large amount of observations over the years and establishes that the distribution of matter at large scales must be homogeneous and isotropic, will be recursive along the development of the project to define the boundary conditions of the simulated Universe and check the properties of the simulation at large scale, which must be in concordance with an homogeneous distribution of matter. Last but not least, space-time is not static, it evolves following the General Relativity theory. Hubble's observations pointed that galaxy velocity increases with the distance and recent observations like the Cosmological Microwave Background (CMB) by Planck Collaboration et al. (2016) or measurements in far supernovas show cosmologists that the Universe is in an accelerated expansion which can be considered isotropic. This created the need to introduce a new component, the dark energy, which can be understood as a negative

pressure that curves the geometry of the space in absence of matter.

The study of the large scale structure of the Universe improves our understanding of the same revealing information about how the gravity force influences the distribution of matter, providing a tool to study the formation of cosmological structures like clusters or halos along the time.

---



## 2 Objectives

The main objectives of this bachelor project were:

- To perform cosmological simulations with different mass resolution using the PKDGRAV3 code in a 100 Mpc/h box with the object of analysing the matter density.
- To characterize the main parameters of the standard cosmological model  $\Lambda$ CMB involved in initial conditions and evolution of the matter distribution.
- To develop and understand the theoretical formalism of one-point and two-point statistics in configuration space and in Fourier space to characterize the distribution in different scales and at different instances.
- To implement and discuss the difference NGP and CIC mass assignment methods to filter the discrete distribution of matter.
- Measure the density probability function of matter distribution and study the evolution of early moments with the redshift.
- To measure correlation function for a low resolution simulation using different estimator methods and study its evolution.
- To measure power spectrum for low resolution simulations and compare simulation results with theoretical models.

# 3 The Standard Cosmological Model

*El presente trabajo se desarrollará en el marco del modelo cosmológico estándar,  $\Lambda$ CDM, el cual se sustenta la Teoría de la Relatividad General y ciertas consideraciones en tensor energía-momentum, así como en la consideración de un espacio homogéneo e isótropo, describe de forma satisfactoria las observaciones conocidos seis parámetros: la densidad bariónica, la densidad de materia oscura (CDM), la edad del universo, el índice espectral, la fluctuación en la amplitud de la curvatura y la profundidad óptica de reionización.*

We first introduce the theoretical framework within which this study has been conducted. The development of the General Relativity theory (Weinberg, 1972; Ridler, 1977) to explain the gravity interaction and the advances on observational cosmology allow us to construct a cosmological model to describe the first instants, the evolution and the fate of the Universe. The most accepted model is  $\Lambda$ CDM, which is mainly based in two concepts: the cosmological constant  $\Lambda$  is a fundamental constant which explains the accelerated expansion of the Universe introducing a vacuum energy, which we refer as dark energy, and the Cold Dark Matter (CDM, Planck Collaboration et al., 2016; Dodelson, 2003; Peebles, 1980), a type of dark matter whose velocity is well below the relativistic limit, necessary condition to structure formation. Some of the main assumptions of the Standard Cosmological Model are that the correct theory to describe the gravity interaction on cosmological scales is General Relativity, and to consider that the Universe is approximately, statistically homogeneous and isotropic, apart from spatially flat, on large scales. It is considered also that initial conditions were Gaussian, i.e., the gravitational interaction had no influence over the initial matter distribution, thus it follows a random distribution. For simplicity, dark matter is collisionless, and electromagnetic forces are negligible at large scales for baryonic matter.

The Universe is in expansion and this expansion only depends on time, being the same for all the points of the space. Therefore, it is useful to express the proper coordinates as a function of the scale factor,  $a(t)$ , and the comoving coordinates,  $x(t)$ , (Padmanabhan, 1993; Dodelson, 2003; Cepa, 2007):

$$\mathbf{r}(t) = a(t)\mathbf{x}(t) \tag{3.1}$$

where  $\mathbf{r}(t)$ , the new vector of coordinates, owns to physical space.

We begin from Einstein field equations, which establish a relation between the space

---

geometry and the amount of matter and energy that it contains by the equation (Weinberg, 1972; Cepa, 2007; Pérez Hernández, 2018; Ridler, 1977):

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 8\pi T_{\mu\nu} - \Lambda g_{\mu\nu} \quad (3.2)$$

where  $R_{\mu\nu}$  is the Ricci curvature tensor,  $R$  is the Ricci scalar,  $g_{\mu\nu}$  is the metric tensor and  $T_{\mu\nu}$  is the stress-energy tensor. The gravitational constant,  $G$ , and the speed of light in vacuum,  $c$ , are taken in natural scale where both are equal to one. The right side of the equation contains the geometrical terms, whose solution we are interested in and, the left side, the sources of field that we have to introduce to solve the equations. The Standard Cosmological Model characterizes matter and energy distribution by a set of parameters which we will explain at the end of the section. Considering cosmological parameters from  $\Lambda$ CDM (Planck Collaboration et al., 2016; Dodelson, 2003) and imposing an homogeneous and isotropic space, we can introduce an ansatz for the metric and obtain a solution for Einstein field equations, which gives as a result the set of equations that governs the evolution of a universe with Friedmann-Lemaître-Robertson-Walker metric, Friedmann equations (Cepa, 2007; Padmanabhan, 1993; Peebles, 1980; Weinberg, 1972):

$$\begin{aligned} H^2 &= \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3}\rho - \frac{k}{a^2} + \frac{\Lambda}{3}, \\ \left(\frac{\ddot{a}}{a}\right) &= -\frac{4\pi}{3}(\rho + 3p) + \frac{\Lambda}{3} \end{aligned} \quad (3.3)$$

The expansion rate of the Universe  $H(t)$  is defined by the first equation as a function of the scale factor,  $a(t)$ , and of the density of matter  $\rho$  considered in the model. The parameter  $k$  is a constant of integration, and is not determined by the theory. Its value has relevant implications in the curvature factor, and its sign determines the type of geometry of the universe, being a de Sitter space if  $k > 0$ , anti de Sitter if  $k < 0$  and a Minkowski space for  $k = 0$ . The second equation quantifies the accelerated expansion, which depends on the matter density and on the pressure field  $P$ . In addition, to solve the above equations and obtain the scale factor as a function of time, we need to add another relation which relates pressure and density, the state equation. For dark matter fluid, this equation is given by  $P = 0$ . The full calculation is developed with more detail in Appendix I (8).

Another interesting parameter that can be defined from the Friedmann equations is the critical density of matter,  $\rho_{crit}$ , the limit from matter starts to collapse, whose expression is given in terms of two constants:

$$\rho_{crit} = \frac{3H_0^2}{8\pi G} \quad (3.4)$$


---

In order to give a description of matter distribution of the Universe, it is necessary to measure the values of cosmological parameters required by the model. These are inferred experimentally from observations realized by (Planck Collaboration et al., 2016), which measured the anisotropy of temperature and polarization of the Cosmic Microwave Background (CMB) radiation. Considering the Hubble reduced parameter  $h = 0.73$ , defined as the value of the expansion rate at present days normalized to an expansion rate of  $100 \text{ km s}^{-1} \text{ Mpc}^{-1}$ , and  $\Omega$ , the fraction of the matter density relative to the critical density, the  $\Lambda$ CDM model can be described by six parameters (Planck Collaboration et al., 2016):

- the baryon fraction  $\Omega_b h^2 = 0.02227 \pm 0.00020$ ;
  - the CDM fraction  $\Omega_c h^2 = 0.1184 \pm 0.0012$ ;
  - the age of the Universe  $t_0 = 13.799 \pm 0.021 \text{ Gyr}$ ;
  - the scalar spectral index  $n_s = 0.9667 \pm 0.0040$ , which describes how density fluctuations vary with the scale;
  - the curvature fluctuation amplitude  $\Delta_R^2 = 2.441_{-0.092}^{+0.088} \times 10^{-9}$  which measures the deviation from flatness;
  - the reionization optical depth  $\tau = 0.066 \pm 0.012$  which offers a measure of the line-of-sight free-electron opacity to CMB radiation (Griffiths et al., 1999).
-

## 4 Statistical analysis of the large scale structure of the Universe

*El primer objetivo del proyecto será caracterizar la distribución estadística de las fluctuaciones de la densidad en distintos momentos de la evolución del Universo. Para la aplicación de un formalismo analítico en la modelización, se introducirán dos funciones filtro correspondientes a los criterios NGP y CIC, para esparcir la información asociada a la distribución de materia, generando un muestreo cuasicontinuo de los valores de estas fluctuaciones. A continuación, se plantea el modelo lognormal para caracterizar la distribución estadística a un punto y se dan argumentos que justifican su uso. Para caracterizar la influencia de la gravedad en la formación de estructuras a distintas escalas, se realiza un estudio de estadística a dos puntos realizando una estimación de la función de correlación empleando el método de Landy-Szalay. En virtud de las propiedades del espacio de Fourier, se estudiará el espectro de potencia de materia que no es sino la transformada de Fourier de la función de correlación, mucho más conveniente computacionalmente. Se pondrán a discusión las diferentes correcciones a esta metodología o ralcionadas con el shot-noise, el efecto de asignación de masa y aliasing derivados de la implementación de la transformada rápida de Fourier (FFT) para la medición.*

### 4.1 Introduction of mass assignment schemes to measure the fluctuations of density field

After characterizing the main aspects that governs the evolution of the Universe, it is time to analyse the dark matter behaviour under the conditions imposed by the model. Assuming a primordial matter distribution at redshift 49, we will study the evolution of it in the context of a N-body problem, defining a cubic volume in a comoving space as the stage where we will put a set of point masses in, interacting under the universal gravity law.

Firstly, we define the main concept of the analysis, the overdensity distribution, which describes the fluctuations of the density field in relation to its average value. In this description of the field, it is expected that in regions whose density value is higher than the critical density of the Universe, the matter will begin to collapse, so we can identify this places with positive values in overdensity field. On the other hand, the regions with negative values will be related to regions in space that are 'drained', i.e.,

the matter in these regions will tend to flow to regions with higher density because of the attractive nature of gravity. The critical density of Universe depends on the parameters of the Cosmological model used and, in case of a N-body simulation, as all the test particles have the same mass, this will be assigned in concordance with the critical density, so the average density which depends on the number of particles contained in the total volume will play the same roll. The continuous overdensity field is defined as (Peebles, 1980; Padmanabhan, 1993; Jing, 2005; Cui et al., 2008):

$$\delta(\mathbf{r}) = \frac{\rho(\mathbf{r}) - \bar{\rho}}{\bar{\rho}}. \quad (4.1)$$

The Universe we observe can be considered as point-process (e.g. a Poisson realization) of an underlying continuous density field. For simplicity, we consider that all the particles have the same mass, so we can define the overdensity in terms of the number of particles per unit of volume like:

$$\delta(\mathbf{r}) = \frac{n(\mathbf{r}) - \bar{n}}{\bar{n}}. \quad (4.2)$$

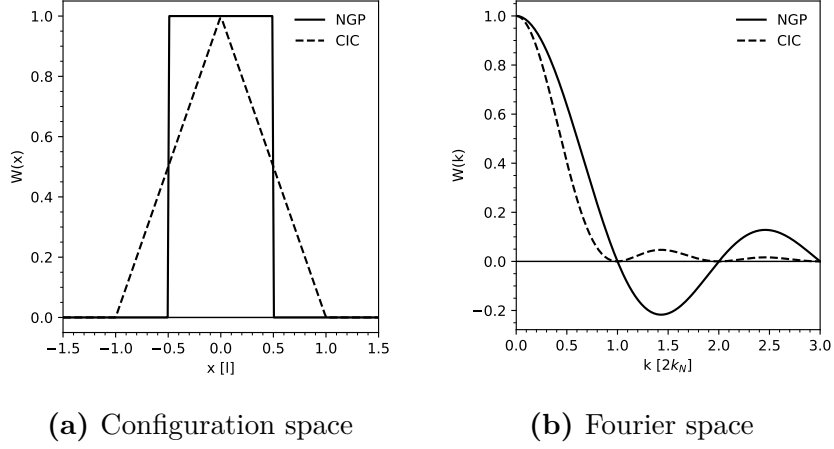
To make an analytical discussion of the problem, it is convenient to define the overdensity function as a continuum. However, in real cases, it would be describe as a collection of Dirac-deltas attending to its discrete nature. In order to apply a continuous analytical formalism, we will smooth the discrete sampling of particles introducing a grid to divide the total volume of our simulation in cubic cells and then follow a scheme to assign mass to each one, assuming the compromise of lost part of the information in favour of spreading it. Mathematically, the transformation is described as a convolution product (Peebles, 1980; Padmanabhan, 1993; Jing, 2005; Cui et al., 2008):

$$\delta(\mathbf{r}_g) = \int \delta(\mathbf{r})W(\mathbf{r} - \mathbf{r}_g)d\mathbf{r}. \quad (4.3)$$

In the expression above, we introduce the vector  $\mathbf{r}_g$  which takes the values of the centers of the cells in the discrete configuration space, being  $g = \frac{L}{N_b}$  the minimum distance between two centers, which is the minimum module of  $\mathbf{r}_g$ , where  $L$  is the size of the box and  $N_b$  is the number of cells along each axis (Cui et al., 2008).

In particular, we will introduce the most usual schemes to make the assignment doing a brief description for one-dimensional cells through the mass assignment function or window function, but the result can be generalized as a product of the values of this function for each axis (Hockney & Eastwood, 1981; Cui et al., 2008; Jing, 2005).

- **Nearest Grid Point (NGP):** This is the simplest method to assign mass to each cell. Conceptually, we associate the information of the particle to the cell



**Figure 4.1:** Window functions in configuration space and in Fourier space, where x-axis values are expressed in terms of  $l$  and k-axis in units proportional to Nyquist frequency. The information contained in a certain range in configuration space is spread throughout all Fourier space.

whose center is closer to it. NGP window function can be written like:

$$W(x) = \begin{cases} 1, & \text{if } |x| < \frac{g}{2} \\ 0, & \text{if } |x| > \frac{g}{2} \end{cases} \quad (4.4)$$

- **Cloud in Cell (CIC):** This scheme spreads the information of the localization of a particle over the nearest cell and over the closest cells to this. To consider the relative orientation of the particle inside a cell, the window function is defined in such a way that only the two closest cells are assigned a fraction of the mass (in one-dimensional description), being eight cells in the case of a three-dimensional space. CIC window function can be written like:

$$W(x) = \begin{cases} 1 - |x|, & \text{if } |x| < g \\ 0, & \text{if } |x| > g \end{cases} \quad (4.5)$$

We could have chosen any other definition for the CIC window function but, considering the convolution theorem, eq. (4.3) can be rewritten as a product in Fourier space, the Fourier transform of the window function can be generalised as (Cui et al., 2008):

$$W(k_x) = \left( \frac{\sin \frac{\pi k_x}{2k_N}}{\frac{\pi k_x}{2k_N}} \right)^p \quad \begin{cases} p = 1, & \text{NGP} \\ p = 2, & \text{CIC} \end{cases} \quad (4.6)$$

These functions depend on the Nyquist frequency  $k_N = \frac{\pi}{l}$  (Nyquist, 1928), which is

---

defined as the half of the maximum frequency that satisfies the boundary conditions of the considered volume and the spacial resolution of configuration space.

## 4.2 A model for the probability density function

Once the overdensity distribution is obtained, we wonder about what is the best mathematical description of the density fluctuations. The lognormal model is proposed to fit the distribution of density fluctuations in a continuous approach and will be discussed in detail. Before continuing with theoretical development, it is important to explain why we are theoretically motivated to use the lognormal model to reproduce the PDF (Probability density Function) of the overdensity random field (Jones, 1990).

On one side, the observations performed by Hubble (1934) led him to notice that the number of galaxies in a two-dimensional grid on the sky follows a distribution like (4.14). Recent cosmological simulations show that this model is a reasonable approximation for a continuous density field, but it is necessary to improve it to alleviate the bias introduced by the discreteness of the sampling (Jones, 1990).

On the other side, for large scales we can consider linear fluctuations in density field that correspond to a Gaussian distribution, but this approach is not suitable for small scales, where clustering processes are important and a non linear noise is expected. Inflation models (Barrow & Coles, 1990) consider that at the beginning of the Universe, the fluctuations on the density field were Gaussian, which implies that Fourier modes were independent but, in the pass of time, the effect of gravity force caused that these modes become to couple, mostly in small scales, where its influence is bigger. This analysis invites us to consider that the main fluctuations depend on the product of local random phases of Fourier space as:

$$Y = \prod_i X_i$$

In addition, if we consider the Central Limit Theorem, taking natural logarithms in both sides of the expression, we will find that:

$$\ln Y = \sum_i \ln X_i \quad (4.7)$$

According to this theorem, the variable  $\ln Y$  follows a Gaussian distribution, which implies that  $Y$  must follow a lognormal distribution.

If we put our attention in the dynamic evolution of the system, considering dark matter as an ideal fluid, the mass conservation will be given by the continuity equation (see more details in Appendix I 8). If we take a comoving reference system, we will obtain the relation:

$$\frac{1}{\rho} \frac{d\rho}{d\tau} = -\nabla \cdot \mathbf{v} \quad (4.8)$$



Considering a initial Gaussian peculiar velocity field whose divergence scales with  $\tau$ , following the law  $(\nabla \mathbf{v})_\tau = (\nabla \mathbf{v})_{\tau_0} \frac{\tau}{\tau_0}$ , we can find a solution for the equation above like:

$$\rho(\mathbf{x}) = \rho_0 e^{\epsilon(\mathbf{x})\tau^2} \quad (4.9)$$

where  $\epsilon(\mathbf{x})$  is a Gaussian field, obtaining that the density field follows a lognormal distribution.

Another indicator to use a lognormal model is the study of biased model for galaxy or cluster formation, which suggests a exponential relation between the number of galaxies per unit of mass inside and outside of superclusters (Jones, 1990).

Last but not least, the lognormal model is a simple way to characterize the non-linear fluctuations, whose sampling spectrum is defined positive and whose behaviour can be described analytically, having the same free parameters as Gaussian random field.

It is important to note that, considering the Moment Generation Function  $M(s)$  defined by the expression:

$$M(s) = \sum_0^{\infty} \frac{\mu'_n s^n}{n!} \quad (4.10)$$

where  $\mu'_n$  is the moment  $n$  of the distribution, leads to a divergence for positive values of  $s$  for this kind of PDF. As a result, lognormal distribution is not completely determined by its moments, so one can construct at least another different distribution with the same moments. Physically, we can understand this non-uniqueness property as consequences of the influence of high-density regions on high-order moments (Jones, 1990). To treat strong clustering regions, it is possible to apply Fry's hierarchical distributions (Fry, 1985) which are completely characterized by their moments, but the question of the non-uniqueness nature of the matter distribution is still open.

Therefore, we develop in detail the main model aspects. Let us consider a variable  $X(\mathbf{r})$  that follows a normal distribution with PDF (Jones, 1990):

$$P(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (4.11)$$

where  $\mu$  is the main value of  $X(\mathbf{r})$  which we will denote as  $x$  in equations and  $\sigma^2$  is the variance of the distribution.

A variable  $y$  will follow a lognormal distribution if its shape is  $y = e^{X(\mathbf{r})}$ . In order to obtain the PDF of a lognormal distribution, we rewrite (4.11) in terms of  $\ln y$ :

$$P(\ln y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\ln y - \mu)^2}{2\sigma^2}} \quad (4.12)$$

The differential probability of measure a certain value of  $\ln y$  is given by:

$$f(\ln y) d \ln y = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\ln y - \mu)^2}{2\sigma^2}} d \ln y = \frac{1}{\sqrt{2\pi}\sigma y} e^{-\frac{(\ln y - \mu)^2}{2\sigma^2}} dy \quad (4.13)$$

From the last term, we can obtain the expression of the PDF for a lognormal distribution:

$$P(y) = \frac{1}{\sqrt{2\pi\sigma}y} e^{-\frac{(\ln y - \mu)^2}{2\sigma^2}} \quad (4.14)$$

For the overdensity PDF, we will follow the approach of (Hongguang Bi, 2003) to obtain the expression considering physical constraints. If we consider  $y = \frac{\rho}{\rho_0}$ , the mean value of  $y$  will be:

$$\bar{y} = \frac{1}{\sqrt{2\pi\sigma}} \int e^x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \quad (4.15)$$

Developing and reorganising the exponential argument, we can solve the integral in a simple way:

$$\begin{aligned} \bar{y} &= \frac{1}{\sqrt{2\pi\sigma}} \int e^{-\left(\frac{2\sigma^2 x + x^2 - 2x\mu + \mu^2}{2\sigma^2}\right)} dx \\ &= \frac{1}{\sqrt{2\pi\sigma}} \int e^{-\left(\frac{x^2 - 2(\sigma^2 + \mu)x + \mu^2}{2\sigma^2}\right)} dx \\ &= \frac{1}{\sqrt{2\pi\sigma}} \int e^{-\left(\frac{x^2 - 2(\sigma^2 + \mu)x + (\sigma^2 + \mu)^2 - (\sigma_0^2 + \mu)^2 + \mu^2}{2\sigma^2}\right)} dx \\ &= \frac{1}{\sqrt{2\pi\sigma}} \int e^{-(\mu + \frac{\sigma^2}{2})} e^{-\frac{(x - (\sigma^2 + \mu))^2}{2\sigma^2}} dx \\ &= e^{-(\mu + \frac{\sigma^2}{2})} \end{aligned} \quad (4.16)$$

The mass conservation implies that  $\rho$  must satisfy the relation  $\bar{\rho} = \bar{\rho}_0$  and  $\bar{y} = 1$ , so:

$$\mu = -\frac{\sigma^2}{2} \quad (4.17)$$

Substituting this relation in (4.14), we finally obtain that:

$$P\left(\frac{\rho}{\rho_0}\right) = \frac{1}{\sqrt{2\pi\sigma} \frac{\rho}{\rho_0}} e^{-\frac{\left(\ln \frac{\rho}{\rho_0} + \frac{\sigma^2}{2}\right)^2}{2\sigma^2}} \quad (4.18)$$

This equation characterizes the lognormal distribution under physical conditions, which gives us a theoretical tool to predict the expected behaviour of the overdensity density PDF.

### 4.3 The two-points correlation function

The two-points statistics measures the statistical dependence between the values of two points of a field (Peebles, 1980; Dodelson, 2003; Padmanabhan, 1993). Let us consider a

random process whose values can be characterized by one parameter  $r$ . Then, construct a vector whose components are values of the random process shifted by a certain lag in the parameter. Considering the array which contains the values of the random process as a vector, the autocorrelation function is defined as the main value of the normalized scalar product between the two vectors and each value of this is characterized by the lag in the parameter. Values of autocorrelation function close to one are related with a strong linear dependence for processes related by a certain lag, otherwise, for values near zero the points will not be correlated, i.e., the values of the field related by the shift are independent. In general, we will refer to this function as the correlation function, which is very usual in statistical mechanics and, in our case, the lag in the parameter will be the distance between to particles of the sample. We can consider this simplification (in general, this parameter is the vector which connects the particles) because of the condition of isotropy imposed by the cosmological principle.

Let  $X(r)$  be a scalar field, by definition, the differential probability of measuring a certain value of the field in the position  $r_1$  and another certain value in position  $r_2$  is given by (Landy & Szalay, 1993; Peebles, 1980; Dodelson, 2003):

$$\delta P = N^2(1 + \xi_{12}(r))d\mathbf{r}_1d\mathbf{r}_2 \quad (4.19)$$

where the two-points correlation function,  $\xi_{12}(\mathbf{r})$ , introduced in the identity above is defined as:

$$\xi_{12}(r) = \frac{\langle (X_1 - \langle X \rangle)(X_2 - \langle X \rangle) \rangle}{\langle X \rangle^2} \quad (4.20)$$

being  $X_1$  and  $X_2$  two elements of the sampling of values of  $X$  field. If we take the density field, changing  $X_1$  for  $\rho(r_0)$  and  $X_2$  for  $\rho(r + r_0)$ , having in account the definition of overdensity field (4.1), we can write the correlation function as follows:

$$\xi_{12}(r) = \langle \delta(r_0)\delta(r + r_0) \rangle \quad (4.21)$$

where the mean value denotes an integration over  $r_0$ .

Physically, the correlation function can be interpreted as a measurement of the gravity influence on matter distribution at different spatial scales. In cosmology, it is common to define it as the probability of finding a pair of galaxies separated by certain distance  $r$ . This probability is referred to a random distribution in order to highlight the difference between an Universe with and without gravity interaction. The random reference is not included in the mathematical definition of correlation function, so it is useful to develop estimator methods to introduce this information in the definition. In relation with the computational cost, these estimators will allow us to perform the measurement for a large sample in a reasonable time. The estimator methods studied are based on the arrays DD (Data-Data), DR (Data-Random) and RR (Random-Random), whose measurement will be explained with more detail in the analysis section. There are several ways to construct reasonable estimators using these

arrays, but we will focus our attention on those proposed by Landy & Szalay (1993):

$$1 + \xi_{12}(\mathbf{r}) \begin{cases} \frac{DD}{RR} \\ \frac{DD}{DR} \\ \frac{DD + RR - 2DR}{RR} \end{cases} \quad (4.22)$$

The last of the estimators is known as the Landy-Szalay estimator, and is the most used in the literature because it reduces the uncertainty in the measure (Landy & Szalay, 1993).

## 4.4 Power Spectrum

To characterize the two-points statistics of a distribution, the natural way is to calculate the correlation function, but it is not the best way to approach the task. If we apply the Fourier transform over it, we find that it can be expressed as the square of the module of the overdensity function in Fourier space. This function is called Power Spectrum, whose definition is (Padmanabhan, 1993; Peebles, 1980; Jing, 2005):

$$P(k) = \langle |\delta(\mathbf{k})|^2 \rangle \quad (4.23)$$

where  $\langle \dots \rangle$  means the ensemble average. When we developed the correlation function formalism, we directly simplify the dependence on the direction but, considering that we will work over simulation data, we must to deal with the general case in theoretical development and the approach will be introduced later, in the practical section, performing the measurement of the mean value of the function in spherical caps. The Fourier transform of the overdensity is given by:

$$\delta(\mathbf{k}) = \frac{1}{V_\mu} \int_{V_\mu} \delta(\mathbf{r}) e^{-i\mathbf{k}\mathbf{r}} d\mathbf{r} \quad (4.24)$$

where  $V_\mu$  is a large volume in which we can consider that the density field is periodic, according to cosmological principle. As we discussed in section 4, the distribution of matter obtained from the simulation is discrete so, taking the expression (4.2), we define the discrete Fourier Transformation of  $\delta(\mathbf{r})$  as follows:

$$\delta^d(\mathbf{k}) = \frac{1}{V_\mu \bar{n}} \int_{V_\mu} n(\mathbf{r}) e^{-i\mathbf{k}\mathbf{r}} d\mathbf{r} - \delta_{\mathbf{k},\mathbf{0}}^K \quad (4.25)$$

As consequence of using a bound volume, the sample of Fourier modes is finite, so we introduce the Kroneker delta as  $\delta^K$  as the Fourier transformation of a constant. For

simplicity, we will consider that the normalization factor is independent of the definition of the Fourier Transform, considering how definition for the same only the integral, which will simplify the lecture of the development. The distribution of particles can be described as a sample of Dirac deltas such  $n(\mathbf{r}) = \sum_i \delta^D((r) - (r)_i)$ . Developing the operation above, we obtain that:

$$\delta^d(\mathbf{k}) = \frac{1}{N} \sum_i n_i e^{-i\mathbf{k}\mathbf{r}_i} - \delta_{\mathbf{k},\mathbf{0}}^K \quad (4.26)$$

where  $N = V_\mu \bar{n}$  and we have divided the volume  $V_\mu$  in differential elements  $dV_i$ , taking them so small that the number of particles inside of each box will be  $n_i \in \{0, 1\}$  and, as consequence,  $n_i = n_i^2 = n_i^3 = \dots$ . Keeping in mind the fact that  $\langle n_i \rangle = \bar{n} dV_i$  and  $\langle n_i n_j \rangle_{i \neq j} = \bar{n}^2 dV_i dV_j [1 + \langle \delta(\mathbf{r}_i) \delta(\mathbf{r}_j) \rangle]$ , we develop the following calculation (Jing, 2005; Hockney & Eastwood, 1981):

$$\begin{aligned} \langle \delta^d(\mathbf{k}_1) \delta^{d*}(\mathbf{k}_2) \rangle &= \frac{1}{N^2} \sum_{i,j} \langle n_i n_j \rangle e^{-i(\mathbf{r}_i \mathbf{k}_1 - \mathbf{r}_j \mathbf{k}_2)} - \delta_{\mathbf{k}_1, \mathbf{0}}^K \delta_{\mathbf{k}_2, \mathbf{0}}^K \\ &= \frac{1}{N^2} \sum_{i \neq j} \langle n_i n_j \rangle e^{-i(\mathbf{r}_i \mathbf{k}_1 - \mathbf{r}_j \mathbf{k}_2)} + \frac{1}{N^2} \sum_i \langle n_i \rangle e^{-i\mathbf{r}_i(\mathbf{k}_1 - \mathbf{k}_2)} - \delta_{\mathbf{k}_1, \mathbf{0}}^K \delta_{\mathbf{k}_2, \mathbf{0}}^K \\ &= \langle \delta(\mathbf{k}_1) \delta^*(\mathbf{k}_2) \rangle + \frac{1}{N} \delta_{\mathbf{k}_1, \mathbf{k}_2}^K \end{aligned} \quad (4.27)$$

In the last step, we introduce the complex exponential inside the ensemble average and then use the discrete Fourier properties to express  $\langle n_i n_j \rangle$  in Fourier space, neglecting the origin contribution term. This approach was chosen in order to illustrate the dependence on continuous Power Spectrum in the discrete development. Finally, if we take  $\mathbf{k}_1 = \mathbf{k}_2 = \mathbf{k}$ , we have found the expression of the Power Spectrum in terms of the DFT (Jing, 2005; Hockney & Eastwood, 1981; Peebles, 1980):

$$P(k) = \langle |\delta(\mathbf{k})|^2 \rangle - \frac{1}{N} \quad (4.28)$$

The difference between the continuous and discrete Power Spectrum is called shot noise. This term has its origin in the discretization nature of the sampling of particles, being associated to a Poissonian noise.

## 4.5 Measurement of Power Spectrum with FFT, the aliasing effect

It is important to remember that we want to characterize the statistical behaviour of an N-body simulation with N a large number, so the direct summation previously

---

described is not appropriate to measure the Power Spectrum due to the high computational time. The most used method is the Fast Fourier Transform (FFT), which uses window functions with the aim of smoothing the information and reducing the calculations as follows (Jing, 2005; Hockney & Eastwood, 1981):

$$\delta^f(\mathbf{k}) = \frac{1}{N} \sum_{\mathbf{g}} n^f(\mathbf{r}_{\mathbf{g}}) e^{i\mathbf{r}_{\mathbf{g}}\mathbf{k}} - \delta_{\mathbf{k},\mathbf{0}}^K \quad (4.29)$$

where  $n^f(\mathbf{r}_{\mathbf{g}})$  is the convolved density defined in the same way as (4.3). Then, we will apply a process similar to that used before to obtain the contribution of the window function to the Power Spectrum measurement. In theoretical development of this section, we will take  $H = g$  to denote the grid spacing. Following the approach of Jing (2005), we introduce the sampling function  $\Pi(\mathbf{r}) = \sum_{\mathbf{n}} \delta^D(\mathbf{r} - \mathbf{n})$  in the convolution product definition, where  $\mathbf{n}$  is an integer vector:

$$n'(\mathbf{r}) = \Pi\left(\frac{\mathbf{r}}{H}\right) \int n(\mathbf{r}_1) W(\mathbf{r}_1 - \mathbf{r}) d\mathbf{r}_1 \quad (4.30)$$

The introduction of  $H^{-1}$  factor is very important with a view to construct a dimensionless sampling function. If we consider a continuous definition for FFT,  $\delta'^d(\mathbf{r})$ , it is easy to prove that  $\delta'^d(\mathbf{r}) = \delta^d(\mathbf{r})$ :

$$\begin{aligned} \delta'^f(\mathbf{r}) &= \frac{1}{N} \int n'^f(\mathbf{r}) e^{-i\mathbf{r}\mathbf{k}} d\mathbf{r} - \delta_{\mathbf{k},\mathbf{0}}^K \\ &= \frac{1}{N} \int \sum_{\mathbf{g}} \delta^D\left(\frac{\mathbf{r}}{H} - \mathbf{g}\right) \int n(\mathbf{r}_1) W(\mathbf{r}_1 - \mathbf{r}) e^{-i\mathbf{r}\mathbf{k}} d\mathbf{r}_1 d\mathbf{r} - \delta_{\mathbf{k},\mathbf{0}}^K \\ &= \frac{1}{N} \sum_{\mathbf{g}} \int n(\mathbf{r}_1) d\mathbf{r}_1 \int \delta^D\left[\frac{1}{H}(\mathbf{r} - \mathbf{r}_{\mathbf{g}})\right] W(\mathbf{r}_1 - \mathbf{r}) e^{-i\mathbf{r}\mathbf{k}} d\mathbf{r} - \delta_{\mathbf{k},\mathbf{0}}^K \\ &= \frac{1}{N} \sum_{\mathbf{g}} \int n(\mathbf{r}_1) W(\mathbf{r}_{\mathbf{g}} - \mathbf{r}_1) e^{-i\mathbf{r}_{\mathbf{g}}\mathbf{k}} d\mathbf{r}_1 - \delta_{\mathbf{k},\mathbf{0}}^K \\ &= \frac{1}{N} \sum_{\mathbf{g}} n^f(\mathbf{r}_{\mathbf{g}}) e^{-i\mathbf{r}_{\mathbf{g}}\mathbf{k}} - \delta_{\mathbf{k},\mathbf{0}}^K = \delta^f(\mathbf{r}) \end{aligned} \quad (4.31)$$

where  $\mathbf{g}$  is an integer vector.

Before continuing with the main demonstration, we will rewrite the sampling function to show explicitly the origin of aliasing effect. According to its definition, it is easy to see that the sampling function can be described by a sinusoidal function, whose frequency is equal to the double of Nyquist frequency  $2k_N = \frac{2\pi}{H}$  (Nyquist, 1928), but the configuration space is discrete, which implies that this is not the only choice, any frequency that can be written as  $k = nk_N$  is possible too. Therefore, the sampling

function can be read as (Kirchner, 2005):

$$\Pi(\mathbf{r}) = \sum_{\mathbf{n}} c_{\mathbf{n}} e^{-i2k_N \mathbf{n} \mathbf{r}} \quad (4.32)$$

If we measure the coefficients of Fourier Serie, we will obtain that  $c_{\mathbf{n}} = 1$  for  $\Pi\left(\frac{\mathbf{r}_1}{H}\right)$ , so:

$$\Pi\left(\frac{\mathbf{r}_1}{H}\right) = \sum_{\mathbf{n}} e^{-i2k_N \mathbf{n} \mathbf{r}} \quad (4.33)$$

Therefore, we divide the total volume  $V_\mu$  in differential volumes  $dV_i$  as in the previous case, so that the FFT of the overdensity field can be written as:

$$\delta^f(\mathbf{k}) = \frac{1}{N} \int_{V_\mu} \Pi\left(\frac{\mathbf{r}}{H}\right) \sum_i n_i W(\mathbf{r} - \mathbf{r}_i) e^{-i\mathbf{r} \mathbf{k}} d\mathbf{r} - \delta_{\mathbf{k}, \mathbf{0}}^K \quad (4.34)$$

Then, the ensemble average of  $\delta^f(\mathbf{r}_1) \delta^{f*}(\mathbf{r}_2)$  will be:

$$\begin{aligned} \langle \delta^f(\mathbf{r}_1) \delta^{f*}(\mathbf{r}_2) \rangle &= \frac{1}{N^2} \int_{V_\mu} \Pi\left(\frac{\mathbf{r}_1}{H}\right) \Pi\left(\frac{\mathbf{r}_2}{H}\right) \left[ \sum_{i \neq j} \langle n_i n_j \rangle W(\mathbf{r}_{1i}) W(\mathbf{r}_{2j}) + \sum_i \langle n_i \rangle W(\mathbf{r}_{1i}) W(\mathbf{r}_{2i}) \right] \\ &\times e^{-i(\mathbf{r}_1 \mathbf{k}_1 - \mathbf{r}_2 \mathbf{k}_2)} d\mathbf{r}_1 d\mathbf{r}_2 - \frac{1}{N} \int_{V_\mu} \Pi\left(\frac{\mathbf{r}_1}{H}\right) \sum_i \langle n_i \rangle W(\mathbf{r}_{1i}) e^{-i\mathbf{r}_1 \mathbf{k}_1} \delta_{\mathbf{k}_2, \mathbf{0}}^K d\mathbf{r}_1 \\ &- \frac{1}{N} \int_{V_\mu} \Pi\left(\frac{\mathbf{r}_2}{H}\right) \sum_i \langle n_i \rangle W(\mathbf{r}_{2i}) e^{i\mathbf{r}_2 \mathbf{k}_2} \delta_{\mathbf{k}_1, \mathbf{0}}^K d\mathbf{r}_2 + \delta_{\mathbf{k}_1, \mathbf{0}}^K \delta_{\mathbf{k}_2, \mathbf{0}}^K \end{aligned} \quad (4.35)$$

where  $\mathbf{r}_{ik} = \mathbf{r}_i - \mathbf{r}_k$ .

Next step will be to develop the expression term by term. For the first, we obtain:

$$\begin{aligned} &\frac{1}{N^2} \int_{V_\mu} \Pi\left(\frac{\mathbf{r}_1}{H}\right) \Pi\left(\frac{\mathbf{r}_2}{H}\right) \sum_{i \neq j} \langle n_i n_j \rangle W(\mathbf{r}_{1i}) W(\mathbf{r}_{2j}) e^{-i(\mathbf{r}_1 \mathbf{k}_1 - \mathbf{r}_2 \mathbf{k}_2)} d\mathbf{r}_1 d\mathbf{r}_2 = \\ &\sum_{\mathbf{n}_1, \mathbf{n}_2} \sum_{i \neq j} \int_{V_\mu} [1 + \langle \delta(\mathbf{r}_i) \delta(\mathbf{r}_j) \rangle] W(\mathbf{r}_{1i}) W(\mathbf{r}_{2j}) e^{-i(\mathbf{r}_1 \mathbf{k}'_1 - \mathbf{r}_2 \mathbf{k}'_2)} d\mathbf{r}_1 d\mathbf{r}_2 = \\ &\sum_{\mathbf{n}_1, \mathbf{n}_2} \sum_{i \neq j} [1 + \langle \delta(\mathbf{r}_i) \delta(\mathbf{r}_j) \rangle] \int_{V_\mu} W(\mathbf{r}_{1i}) e^{-i\mathbf{r}_1 \mathbf{k}'_1} d\mathbf{r}_1 \int_{V_\mu} W(\mathbf{r}_{2j}) e^{i\mathbf{r}_2 \mathbf{k}'_2} d\mathbf{r}_2 = \quad (4.36) \\ &\sum_{\mathbf{n}_1, \mathbf{n}_2} \sum_{i \neq j} [1 + \langle \delta(\mathbf{r}_i) \delta(\mathbf{r}_j) \rangle] e^{-i(\mathbf{r}_i \mathbf{k}'_1 - \mathbf{r}_j \mathbf{k}'_2)} W(\mathbf{k}'_1) W^*(\mathbf{k}'_2) = \\ &\sum_{\mathbf{n}_1, \mathbf{n}_2} [\delta_{\mathbf{k}'_1, \mathbf{0}}^K \delta_{\mathbf{k}'_2, \mathbf{0}}^K + \langle \delta(\mathbf{k}_1) \delta^*(\mathbf{k}_2) \rangle] \delta_{\mathbf{k}'_1, \mathbf{k}'_2}^K |W(\mathbf{k}'_1)|^2 \end{aligned}$$

where  $\mathbf{k}' = \mathbf{k} + 2k_N \mathbf{n}$ . Following the development of this term, we obtain for the second:

$$\begin{aligned}
\frac{1}{N^2} \int_{V_\mu} \Pi\left(\frac{\mathbf{r}_1}{H}\right) \Pi\left(\frac{\mathbf{r}_2}{H}\right) \sum_i \langle n_i \rangle W(\mathbf{r}_{1i}) W(\mathbf{r}_{2i}) e^{-i(\mathbf{r}_1 \mathbf{k}_1 - \mathbf{r}_2 \mathbf{k}_2)} d\mathbf{r}_1 d\mathbf{r}_2 = \\
\frac{1}{N} \sum_{\mathbf{n}_1, \mathbf{n}_2} \sum_i \int_{V_\mu} W(\mathbf{r}_{1i}) W(\mathbf{r}_{2i}) e^{-i(\mathbf{r}_1 \mathbf{k}'_1 - \mathbf{r}_2 \mathbf{k}'_2)} d\mathbf{r}_1 d\mathbf{r}_2 = \\
\frac{1}{N} \sum_{\mathbf{n}_1, \mathbf{n}_2} \delta_{\mathbf{k}'_1, \mathbf{k}'_2}^K |W(\mathbf{k}'_1)|^2
\end{aligned} \tag{4.37}$$

We obtain the third and the fourth terms as:

$$\begin{aligned}
\frac{1}{N} \int_{V_\mu} \Pi\left(\frac{\mathbf{r}_1}{H}\right) \sum_i \langle n_i \rangle W(\mathbf{r}_{1i}) e^{-i\mathbf{r}_1 \mathbf{k}_1} \delta_{\mathbf{k}_2, \mathbf{0}}^K d\mathbf{r}_1 = \\
\delta_{\mathbf{k}_2, \mathbf{0}}^K \sum_{\mathbf{n}_1, \mathbf{n}_2} \sum_i \int_{V_\mu} W(\mathbf{r}_{1i}) e^{-i\mathbf{r}_1 \mathbf{k}'_1} d\mathbf{r}_1 = \\
\delta_{\mathbf{k}_2, \mathbf{0}}^K \sum_{\mathbf{n}_1} \delta_{\mathbf{k}'_1, \mathbf{0}}^K W(\mathbf{k}'_1)
\end{aligned} \tag{4.38}$$

From the results (4.36), (4.37) and (4.38), we found:

$$\begin{aligned}
\langle \delta^f(\mathbf{r}_1) \delta^{f*}(\mathbf{r}_2) \rangle = \sum_{\mathbf{n}_1, \mathbf{n}_2} [\langle \delta(\mathbf{k}_1) \delta^*(\mathbf{k}_2) \rangle \delta_{\mathbf{k}'_1, \mathbf{k}'_2}^K |W(\mathbf{k}'_1)|^2 + \frac{1}{N} \delta_{\mathbf{k}'_1, \mathbf{k}'_2}^K |W(\mathbf{k}'_1)|^2 \\
- \delta_{\mathbf{k}'_1, \mathbf{0}}^K \delta_{\mathbf{k}'_2, \mathbf{0}}^K |W(\mathbf{k}'_1 = \mathbf{0})|^2 - \delta_{\mathbf{k}'_1, \mathbf{0}}^K \delta_{\mathbf{k}'_2, \mathbf{0}}^K W(\mathbf{k}'_1 = \mathbf{0}) - \delta_{\mathbf{k}'_1, \mathbf{0}}^K \delta_{\mathbf{k}'_2, \mathbf{0}}^K W^*(\mathbf{k}'_2 = \mathbf{0}) + \delta_{\mathbf{k}_1, \mathbf{0}}^K \delta_{\mathbf{k}_2, \mathbf{0}}^K]
\end{aligned} \tag{4.39}$$

According to the Fourier Transform definition,  $W(\mathbf{k}' = \mathbf{0})$  can be read as:

$$W(\mathbf{k}' = \mathbf{0}) = \int W(\mathbf{r}) d\mathbf{r} \tag{4.40}$$

The window function is normalized by definition, which means that  $W(\mathbf{k}' = \mathbf{0}) = 1$ , so:

$$\langle \delta^f(\mathbf{r}_1) \delta^{f*}(\mathbf{r}_2) \rangle = \sum_{\mathbf{n}_1, \mathbf{n}_2} \delta_{\mathbf{k}'_1, \mathbf{k}'_2}^K |W(\mathbf{k}'_1)|^2 P(\mathbf{k}'_1) + \frac{1}{N} \delta_{\mathbf{k}'_1, \mathbf{k}'_2}^K |W(\mathbf{k}'_1)|^2 \tag{4.41}$$

Finally, for  $\mathbf{k}_1 = \mathbf{k}_2 = \mathbf{k}$ , we conclude that (Cui et al., 2008; Jing, 2005):

$$\langle |\delta^f(\mathbf{k})|^2 \rangle = \sum_{\mathbf{n}} [|W(\mathbf{k}'_1)|^2 P(\mathbf{k}') + \frac{1}{N} |W(\mathbf{k}')|^2] \tag{4.42}$$

This expression shows how mass assignment function contributes to Power Spectrum



and shot noise and how the discreteness of configuration space and periodic boundary conditions introduces the alias sums.

---

## 5 PKDGRAV3

*Para iniciar la parte práctica del trabajo, se compila el código de simulación de  $N$ -cuerpos PKDGRAV3 y se corren simulaciones de distinta resolución espacial en un volumen cúbico de lado 100 Mpc/h desde un redshift inicial 49 hasta un redshift a tiempo actual almacenando diez pasos temporales y se muestra en un mapa de color la distribución de densidades obtenidas a distintos redshift en distintas regiones del volumen.*

The first step of the practical component of the bachelor project was to generate the sampling of particles of our particular Universe in different time instants using PKDGRAV3 <sup>1</sup> (Douglas Potter, 2016), a parallel code to perform  $N$ -body simulations.

The number of particles used in  $N$ -body simulations has grown enormously in order to improve the accuracy of the results and better compare with observations. Initially, codes calculated gravity force using the particle-particle (PP) methods, implementation whose computational time scales with  $O(N^2)$ . Therefore, for large amount of particles it was impossible to perform simulations. Codes that use particle-mesh (PM) methods based on FFT, allow to reduce computational time, scaling with  $O(N \ln N)$ , at the cost of resolution for the same number of particles. PKDGRAV3 uses the Fast Multipole Method (FMM) to perform the interaction calculation, computational time scales with  $O(N)$  and makes possible to run simulations with trillions of particles.

The FMM method consists on approximating interaction forces with a multipolar expansion. It makes use of PP method to calculate with high accuracy the interaction with the nearest particles and, taking into account that the weight of the terms of the expansion decreases with the distance, approximates the large-range forces by cell-cell (CC) method. This process can be performed by a tree-code, a binary tree-code in case of PKDGRAV3 where parent cells, i.e., the closest cells to the base of the tree, are divided along the longest axis into two equal volume child cells.

Another interesting characteristic of PKDGRAV3 is that it generates the initial conditions at initial redshift, defined by a set of cosmological parameters, in contrast with most cosmological codes that need to use auxiliary codes to obtain them. Apart from  $N$ -body simulations, PKDGRAV3 can also perform hydrodynamical simulations, but this is not the object of our study.

---

<sup>1</sup>The code is publicly available at <https://bitbucket.org/dpotter/pkdgrav3/src/master/>

## 5.1 Simulations

PKDGRAV3 was compiled on a MacBook Air by the student and then was used to perform four simulations with the same volume but different number of particles. The set of the initial parameters of the simulation are listed in table 5.1.

Parameter	Value
L	100 Mpc/h
$z_{in}$	49
h	0.67
$\Omega_0$	0.32
$\Omega_\Lambda$	0.68
$\sigma_8$	0.83
$n_s$	0.96

**Table 5.1:** Set of cosmological parameters used in simulations.

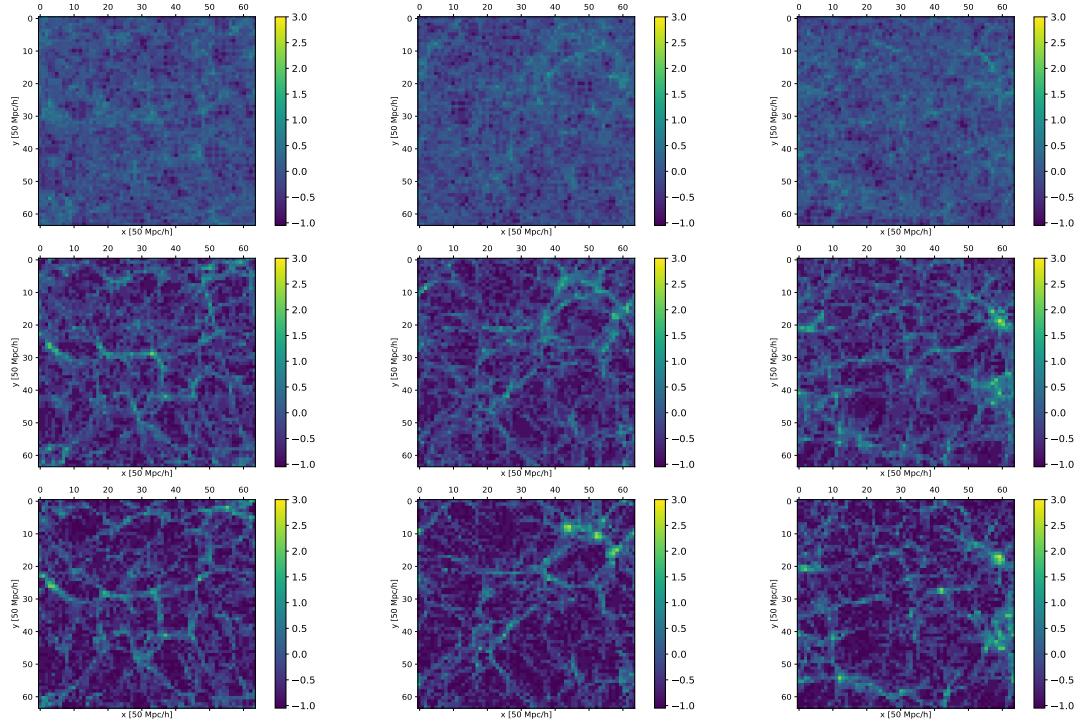
The initial redshift of the simulation is 49, saving matter distribution properties like the positions of the particles for ten time steps. The file format used in this work is HDF5<sup>2</sup>.

The first simulation was run with  $N = 128^3$ , a bit more than 2 million particles. It was considered that this number was reasonable to have a statistical sampling sufficiently large, taking into account the computational limitations of the laptop. The second simulation was performed with 8 million particles in order to increase the spatial resolution, given by the inverse of the cubic root of the number of particles, if the volume of the box simulation is normalized. In theoretical sections, we will discuss about the excessive computational cost of the correlation function and the direct sum of Fourier Transform measurements, so we performed other two simulations with less resolution using over  $N = 32^3$  and  $N = 64^3$  particles respectively to be able to measure them in a reasonable period of time.

Before describing the statistical analysis of the simulations, we plot the distribution of overdensity obtained from different regions of the volume and at different times (Figure 5.1). In this way, we display the formation of structures, paying attention to the yellow filaments of the colour maps which are associated with large matter concentration.

---

<sup>2</sup>The official website of HDF5 is <http://www.h5py.org>



**Figure 5.1:** Colour maps of  $\ln(\delta + 1)$  for the simulation with  $N = 128^3$ . On columns, we show three slices along  $z$  axis, where the central column is associated with the center of the volume and the other columns show symmetric regions in relation to the center. On rows, the time of the slices varies from the first redshift saved to actual time, passing through an intermediate epoch. To measure the overdensity function, a  $64^3$  grid was used.

# 6 Statistical analysis of the N-body simulation

*En la presente sección se detallarán las mediciones estadísticas realizadas con la distribución de partículas de la simulación de  $N$  cuerpos. En primer lugar se introduce la implementación de los esquemas de interpolación en tres dimensiones NGP y CIC así como sus efectos sobre los histogramas medidos. Se discute el modelo log-normal para la distribución de densidades de la materia y su comportamiento como función del tiempo cosmológico. Se observa el aglutinamiento de materia y su efecto en la distribución de las densidades en el volumen cosmológico simulado. A continuación se detalla la metodología seguida para medir la función de correlación de dos puntos, usando estimadores basados en la distribución de pares de objetos en la simulación. Finalmente, se mide el espectro de potencias de la distribución empleando las funciones de FFT implementadas en Python, discutiéndose los efectos de shot noise, asignación de masa y aliasing empleando los esquemas NGP y CIC y se comparan con la medida ofrecida por la DFT (Discrete Fourier Transform). Se realiza una verificación de que la simulación se ejecuta correctamente al comparar los espectros medidos con los teóricos ofrecidos por CAMB y se muestra la evolución temporal del espectro de potencia observando su desviación del comportamiento lineal conforme avanza el tiempo, describiendo el comportamiento esperado.*

## 6.1 Difference between NGP and CIC mass assignment

Once we have obtained the results of the simulation, we start the practical part of the project analysing the statistical properties of the data using the NGP and CIC mass assignment schemes. In order to measure the overdensity function, we divide the original volume of the simulation (100 Mpc/h) in  $64^3$  cubes because the FFT algorithm that we will apply later uses a radix-2 implementation (Cooley & Turkey, 1965) which scales with  $N \log_2 N$  and it was considered that this amount is enough to bin the space and perform the calculation in a reasonable time.

The most simple scheme is the NGP method, which is easy to implement, reading the idea behind the window function as an histogram, so one only have to define the limits of each bin, i.e., the cubes in which the whole space was divided and then count how many particles are inside each one. It is not only the easiest method to implement in Python, but also the most efficient because the function histogram2D of the Numpy

library allows to construct a 2D histogram of a data sample, so it is enough to use a single loop along one of the axis to complete the operation.

On the other side, the CIC algorithm is a bit more complicated to construct. The main scheme of the code developed by the student is the following: first, a particle is selected and the index of the box that contains it is obtained; then, we measure the window function associated with its position, obtaining the index of the cells over the information will be spread as well as the contribution of the particle to each one, and finally, the result is added to the whole space matrix. After performing a loop over all particles, the mean density of particles is introduced. The code was implemented as follows:

```

1 '''CIC window function analysis'''
2 #Matrix needed in W_CIC
3 bin7=np.array↵
   ↵ ([ [0,0,1], [0,1,0], [0,1,1], [1,0,0], [1,0,1], [1,1,0], [1,1,1] ])
4 Zeros=np.zeros((7,3))
5
6 def W_CIC(r,index):
7     #Calculate the position of the center of the cell and
8     #the relative orientation of the particle in the box
9     rc=(index+0.5)*2*l/bins-1
10    sg=np.sign(r-rc)
11
12    #Create Index matrix of the relevant cells
13    Index=index+Zeros
14    Index+=sg*bin7
15
16    #Degenerate cases
17    if 0 in sg:
18        Index=Index[(bin7[:,sg==0]==1).flatten()]
19
20    #Add the first cell
21    Index=np.vstack((index,Index))
22
23    #Calculate the position of the center of the other cells
24    Rc=(Index+0.5)*2*l/bins-1
25
26    #Get W values
27    Rr=abs(r-Rc)*bins/(2.*l)
28    W=1-Rr
29    W=W[:,0]*W[:,1]*W[:,2]
30
31    #Correct Index
32    Index[Index==bins]*=0

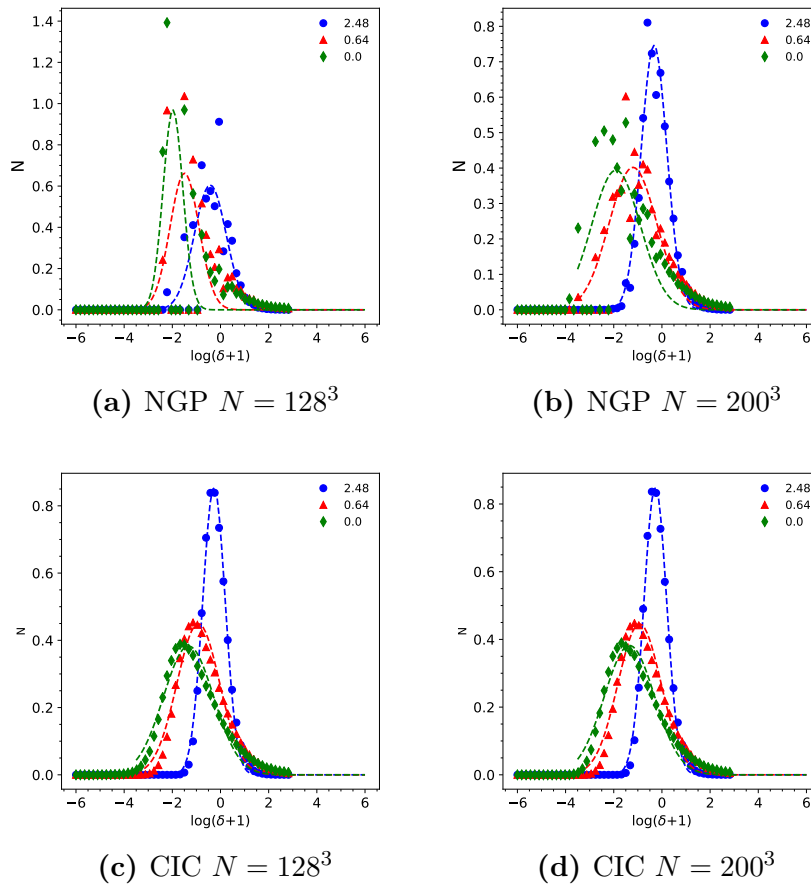
```

```

33     Index=Index.astype(int)
34     return Index,W
35
36 #CIC implementation weighting particle by particle
37 def CIC(F):
38
39     F_Overd=[]
40     V_norm=cuad*(bins)**(-3)#Normalize volume
41
42     for n in range(len(F)):
43         Overd=np.zeros((bins,bins,bins))
44         R=np.array(F[n]['dark']['position'])#Take position array
45
46         for Rn in R:
47             #Obtain indexes. Then, to correct extremal cases.
48             index=(Rn+1)*bins/(2*1)
49             index[index==bins]-=1
50             index=index.astype(int)
51
52             #Obtain the values of W to put in Overd and indexes of the ←→
53             ↔ cells to fill
54             I,W=W_CIC(Rn,index)
55
56             #Fill Overd matrix
57             for i, w in zip(I, W):
58                 Overd[i[0],i[1],i[2]]+=w
59
60             #Overdensity values:
61             Overd/=len(R[:,0])*V_norm
62             Overd-=1.
63             #Storing
64             F_Overd.append(Overd)
65
66     return F_Overd

```

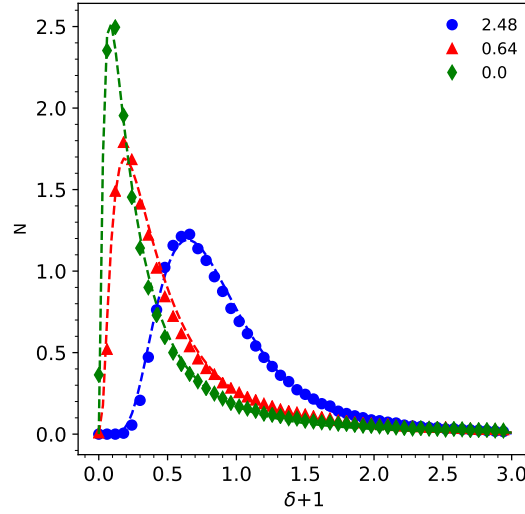
This CIC scheme assigns mass to the cells neighbouring the particle's cell. Therefore, in accordance with the finite dimensions of the simulated volume, periodic boundary conditions must be introduced to measure correctly the values of the window function in the boundary cells. This is introduced in the code changing the box index with value *bins* (i.e., the number of bins, which will be the index of the following box of the last box) for 0, getting that the adjacent cell in the right side of the last cell was the first for each axis. The same correction should be done in the other side of the array of cells but, by default, Python considers the index -1 is associated to the last position of an array.



**Figure 6.1:** Histograms obtained for  $\ln(\delta + 1)$  for different epochs applying NGP and CIC mass assignment schemes for the largest simulations. A Gaussian fit was done to the non-zero values.

The most complex part of the code was the design of the window function. First, we constructed the matrix that contains the indexes of the cells into which we distribute the mass of the selected particle. To determine these cells, one puts a reference system in the center of the cell of the selected particle and obtain the signs of the coordinates of the particle's relative position vector to this system. This way, we can locate the particle in one of the octants in which this system divides the box. The next step will be to construct an array that contains the indexes of the cells closer to the first octant and then to multiply the coordinates of the vectors by the sign obtained for each coordinate in the step before. Now, we have the indexes of the desired boxes for the general case but, although it is highly unlikely, in order to avoid troubles in the performance, degenerate cases must be treated. The sinus function is zero if the argument is zero, therefore, we only have to exclude the indexes that imply a movement along the axis where the particle coordinates match with the box coordinates using the lines specified





**Figure 6.2:** The overdensity PDF, using the log-normal model to fit the data of the simulation with  $N = 200^3$ .

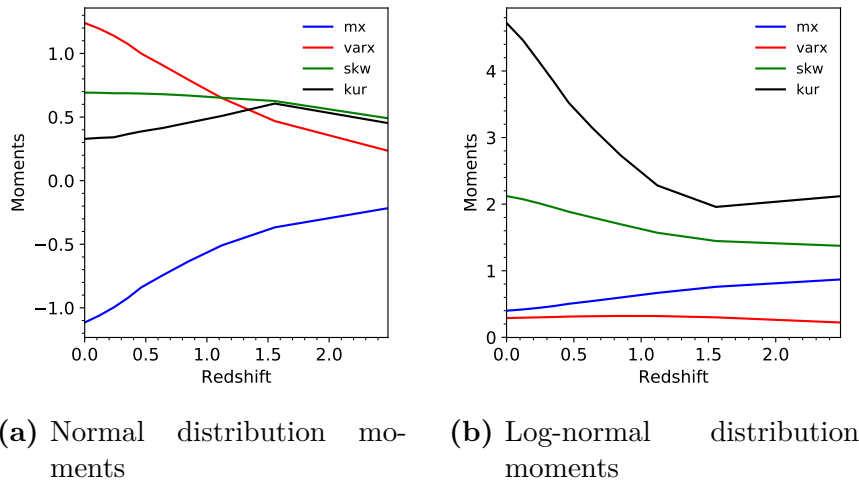
in the code. Then, we evaluate the window function and the implementation of the same will be completed.

Both mass assignment schemes were applied over the largest simulations, Figure 6.1, in order to sample the matter density field on a 3-dimensional Cartesian grid. In case of NGP, we found that the low densities sampling is rather noisy. The reason is the discrete values that overdensity function can take according to the scheme chosen, thus it only offers the possibility to count particles in each box. The smoothing of the CIC method provide a more continuous representation of the density field closer to the result expected. Considering the advantages of the CIC mass assignment scheme, we will use only CIC in the analysis of one-point statistical distribution.

## 6.2 The log-normal model for the distribution of matter

A Gaussian model for the distribution of matter was used in Figure 6.1. We chose to show the distribution of the natural logarithm of the overdensity function (plus one in order to avoid negative arguments) in first place because the normal distribution is easier to distinguish. If we plot the histogram of the overdensity, we will find the results of Figure 6.2. As expected, the log-normal model fits with accuracy the distribution for lower redshifts. It is also suitable to describe the behaviour at higher redshifts, but at time is passing, gravity continues increasing the non-linearity of the distribution so, as consequence, the discrepancy between measurements and model increases.

Therefore, the next task is to study the evolution of the distribution moments with



**Figure 6.3:** Evolution of the distribution moments with the redshift. CIC mass assignment was applied over the simulation with  $N = 200$  to measure them.

the advance of time. At the beginning of the Universe, the distribution of matter density was Gaussian. Then, the gravitational force starts to couple Fourier modes and the PDF evolves to a log-normal distribution in times of the order of the first redshift saved by the simulation. The structure formation can be discussed in terms of the early moments: mean value, variance, skewness and kurtosis. As expected, the mean value decreases with time, which means that gravity makes the regions with a value of density upper than the critical density of the Universe collapse and then these tend to attract matter to themselves so, the most part of the Universe matter starts to concentrate in specific regions and the volume of low density regions increases considerably. As gravity forces are attractive, the trend is set to continue.

The discussion in terms of the variance will be interesting when we consider a normal distribution. Under this conditions, the variance is related with the amount of values closer to the mean value that have a highly probability to be found. The increment of this moment with time indicates that the amount of regions with different structure formation increases too. Regarding the skewness, this moment measures the asymmetry of the normal distribution, pointing that if its value is positive, the values with more probability to be found will be shifted to the left with respect to the center of the Gaussian PDF, skewing the shape of the function. On the other hand, kurtosis measures the weight of the values far from the mean value in the distribution. Attending to Figure 6.3, the growth of skewness can be interpreted as the regions that concentrate most of the matter in Universe are very small in comparison with the total volume, whilst the kurtosis shows that the sample of values of overdensity increases with time, increasing also the richness of structures that we can observe.

## 6.3 Measuring the correlation function

To measure the correlation function, we start from the array of positions obtained from the simulation with  $N = 32^3$ , at  $z = 0$ , in order to make possible its calculation, which scales with  $N^2$ . Apart from this, we distribute randomly  $64^3$  particles in the same volume to measure the contrast of a Universe without gravity. As we explained in the theoretical development, the correlation function can be obtained as a combination of the arrays DD, DR and RR.

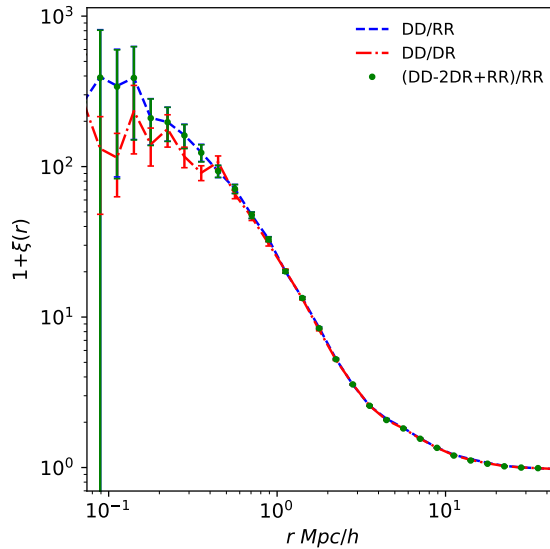
Let us consider two samples of particle positions, D and R. We are interested to know the probability of finding a couple of particles separated by a distance  $r$ . Considering that one of the particles owns to D and the other to R, Landy-Szalay (Landy & Szalay, 1993) proposed to measure the arrays as the following: firstly, we focus our attention in one particle in D, then we consider spherical caps with center in the position of the selected particle and count the number of particles inside each cap, considering the distribution R in the measurement. This calculation can be evaluated as an histogram and stored in an array. This process must be repeated for each particle in D, adding the result to the previous array. This way, we will obtain a final array with the number of couples of particles found in each interval of distance defined by the spherical caps. The limits of each cap will be separated by the same distance in a decimal logarithmic scale and each interval will be identified by the distance of its center to the center of the sphere. In this description an important detail is missing: the maximum radius of a spherical cap must be the half of the length of the box to be consistent with the length of the box considered and to avoid to deal with the corners of the same. Then, if we consider a particle that is not in the center, some caps will be outside of our volume. Keeping in mind again the cosmological principle, we will assume periodic boundary conditions and replicate the volume in order to complete the measurement. Another option, which is the followed in the implementation, is to shift the center of the box to the position of the particle considered changing the reference system and relocating the necessary particles.

If we define the measurement as  $\text{hist}(\text{DR})$ , i.e., the number of pairs contained in range  $[r, r+dr]$  and considering the procedure for the pairs DD and RR, the normalized expressions for these arrays are (Landy & Szalay, 1993):

$$\begin{aligned} DD &= \frac{2}{n(n-1)} \text{hist}(DD) \\ RR &= \frac{2}{n_r(n_r-1)} \text{hist}(RR) \\ DR &= \frac{1}{nn_r} \text{hist}(DR) \end{aligned} \tag{6.1}$$

where  $n$  is the number of pairs in the data sample and  $n_r$  the number of pairs in the random sample. Combining these estimators as in (4.22), we find agreement among

---



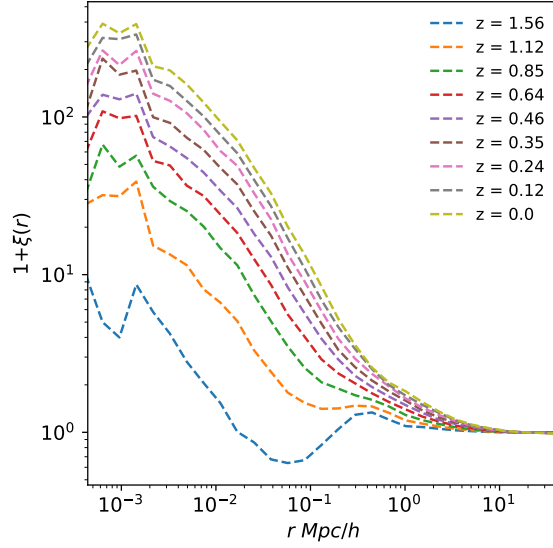
**Figure 6.4:** Two-point correlation function measured using three different estimators suggested at redshift zero.

them, as we show in Figure 6.2.

The shape of the measured two-point correlation function is as we expected: in small scales, we found an overcrowding of objects as consequence of the gravity domain and, on the other hand, the value of the correlation function for pairs separated by large distances tends to one (considering  $1 + \xi$ ), which means that they follow a random distribution. This observation agrees with the gravity law, the magnitude of the force decreases with the square of the distance so we expect that at large scales the contribution of it is negligible. Therefore, the Universe at large scale seems to be homogeneous, in concordance with the Cosmological Principle. It is important also to note the discrepancies observed for the different methods at small scales. They are associated to the resolution of the simulation what is proportional to the softening length used by PKDGRAV3 (Douglas Potter, 2016) to smooth force evaluations of closer particles.

The most extended estimator in Cosmology is the Sandy-Szalay purpose because of the reduction of the error bars associated with the measurements. Unfortunately, we cannot prove this improvement (see Figure 6.4). This is because we only take one random distribution as reference and the errorbars that we considered were measured assuming a Poissonian distribution for the values of the histograms and performing the standard dispersion of the errors according to the expression of each estimator. The errors considered in (Landy & Szalay, 1993) are obtained as the variance of the average of the measurements of the estimator using one hundred of random references, a calculation that cannot be performed in this project because of the computation time.

Finally, we chose the Landy-Szalay estimator to show the evolution of the correlation function with redshift (see Figure 6.5). On scales of  $10h^{-1}$  Mpc, the behaviour of the the

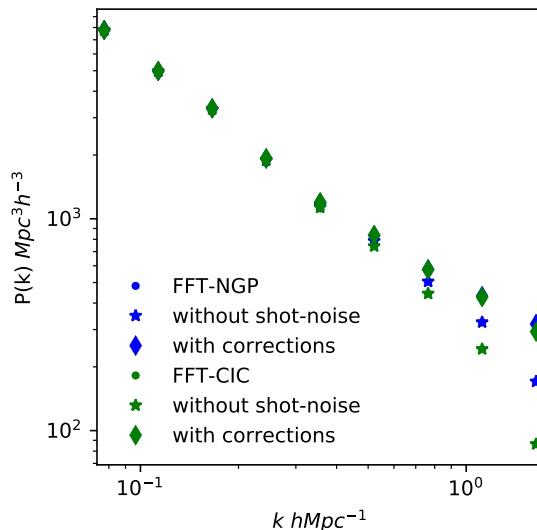


**Figure 6.5:** Evolution of correlation function with redshift. The curve with less amplitude is related with early times in the simulation and the curve with the most with present days.

distribution does not vary significantly with redshift, which introduce an upper limit to the range of influence of gravity in matter distribution. On the other side, the excess of matter in closer regions grows with time and, the shortest the distance between the pair, the greater the accumulation will be. It is recommended (Landy & Szalay, 1993) to use a random sampling with large number of particles than the distribution in order to improve the comparison. Unexpected oscillations and behaviours observed, specially in curves for early epochs, can be related with a low quality of the random sample. This can be improved generating more random samplings and then obtaining the average of them (Landy & Szalay, 1993), but this could not be implemented for lack of time.

## 6.4 Measuring the matter Power Spectrum

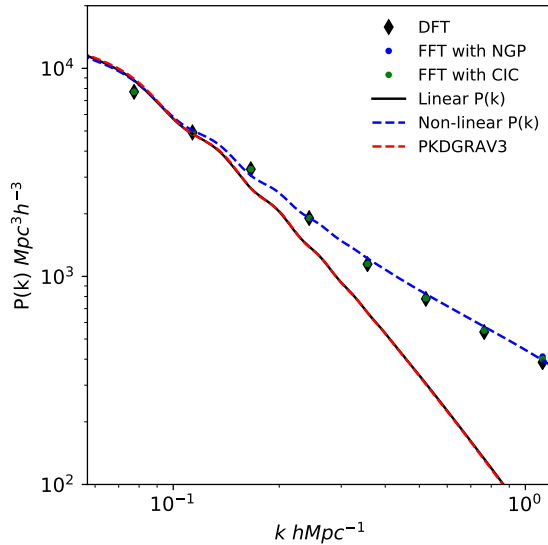
For the last task of the project, we take the matter overdensity distribution calculated using the NGP and CIC methods, and then apply the FFT implemented in the FFTWpy library of Python to obtain the Fourier transformation of it. The Power Spectrum was measured according to equation (4.23), performing an average in spherical shells, whose width is constant in the decimal logarithm of the modes, without applying any of the corrections discussed in theoretical section over the measurement. To observe the effect of corrections in the Power Spectrum (see Figure 6.6), we apply the shot-noise correction which is negligible because  $N$  is large and then the mass assign-



**Figure 6.6:** Measurements of Power Spectrum obtained using NGP and CIC schemes, where we compare the FFT measurement with its correction by shot noise and mass assignment.

ment correction, dividing the Power Spectrum corrected by shot noise by the square of the window function in Fourier space (equation (4.6)). It is only an approximation to equation (4.42), where we do not consider corrections to aliasing effect. Comparing the spectra after and before corrections, it is observed that, at large scales, the mass assignment have no relevant effects on the shape, but it produces a loss of power at smaller scales. This result is what we expected considering that at large scales the distribution of matter is homogeneous, so the criteria used to assign mass to cells will not change significantly the measurement but, when we go to small scales where the distribution of matter is more complex in comparison with structure formation, the way chosen to compute the density field is relevant. Once the correction is done, NGP and CIC measurements offer closer results which means, in principle, that the correction was implemented correctly. According to the definition (4.23), Power Spectrum is an dimensionless function but the normalization used in Cosmology (Peebles, 1980; Dodelson, 2003; Padmanabhan, 1993) is given by  $\frac{V}{N_m}$ , where  $V$  is the peculiar comoving volume of the simulation and  $N_m$  is the number of modes of the Fourier space.

At this point we verify whether the simulation performed offers acceptable results for the evolution of the matter distribution is not possible yet. Firstly, we must to be sure that the result obtained for the corrected Power Spectra is what we expected. Keeping in mind the theoretical development of Power Spectrum, the most direct way to measure it is performing DFT, which is also the most expensive. DFT was implemented by the student and the result, corrected by shot-noise, is shown in Figure 6.7. FFT and DFT results are closer, which means that the power spectrum of the simulated



**Figure 6.7:** Comparison between theoretical Power Spectrum obtained from CAMB and the simulation measurements corrected by shot noise and mass assignment at present days. PKDGRAV3 lineal Power Spectrum is plotted too, verifying that it is with concordance with CAMB.

matter distribution was measured successfully.

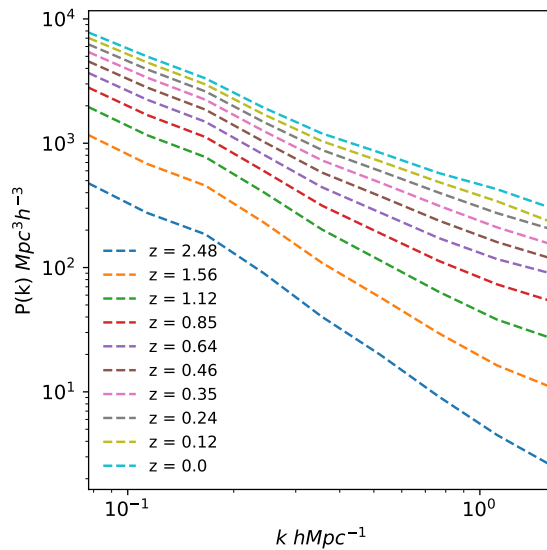
Now, we have all the ingredients to check if the simulation obtained from PKDGRAV3 code is in concordance with cosmological models. To do so, we use the online tool CAMB <sup>1</sup> to obtain two theoretical curves for power spectrum. The first is the linear power spectrum, which is related with the matter distribution with Gaussian fluctuations, and the other is a non-linear power spectrum, which differs from the previous at lower scales where high values of overdensity function differs from the lineal approximation, which results in an excess of power (for more details, see Appendix II 9). Cosmological parameters used in PKDGRAV3 were introduced in the configuration. This way, we can check if the behaviour of the measured power spectrum is closer to linearity or non-linearity. In addition, we introduce the Power Spectrum used by PKDGRAV3 to generate the initial conditions, where it is expected that the power spectrum was linear.

Before present the results, we need to introduce the theoretical shape of a linear power spectrum (see Appendix II (9)), given by Hu (1997):

$$P(k) = Ak^n T(k, z)^2 \left( \frac{D_1(z)}{D_1(0)} \right)^2 \quad (6.2)$$

where  $A$  is a normalization constant,  $n$  is the spectral index,  $T(k, z)$  is the transfer

<sup>1</sup>[https://lambda.gsfc.nasa.gov/toolbox/tb\\_camb\\_form.cfm](https://lambda.gsfc.nasa.gov/toolbox/tb_camb_form.cfm)



**Figure 6.8:** Evolution of Power Spectrum with redshift. The spectrum at redshift zero is at the top, and the obtained in first time saved is at the bottom.

function defined as the ratio of time-integrated growth on a particular scale as compared to that on scales far larger than the Jeans scale (Jeans, 1902).  $D_1(z)$  is the growth factor of the Universe and is normalized with its value at redshift zero.

Historically, it has been adopted the use of the dimensionless power spectrum as  $\Delta^2 = \frac{k^3 P(k)}{2\pi^2}$ , which measures the contribution of perturbations per unit logarithmic interval at wavenumber  $k$  to the variance in the matter density fluctuations (Peebles, 1980; Padmanabhan, 1993). To normalize this function, the literature suggests to use the parameter  $\sigma_8$  defined as the amplitude of the linear power spectrum at present days in a scale  $R = 8 \text{ Mpc/h}$ :

$$\sigma_8^2 = \int \frac{dk}{k} \Delta^2 |W(R = 8, k)|^2 \quad (6.3)$$

where  $W(R, k)$  is the Fourier transform of the top hat function, given by:

$$W(R = 8, k) = \frac{3}{(8k)^3} (\sin(8k) - 8k \cos(8k)) \quad (6.4)$$

To normalize the theoretical curves, we obtain  $\sigma_8$  for the theoretical function and then the  $\sigma_8$  of the simulation is imposed in this way:

$$P_{th} = \left( \frac{\sigma_8^{sim}}{\sigma_8^{CAMB}} \right)^2 P(k) \quad (6.5)$$



---

As we explained previously, PKDGRAV3 uses a file which contains the transfer function that used to generate the initial conditions, so we take it too and use eq. (6.2) and eq. (6.3) to measure the linear power spectrum of the code 6.7. Once all power spectra are correctly normalized, we compare them in a plot, obtaining that the non-linear power spectrum fits satisfactory simulation measurements. Finally, we can be sure that Power Spectrum measurement and PKDGRAV3 performance are right.

Finally, we discuss the evolution of Power Spectrum with redshift. As Figure 6.8 shows, the spectral amplitude grows with pass of time, preserving its shape for large scales and deviating from the liner behaviour for low scales. It is a prove of clustering processes at low scales but we must interpret this result carefully. In scales closer to Nyquist frequency aliasing effect is important so a limit to trust the measurement must be defined.

---

## 7 Conclusion

The statistical analysis of the matter distribution shows that one-point statistics is very useful to make a quantitative description of its evolution, beginning from a log-normal distribution at redshift of the first snapshot stored which can be understood as a normal distribution taking the natural logarithm of the overdensity, whose main value moves to low densities with time. Matter tends to collapse in regions whose density is higher than the critical density and it accumulates progressively in these spots. The two-point statistical analysis reveals how gravity influences at different scales to structure formation, having more importance in the smallest scales. Two-point correlation function results show us that the three estimators proposed offer close results which agree with the expected distribution of matter for redshift zero and the experience of the implementation shows the difficulty of the task, which is related with the computational efficiency of the code for a large volume of calculations. This fact motivates the study of two-point statistics through the Power Spectrum formalism, whose technical details were discussed in depth in order to obtain the physical information of the system with high quality, i.e. avoiding as much as possible the noise introduced in the measurement by the technique.

Regarding the technical part of the development, we have shown that the N-body code PKDGRAV3 performs the evolution of matter distribution according to standard cosmological model  $\Lambda$ CDM predictions. The detailed knowledge about how window functions assign mass to cells was essential to predict its influence in the shape of power spectrum measured, specially in small scales.

Throughout the three months of the project, has acquired the basis to understand the current cosmological theories as well as the limitations in parameter constraints. The theoretical development of the tasks not only expands his wisdom about statistical techniques and Fourier space properties, it also shows how to apply this mathematical tools to analyze the gravity effects over matter distribution. Apart from this, programming skills in Python language have been improved to be able to approach the calculations required by the study, specially skills related with code efficiency.

In order to improve as the results obtained, as well as the methodology applied in this bachelor's project, it is proposed to implement the measurement of correlation function and direct Fourier transform in parallel code using C++ to reduce the computational time, which was one of the most important difficulties in the development of work. This way, it could be possible to try to use more random references in order to discuss the L-S error results. It would be interesting to increment the size of the box of the simulation, this way we could see more details in Power Spectrum and correlation function. A

higher mass resolution, i.e., to run a simulation with larger number of particles (for a fixed volume) will increase the details observed in the analysis of one-point and two-point correlation function, allowing to increment the size of the grid, increasing the amount of modes in Fourier space. In addition, it would be also interesting to discuss the influence of the resolution offered by PKDGRAV3 related with the softening length, allowing to establish a limit henceforth not consider as true the measurements. To go further in cosmological analysis, the analysis of the slope in Power Spectrum in small scales could be done in order to apply a fractal analysis of the matter distribution to show more details about the structure of the survey.

---

# Bibliography

- Barrow J., Coles P., 1990, MNRAS, 244, 188
- Cepa J., 2007, *Cosmología Física*. Ediciones Akal
- Cooley J. W., Turkey J. W., 1965, *Math. Comput.*, 19: 297-301
- Cui W., Liu L., Yang X., Wang Y., Feng L., Springel V., 2008, *ApJ*, 687, 738
- Dodelson S., 2003, *Modern Cosmology*. Academic Press, Elsevier Science
- Douglas Potter Joachim Stadel R. T., 2016, *Computational Astrophysics and Cosmology*
- Eisenhart L. P., 1961, *Continuous Groups of Transformations*. Dover Publications, New York
- Eisenhart L. P., 1962, *Riemannian Geometry*. Princeton University Press, Princeton, N. J.
- Fry J. N., 1985, *ApJ*, 289, 10
- Griffiths L. M., Barbosa D., Liddle A. R., 1999, MNRAS, 308, 854
- Helgason S., 1962, *Differential Geometry and Symmetric Spaces*. Academic Press, New York
- Hockney R. W., Eastwood J. W., 1981, *Computer simulations using particles*. McGraw-Hill
- Hongguang Bi Li-Zhi Fang L. F. Y. J., 2003, *ApJ*, 598, 1
- Hu D. E. . W., 1997, *ApJ*, 511 5
- Hubble E., 1934, *ApJ*, 79, 8
- Jarosik N., et al., 2011, *ApJS*, 192, 14
- Jeans J. H., 1902, *Philosophical Transactions of the Royal Society of London Series A*, 199, 1
- Jing Y. P., 2005, *ApJ*, 620, 559

- 
- Jones P. C. . B., 1990, MNRAS, 248, 1
- Kirchner J. W., 2005, Phys. Rev. E, 71, 066110
- Landy S. D., Szalay A. S., 1993, ApJ, 412, 64-71
- Lewis A., 2014, DAMTP, CITA, IoA, Sussex
- Marchese L., 2018, Friedmann equations. The Newtonian derivation of Friedmann equations and an introductory study of them as a dynamical system
- NASA 2019, CAMB Web Interface, [https://lambda.gsfc.nasa.gov/toolbox/tb\\_camb\\_form.cfm](https://lambda.gsfc.nasa.gov/toolbox/tb_camb_form.cfm)
- Nyquist H., 1928, Transactions of the American Institute of Electrical Engineers, 47, 617
- Padmanabhan T., 1993, Structure formation in the universe. University of Cambridge
- Peebles P., 1980, The large scale structure of the universe. Princeton: Princeton University Press
- Planck Collaboration et al., 2016, A&A, 594, A13
- Pérez Hernández F. J., 2018, Relatividad general (Grado en Física). Departamento de Astrofísica, Universidad de La Laguna
- Requena Plens J. M., Rocamora Zamora J., 2019, Plantilla TFG/TFM
- Ridler 1977, Essential Relativity. Special, general and cosmological. Springer-Verlag
- Weinberg S., 1972, Gravitation and cosmology, principles and applications of the general theory of relativity. John Wiley & Sons Inc
-

## 8 Appendix I: Friedmann equations

Firstly, we will introduce the metric considered in The Standard Cosmological Model and then we will use it to obtain the Friedmann equations from Einstein's field equations.

According to the development of Weinberg (1972), Eisenhart (1961), Eisenhart (1962), Helgason (1962), the most general form for a metric consistent with the Cosmological Principle is given by:

$$ds^2 = g_{ab}(v)dv^a dv^b + f(v) \left( d\mathbf{u}^2 + \frac{k(\mathbf{u} \cdot d\mathbf{u})^2}{1 - k\mathbf{u}^2} \right) \quad (8.1)$$

where  $\mathbf{u}$  and  $\mathbf{v}$  are vectors that belong to the different maximal symmetric subspaces that is composed the general space,  $f(v)$  is a positive function and  $k$  is the sign of the curvature factor  $K$ , which can take the values:

$$\begin{cases} k = +1 & \text{if max. sym. subsp. has } K > 0 \\ k = -1 & \text{if max. sym. subsp. has } K < 0 \\ k = 0 & \text{if max. sym. subsp. has } K = 0 \end{cases} \quad (8.2)$$

If we take the coordinates:

$$\begin{cases} \int (-g(v))^{\frac{1}{2}} dv \equiv t \\ u^1 \equiv r \sin \theta \cos \varphi \\ u^2 \equiv r \sin \theta \sin \varphi \\ u^3 \equiv r \cos \theta \end{cases} \quad (8.3)$$

We obtain the Robertson-Walker-Friedmann-Lemaitre metric Ceba (2007), Weinberg (1972), Padmanabhan (1993), Ridler (1977):

$$ds^2 = -dt^2 + a^2(t) \left( \frac{dr^2}{1 - kr^2} + r^2 d\theta^2 + r^2 \sin^2 \theta d\varphi^2 \right) \quad (8.4)$$

where  $a(t)$  is the scale factor. In this process, we obtain the expression for the metric following geometrical arguments, without introducing the distribution of matter and energy of the universe, so the scale factor cannot be determined. In order to obtain it, Einstein's field equations must be solved.

Einstein's field equations Weinberg (1972), Ceba (2007), Pérez Hernández (2018), Ridler

(1977) are given by:

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 8\pi T_{\mu\nu} - \Lambda g_{\mu\nu} \quad (8.5)$$

where  $R_{\mu\nu}$  is the Ricci curvature tensor,  $R$  is the Ricci scalar,  $g_{\mu\nu}$  is the metric tensor and  $T_{\mu\nu}$  is the stress-energy tensor. The gravitational constant  $G$  and the speed of light in vacuum  $c$  are taken in natural scale where both are equal to one. The cosmological parameter  $\Lambda$  can be introduced in the equation according to the range of freedom of the solution of the metric.

For the case of an ideal fluid, i.e., a fluid in which viscosity forces can be neglected, we obtain Cepa (2007), Peebles (1980), Pérez Hernández (2018), Weinberg (1972):

$$R_{\beta}^{\alpha} - \frac{1}{2}R\delta_{\beta}^{\alpha} + \Lambda\delta_{\beta}^{\alpha} = 8\pi[-(p + \rho)\delta_{0}^{\alpha} + p]\delta_{\beta}^{\alpha} \quad (8.6)$$

where  $\delta_{\beta}^{\alpha}$  is the Kroneker delta and  $p$  and  $\rho$  are the pressure and density field respectively.

From (8.4), measuring the Christoffel symbols, the components of the Ricci curvature tensor are given by Cepa (2007), Weinberg (1972), Ridler (1977):

$$\begin{cases} R_0^0 = \frac{3\ddot{a}}{a} \\ R_1^1 = R_2^2 = R_3^3 = \frac{\ddot{a}}{a} + \frac{2\dot{a}^2 + 2\epsilon}{a^2} \\ R_{\beta}^{\alpha} = 0 \end{cases} \quad \text{if } \alpha \neq \beta \quad (8.7)$$

So, the Ricci scalar is obtained as a sum of each component:

$$R = 6 \left( \frac{\ddot{a}}{a} + \frac{\dot{a}^2 + \epsilon}{a^2} \right) \quad (8.8)$$

From (8.6), using (8.7) and (8.8), we can obtain the pressure and density field as function of the scalar factor and its derivatives:

$$\begin{cases} p = -\frac{1}{8\pi} \left( \frac{2\ddot{a}}{a} + \frac{\dot{a}^2 + \epsilon}{a^2} - \Lambda \right) & \text{if } \alpha = \beta = 1, 2, 3 \\ \rho = \frac{1}{8\pi} \left( 3\frac{\dot{a}^2 + \epsilon}{a^2} - \Lambda \right) & \text{if } \alpha = \beta = 0 \end{cases} \quad (8.9)$$

where we replace the pressure expression in the second identity using the first.

If we reorganize the first identity and then combine both equations properly, we finally find the Fridmann equations Cepa (2007), Padmanabhan (1993), Peebles (1980), Wein-

berg (1972):

$$\begin{aligned} H^2 &= \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3}\rho - \frac{k}{a^2} + \frac{\Lambda}{3}, \\ \left(\frac{\ddot{a}}{a}\right) &= -\frac{4\pi}{3}(\rho + 3p) + \frac{\Lambda}{3} \end{aligned} \quad (8.10)$$

The redshift is defined as Weinberg (1972), Cepa (2007), Padmanabhan (1993):

$$z = \frac{\lambda_0 - \lambda_1}{\lambda_1} \quad (8.11)$$

where  $\lambda_0$  is the wavelength of a wave measured at present days and  $\lambda_1$  the wavelength measured in the source.

According to Relativistic Doppler effect, redshift can be written as:

$$z = \frac{a(t_0)}{a(t)} - 1 \quad (8.12)$$

with  $t_0$  the present days time. If we use a Taylor expansion, redshift can be expressed as function of the Fridmann equations evaluated at present days:

$$z \approx H_0(t_0 - t) + H_0^2\left(1 + \frac{1}{2}q_0\right)(t_0 - t)^2 \quad (8.13)$$

where  $H_0$  is the Hubble's constant at present days and  $q_0 = -\frac{\ddot{a}}{aH}\Big|_{t=t_0}$  is the acceleration parameter.

---



## 9 Appendix II: Linear Power Spectrum

Let  $\mathbf{r}$  be the coordinates of a space whose metric is given by (8.4). Consider the set of moments of the Boltzmann equation, composed by Peebles (1980), Dodelson (2003):

- Mass conservation equation:

$$\left(\frac{\delta\rho}{\delta t}\right)_{\mathbf{r}} + \nabla_{\mathbf{r}}(\rho\mathbf{u}) = 0 \quad (9.1)$$

- Motion equation:

$$\left(\frac{\delta\mathbf{u}}{\delta t}\right)_{\mathbf{r}} + (\mathbf{u}\nabla_{\mathbf{r}})\mathbf{u} = -\nabla_{\mathbf{r}}\Phi \quad (9.2)$$

and the Poisson equation, which denotes the source of gravity and closes the system:

$$\nabla_{\mathbf{r}}^2\Phi = 4\pi G\rho \quad (9.3)$$

where  $\mathbf{u}$  is the velocity,  $\phi$  is the potential field of the force,  $\nabla$  is the gradient operator and  $\nabla^2$  is the Laplace operator.

We define the comoving coordinates as follows:

$$\mathbf{x} = \frac{1}{a}\mathbf{r} \quad (9.4)$$

$$\mathbf{u} = \dot{(a)}\mathbf{x} + \mathbf{v}(\mathbf{x}, t) \quad (9.5)$$

Consider a function such as  $f = f(t, \mathbf{x})$ . Then its partial derivatives respect to  $t$  and  $\mathbf{x}$  as a function of  $\mathbf{r}$  coordinates are given by:

$$\left(\frac{\delta f}{\delta t}\right)_{\mathbf{r}} = \left(\frac{\delta f}{\delta t}\right)_{\mathbf{x}} - \frac{\dot{a}}{a}\mathbf{x}\nabla f \quad (9.6)$$

$$\nabla = a\nabla_{\mathbf{r}}$$

where we rename  $\nabla = \nabla_{\mathbf{x}}$ .

We introduce the density field as a function of overdensity function Peebles (1980), Do-  
delson (2003):

$$\rho = \rho_b(t)(1 + \delta(\mathbf{x}, t)) \quad (9.7)$$

From (9.1), (9.6), (9.7), we find:

$$\left( \frac{\delta}{\delta t} - \frac{\dot{a}}{a} \mathbf{x} \nabla \right) [\rho_b(1 + \delta)] + \frac{\rho_b}{a} [(1 + \delta)(\dot{a} \mathbf{x} + \mathbf{v})] = 0 \quad (9.8)$$

Taking time derivatives in (8.10), we find that  $\dot{\rho}_b = -3\frac{\dot{a}}{a}$ , so:

$$\frac{\delta(\delta)}{\delta t} + \frac{1}{a} \nabla [(1 + \delta) \mathbf{v}] = 0 \quad (9.9)$$

For Poisson's equation (9.3), we find:

$$\frac{1}{a^2} \nabla^2 \Phi = 4\pi G \rho_b (1 + \delta) \quad (9.10)$$

If we consider the field  $\phi(\mathbf{x}, t)$  from Fridmann equations (8.10), such as:

$$\Phi(\mathbf{x}, t) = \phi(\mathbf{x}, t) + \frac{2}{3} \pi G \rho_b a^2 \mathbf{x}^2 + \frac{1}{6} \Lambda a^2 \mathbf{x}^2 \quad (9.11)$$

we find:

$$\nabla^2 \phi = 4\pi G a^2 \rho_b \delta \quad (9.12)$$

For motion equation (9.2), we find:

$$\begin{aligned} \left( \frac{\delta}{\delta t} - \frac{\dot{a}}{a} \mathbf{x} \nabla \right) (\dot{a} \mathbf{x} + \mathbf{v}) + \frac{1}{a} (\dot{a} \mathbf{x} + \mathbf{v}) \nabla (\dot{a} \mathbf{x} + \mathbf{v}) = \\ -\frac{1}{a} \nabla \phi - \frac{4}{3} \pi G \rho_b a \mathbf{x} + \frac{1}{3} \Lambda a \mathbf{x} \end{aligned} \quad (9.13)$$

which we can simplify as:

$$\frac{\delta \mathbf{v}}{\delta t} + \frac{\dot{a}}{a} \mathbf{v} + \frac{1}{a} (\mathbf{v} \nabla) \mathbf{v} = -\frac{1}{a} \nabla \phi \quad (9.14)$$

Using a linear perturbation description, we will neglect terms of order  $\mathbf{v}\delta$  and  $\mathbf{v}^2$ ,  
obtaining from (9.9) and (9.14):

$$\begin{cases} \frac{\delta(\delta)}{\delta t} + \frac{1}{a} \nabla \mathbf{v} = 0 \\ \frac{\delta \mathbf{v}}{\delta t} + \frac{\dot{a}}{a} \mathbf{v} + \frac{1}{a} \nabla \phi = 0 \end{cases} \quad (9.15)$$

Finally, using Poisson's equation (9.12), we find:

$$\frac{\delta^2(\delta)}{\delta t^2} + 2\frac{\dot{a}}{a}\frac{\delta(\delta)}{\delta t} = 4\pi G\rho_b\delta \quad (9.16)$$

The general solution of this equation is:

$$\delta = A(x)D_+(t) + B(x)D_-(t) \quad (9.17)$$

where  $D_+(t)$  is a increasing function of time and  $D_-(t)$  a decreasing function of time, both independent. According to observations, we consider that  $D_-(t)$  is not a physical solution, so  $\delta = A(x)D_+(t)$ .

From the calculation in Dodelson (2003), bearing in mind the Power Spectrum definition (4.23), we can write:

$$P(k) = 2\pi\delta_H^2\frac{k^n}{H_0^{n+3}}T^2(k)\left(\frac{D_+(z)}{D_+(z=0)}\right)^2 \quad (9.18)$$

where  $\delta_H$  is a constant,  $n$  is the spectral index,  $T(k)$  is the transfer function and  $D_+(z)$  is called the growth factor.

---