Universidade de São Paulo
Instituto de Física

Técnicas estatísticas para levantamentos futuros: extraindo física fundamental da estrutura em larga escala

Caroline Macedo Guandalin



Orientador: Prof. Dr. Luis Raul Weber Abramo
Co-orientador: Dr. Philip Bull

Tese de doutorado apresentada ao Instituto de Física da Universidade de São Paulo, como requisito parcial para a obtenção do título de Doutora em Ciências.

Banca Examinadora:
Prof. Dr. Luis Raul Weber Abramo (IF-USP)
Prof. Dr. Marcos Vinicius Borges Teixeira Lima (IF-USP)
Prof. Dr. Miguel Boavista Quartin (IF-UFRJ)
Dr. Emanuele Castorina (U. Milano)
Dr. Pedro Ferreira (U. Oxford)

São Paulo
2021

FICHA CATALOGRÁFICA
Preparada pelo Serviço de Biblioteca e Informação do Instituto de Física da Universidade de São Paulo

Guandalin, Caroline Macedo

Técnicas estatísticas para levantamentos futuros: extraindo física fundamental da estrutura em larga escala / Statistical techniques for future Surveys: extracting fundamental physics out of the large-scale structure. São Paulo, 2021.

Tese (Doutorado) – Universidade de São Paulo. Instituto de Física. Depto. de Física Matemática.

Orientador(a): Prof. Dr. Luis Raul Weber Abramo

Área de Concentração: Física

Unitermos: 1. Estrutura do universo; 2. Radiotelescópios; 3. Relatividade geral; 4. Redshifts fotométricos; 5. Universo primordial.

USP/IF/SBI-090/2021

University of São Paulo
Physics Institute

Statistical techniques for future surveys: extracting fundamental physics out of the large-scale structure

Caroline Macedo Guandalin

Supervisor: Prof. Dr. Luis Raul Weber Abramo
Co-supervisor: Dr. Philip Bull

Thesis submitted to the Physics Institute of the University of São Paulo in partial fulfillment of the requirements for the degree of Doctor of Science.

Examining Committee:
Prof. Dr. Luis Raul Weber Abramo (IF-USP)
Prof. Dr. Marcos Vinicius Borges Teixeira Lima (IF-USP)
Prof. Dr. Miguel Boavista Quartin (IF-UFRJ)
Dr. Emanuele Castorina (U. Milano)
Dr. Pedro Ferreira (U. Oxford)

São Paulo
2021

As opiniões, hipóteses e conclusões ou recomendações expressas neste material são de responsabilidade da autora e não necessariamente refletem a visão da FAPESP.

The opinions, hypotheses, and conclusions or recommendations expressed in this material are the author's responsibility and do not necessarily reflect the views of FAPESP.

Acknowledgments

I owe my gratitude to my advisor, Prof. Raul Abramo, who has been mentoring and inspiring me for the past 5 years. Obrigada Raul!

Thanks to Phil Bull for opening countless doors, bringing me to the amazing field of relativistic effects, and to meet the fantastic group at Queen Mary with whom I had the pleasure to collaborate with. Alongside with Phil, I should also thank Chris Clarkson for getting on board and for the will to continue with what lies ahead. Thanks Phil and Chris!

To the three people without whom this thesis would not be what it is, who contribute to my professional growth and provided amazing discussions, help and guidance: Julian Adamek, Rodrigo Voivodic and David Alonso. Danke Julian! Obrigada Rodrigo! Gracias David!

For a thorough revision of the first stages of this work, I would also like to thank Marcos Lima, Elisa Ferreira, and Edivaldo Santos; in particular, Marcos Lima should also be thanked for all the guidance, advises and support. Obrigada Marcos, Elisa e Edivaldo!

For boosting my PhD experience, I thank the whole ICTP-SAIFR group at the Institute for Theoretical Physics (IFT-UNESP). In particular, I thank Hugo Camacho and Antonino Troja for all the discussions, insights and collaboration.

For making the whole process brighter, I thank my amazing cosmologist friends and colleagues, Renan Boschetti, Louis Coates, Antonio Dorta, Arthur Loureiro, Francisco Maion, Thiago Mergulhão, Carolina Queiroz, Natalia Rodrigues, Natali de Santi, Beatriz Tucci, and Rodrigo Voivodic. Special thanks to Henrique Rubira, who reminded me of the PhD application deadline. I also thank Ana Camila Esteves for many exciting discussions.

For the essential financial support, I thank the *Fundação de Amparo à Pesquisa do Estado de São Paulo* (FAPESP), grant number 2018/10396-2, and the Queen Mary University of London grant PHY2420B.

I also thank all the employees of the Institute of Physics, without whom there is no institute. Special thanks to the DFMA's secretariat (obrigada Simone, Marcelo e Cecília!) and the graduate committee (obrigada Andrea e Éber!).

I thank Felipe Prado for all the support, companionship and love during the last years. You should also be thanked for awakening my desire to know things with all the depth they deserve. Obrigada Fê!

I thank my whole family, specially my mother Mary and father Luiz for prioritising my education: obrigada!

Finally, for being my home for the past 10 years, I thank the Institute of Physics and the University of São Paulo.

For all the tenure of humans on Earth, the night sky had been a companion and an inspiration. The stars were comforting. They seemed to demonstrate that the heavens were created for the benefit and instruction of humans. This pathetic conceit became the conventional wisdom worldwide. No culture was free of it. Some people found in the skies an aperture to the religious sensibility. Many were awestruck and humbled by the glory and scale of the cosmos. Others were stimulated to the most extravagant flights of fancy. At the very moment that humans discovered the scale of the universe and found that their most unconstrained fancies were in fact dwarfed by the true dimensions of even the Milky Way Galaxy, they took steps that ensured that their descendants would be unable to see the stars at all. For a million years humans had grown up with a personal daily knowledge of the vault of heaven. I the last few thousand years they began building and emigrating to the cities. In the last few decades, a major fraction of the human population had abandoned a rustic way of life. As technology developed and the cities were polluted, the nights became starless. New generations grew to maturity wholly ignorant of the sky that had transfixed their ancestors and that had stimulated the modern age of science and technology. Without even noticing, just as astronomy entered a golden age most people cut themselves off from the sky, a cosmic isolationism that ended only with the dawn of space exploration.

— Contact (Carl Sagan).

So let there be more light

Abstract

We have seen an unprecedented development in the field of cosmology, in the past decades, with the development of increasingly detailed cosmic microwave background maps. Notwithstanding, the amount of information one can extract from those two-dimensional maps is limited when compared to what can be achieved through the three-dimensional mapping obtained with galaxy surveys. The spatial dark matter distribution inferred with these surveys depends on the highly non-linear gravitational collapse, which can be incorporated in our three-dimensional description through N-body simulations (usually performed within Newtonian gravity) up to the tens of Mpc scales. In contrast, scales at the order of hundreds of Mpc are free from astrophysical effects and can be accurately described with perturbation theory. At these scales, relic features characteristic of primordial universe physics are left in the n-point statistics of dark matter tracers (e.g. galaxies and halos), and we can obtain novel gravitational effects through a theory-observables connection (the latter which are based on a set of fundamental observables such as redshift and observed angles on the sky). Primarily, we explore the relativistic anisotropies in the dark matter halo distribution that emerge after connecting the observed redshift with its theoretical general relativistic prediction. We focus on the power-spectrum dipole (2-point statistics) signal that appears after cross-correlating different halo populations (i.e. different masses) obtained from a weak-field relativistic N-body simulation. We make a complete presentation of the details necessary to extract essential parameters to model and interpret the results on an observed light cone. From an observational perspective, while it is desirable to have high-precision redshifts, we also require a large volume coverage to reach the necessary scales at which relativistic and primordial non-Gaussian effects (the latter is a characteristic feature of inflationary models) are manifested. However, when dealing with discrete dark matter tracers, we must acquire a large number of observations for the n-point signal to overcome its noise. Large volumes can be densely mapped through the so-called photometric redshifts, at the cost of redshifts with spectroscopic precision. Therefore, our second objective is to calibrate photometric redshifts utilising the clustering information of both galaxies and neutral hydrogen intensity mapping. We assess the ability of the bispectrum (3-point statistics) to recover the redshift distribution parameters, and we compare the results with the power spectrum. We also verify, for both 2- and 3-point statistics, how this clustering redshifts method depends upon the foreground contamination present in the neutral hydrogen maps.

Keywords: large-scale structure of the universe; relativistic effects; power spectrum; bispectrum; photometric redshifts.

Resumo

As últimas décadas trouxeram um desenvolvimento sem precedentes para a cosmologia através de mapas cada vez mais detalhados da radiação cósmica de fundo. Contudo, por representar uma informação essencialmente bidimensional, a quantidade de informação disponível é menor do que a obtida através dos mapas tridimensionais extraídos com levantamentos de galáxias. A distribuição de matéria escura inferida por esses levantamentos depende do processo altamente não-linear de colapso gravitacional, que pode ser incorporado na sua descrição através de simulações de N-corpos (geralmente Newtonianas) até escalas da ordem de dezenas de Mpc. Em contraste, escalas maiores (centenas de Mpc) são livres de efeitos astrofísicos e podem ser trabalhadas de forma acurada com teoria de perturbações. Nessas escalas podemos vincular a física do universo primordial através de características específicas deixadas nas estatísticas de n-pontos dos traçadores da matéria (e.g. galáxias e halos) e obter novos efeitos gravitacionais através da conexão entre a teoria e os observáveis (fundamentalmente dados em termos do redshift e ângulos de observação no céu). Primeiramente, exploramos as anisotropias da distribuição de halos de matéria escura devidas aos efeitos relativísticos que emergem ao conectarmos o redshift observado com o fornecido teoricamente pela relatividade geral. Focamos no dipolo do espectro de potência (estatística de 2-pontos) cruzado entre halos de diferentes massas de uma simulação de N-corpos relativística no limite de campos fracos. Apresentamos, em todos os detalhes, a sequência de como extrair parâmetros essenciais para descrever esse efeito, como modelá-lo e interpretá-lo no cone de luz observado. Do ponto de vista observacional, enquanto desejamos obter redshifts altamente precisos, através de levantamentos espectroscópicos, também desejamos cobrir o maior volume possível para alcançarmos as escalas onde efeitos relativísticos e não-Gaussianidades primordiais (característica de modelos inflacionários, por exemplo) podem ser observados. Contudo, ao trabalharmos com traçadores discretos da matéria, devemos ter uma densidade suficiente de objetos para que o sinal das funções de n-pontos supere o ruído. Esses volumes podem ser densamente mapeados através dos chamados levantamentos fotométricos, ao custo da precisão espectroscópica dos redshifts. Como segundo objetivo, desejamos calibrar os chamados redshifts fotométricos utilizando a correlação entre galáxias e mapas de intensidade de hidrogênio neutro. Avaliamos a capacidade do bispectro (estatística de 3-pontos) em recuperar os parâmetros da distribuição de redshift e comparamos com o espectro de potências, verificando como o método depende da contaminação em primeiro plano presente nos mapas de intensidade para ambos os casos.

Palavras-chave: estrutura em larga escala do universo; efeitos relativísticos; espectro de potências; bispectro; redshifts fotométricos.

Contents

1 Introduction **15**
- 1.1 Goals of this thesis . . . 25
- 1.2 Notation and conventions . . . 26

2 Cosmological Surveys: Mapping the Large-Scale Structure **28**
- 2.1 General considerations . . . 28
- 2.2 Galaxy redshift surveys . . . 32
  - 2.2.1 Spectroscopic surveys . . . 34
  - 2.2.2 Photometric surveys . . . 35
- 2.3 21-cm intensity mapping . . . 37
  - 2.3.1 Cosmology with 21-cm IM . . . 39
- 2.4 Multiple-tracers approach . . . 42

3 Statistical Analysis of Cosmological Surveys **45**
- 3.1 The signal and the noise . . . 45
  - 3.1.1 The signal from the second-order statistics . . . 46
  - 3.1.2 The signal from the third-order statistics . . . 48
  - 3.1.3 Shot noise . . . 50
  - 3.1.4 Cosmic variance . . . 54
- 3.2 An overview of estimators . . . 55
  - 3.2.1 Direct methods . . . 56
  - 3.2.2 Traditional methods . . . 59
  - 3.2.3 Quadratic compression . . . 63
  - 3.2.4 Cubic compression . . . 64

4 Primordial Features in the Large-Scale Structure **68**
- 4.1 Introductory remarks . . . 68
- 4.2 Theoretical framework . . . 70
  - 4.2.1 Primordial non-Gaussianity of the local type . . . 71
  - 4.2.2 The effect of PNG on halos . . . 73
- 4.3 Mock catalogues . . . 74

5 Relativistic Analysis of the Large-Scale Structure **76**
- 5.1 Theoretical background and rationale . . . 76
- 5.2 Multipoles of the power spectrum . . . 81
  - 5.2.1 Flat-sky approximation . . . 82
  - 5.2.2 Wide-angle corrections . . . 84

5.3 Summary . . . 84

6 Observing Relativistic Features in the Large-Scale Structure 86
6.1 Relativistic effects in the power spectrum . . . 86
6.2 Simulation . . . 88
6.2.1 The particle catalogue . . . 89
6.2.2 Halo catalogue . . . 91
6.3 Power-spectrum multipoles estimator . . . 99
6.3.1 Window function . . . 103
6.4 Results . . . 107

7 Clustering redshifts with 21-cm bispectra 113
7.1 Motivation . . . 113
7.2 Clustering redshift method . . . 115
7.2.1 Fisher matrix . . . 118
7.3 Forecast specifications . . . 121
7.3.1 Photometric galaxy survey . . . 125
7.3.2 HI intensity mapping . . . 126
7.3.3 Cross-correlations . . . 128
7.4 Application to a Gaussian distribution . . . 129
7.4.1 Power-spectrum constraints . . . 129
7.4.2 Bispectrum constraints . . . 131

8 Conclusions and Final Remarks 134

Bibliography 141

Appendices 158

A Derivation of Fundamental Equations 158
A.1 Geometric description of gravity . . . 158
A.2 Symmetries . . . 160
A.3 Friedmann–Lemaître–Robertson–Walker universes . . . 161
A.3.1 Friedmann equations . . . 164
A.3.2 Single-component solutions . . . 166
A.3.3 ΛCDM class of models . . . 167
A.3.4 Redshift . . . 168
A.3.5 Distances . . . 169

B Overview on Perturbation Theory 171
B.1 Motivation and limits of validity . . . 171
B.2 Perturbation theory in general relativity . . . 173
B.2.1 Scalar-Vector-Tensor (SVT) decomposition . . . 174
B.3 Gauge transformations and its consequences . . . 176
B.3.1 Gauge-invariant variables: metric sector . . . 178

B.3.2 Gauge-invariant variables: energy-momentum tensor . . . 180
B.3.3 Gauge choices . . . 181
B.4 Perturbed Einstein's equations . . . 182
B.4.1 Newtonian gauge . . . 182
B.5 Observables in cosmology . . . 186

C Tracing the Large-Scale Structure of the Universe **189**
C.1 Halo model . . . 189
C.1.1 Halo mass function . . . 190
C.1.2 Halo profile . . . 193
C.1.3 Halo bias . . . 194
C.2 Power spectrum . . . 196

D Fisher Matrix Formalism **197**
D.1 Definitions . . . 197
D.2 Power-spectrum Fisher matrix . . . 199
D.3 Bispectrum Fisher matrix . . . 203

Chapter 1

Introduction

> "*The Universe is everything. What it is, in its own right, independent of our changing opinions, we never fully know. It is all-inclusive and includes us as conscious beings. We are a part or an aspect of the Universe experiencing and thinking about itself.*"
>
> — Edward Harrison, 2003.

Why studying cosmology? As beautifully expressed by Harrison (2000, p, 20-21), the idea of an "Earth of unlimited bounty", where resources are available with an "inexhaustible abundance", with political ideologies assuming an "endless untapped wealth" to be explored no matter the consequences have proven to be untenable. These deceptive ideas can be traced to a cosmological myth that is still present in many cultures around the world. Therefore, knowing our place in the Universe is crucial as it "shapes our histories, inspire civilizations, foment wars, create monarchies, launch empires, and establish political systems". This reasoning can be applied to a general context of what is meant by cosmology (for an in-depth discussion on what cosmology is, or what it can signify to different people, see Harrison, 2000, Chapter 1), and it is thus hoped that, by shifting our cosmological views, we can break the anthropocentric myth of an Earth unconditionally available for our own needs at the expense of everything else.

Because the Universe is everything, we should make explicit within which domain of knowledge this particular work is inserted. Let us thus begin by saying what we will not mean by cosmology. It will not be the study of everything. It will not even be the study of gravity, or galaxies, or black holes. Broadly speaking, it will be a study of the macrodynamics of a model universe: we will be interested in the largest scales of the observable universe, after averaging over its microdynamics (small scales), that is, after blurring galaxies and clusters of galaxies (local inhomogeneities). This defines the core concept of this thesis, which we shall simply name cosmology. In this sense, it deals with the components of the physical universe,

and the processes and interactions that these components are subjected to during the evolution and expansion of the universe[1]

Cosmology will then provide us with models that allow us to understand how the components that were present in the very early universe evolved into the structures we see today. Hence, it has important anthropic consequences: if the constituents of a predefined cosmological model, or the physical processes dictating their evolution, were much different from the reasonable values empirically obtained by current and past observations, then there would be no room for the formation of the Earth and life on it, as we know (Ellis, 2007). Therefore, cosmology is also a subject of concern for other sciences: just like geology sets a bound on the age of the Earth, hence limiting the timescale of biological processes, it is unreasonable to propose the existence of common matter 14 billion years ago, or the formation of galaxies and planets when the universe proposed by our current models was approximately 380000 years old. Therefore, it has profound implications on scientific endeavour as a whole.

By our current physical knowledge, the physical processes operating during the evolution of the universe are defined by four fundamental forces of electromagnetism, the weak and strong interactions, and gravity. Their intricate interplay dictates the dynamics of everything in the natural world: from the formation of atoms and molecules, up to the far away objects we see such as stars and galaxies. It is also known that gravity has the smallest coupling among them (i.e. it is the weakest force). Yet, if we assume the universe to be electrically neutral, which seems a reasonable assumption, then gravity will be the dominant force in the scales that cosmology attempts to understand. Therefore, gravity will be the leading force in this work[2].

Up until now, the best theory we have for gravity is general relativity, proposed 106 years ago by Albert Einstein (for an excellent overview on general relativity, see Stewart, 1993, Chapter 1). Cosmology will then apply this theory when the strength of gravitational fields is weak, setting down the scales involved in this particular branch of science. On the other hand, as we look to the thermal history of the universe, the microphysics (i.e., the things we washed out to focus solely on the scales where gravity is the prominent force) will have a relevant role in the overall picture. At this stage, other forces become equally important. Finally, if we attempt to look at the primordial universe, where temperatures and densities were so high to the point that usual forms of matter were not present, as we believe, then cosmology will apply

[1] Throughout this thesis we use a distinction between the two spellings Universe and universe: the former will be exclusive to the concrete, existing Universe, while universe will be used whenever we refer to the abstract entity that depends upon models, cultures, epochs and societies. The central object of this work is the "modern physical universe", a particular rationalisation of the true Universe (Harrison, 2003).

[2] As already mentioned, cosmology aims at the macrodynamics of the universe after averaging over the microphysics. Despite the simplistic definition given here, the microphysics may impact the large scales, with other forces beyond gravity having important contributions to the physical modelling. For example, baryons interact not only through gravity, and modelling their effect on cosmological scales is an active area of research.

quantum theory of fields in curved spacetimes (for a pedagogical introduction, see Mukhanov and Winitzki, 2007) to treat this specific physical system.

Because general relativity is an intrinsic geometric theory, the scientific programme of observational cosmology is to determine the geometry of spacetime from the observations, with the former being done on our past light cone, which defines our observational past (Manchak, 2009; Ellis et al., 2012), and what the energy content of the universe is (and how it is distributed, at large scales, throughout space and time).

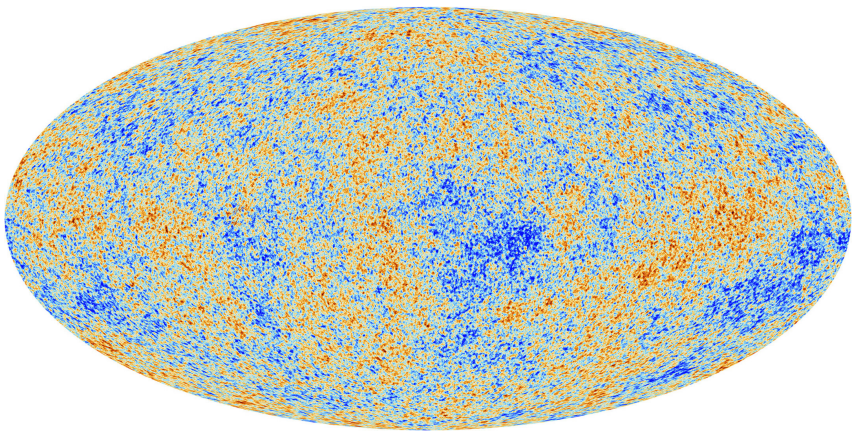


Figure 1.1: The Cosmic Microwave Background radiation from the time when universe was, approximately, 380000 years old. It corroborated the Big Bang theory, as it should leave a relic black-body radiation (Alpher and Herman, 1948, 1949) with temperature of 3 K, as predicted by Peebles (1966), which is in excellent agreement with the current known value of $T \approx 2.7$ K (Fixsen, 2009). In this figure, blue regions correspond to colder spots, with the hottest temperature (in red) differing from the coldest by $\sim 10^{-5}$ K. [Credit: ESA and the Planck Collaboration. Extracted from https://www.esa.int/].

One of the most astonishing observations we have of our past light cone is the cosmic microwave background (CMB, see Figure 1.1), which consists of a thermal relic of the very early universe. It shows a photon bath with a black-body spectrum of 2.7 K and, as discussed shortly, provides a strong evidence for a prediction of general relativity: that the spacetime is expanding. To weave the network of physical theories that altogether explains this observation is beyond the scope of this work, and we refer the interested reader to Giovannini (2008) and Durrer (2020).

This work will focus on another past light-cone observation: the large-scale structure (LSS), which is composed by a plethora of objects such as galaxies. As we shall discuss in Chapter 2, these galaxies are surveyed by telescopes, which furnish us with three-dimensional (3D) maps of patches of the Universe, after some set of assumptions are made. In this work we explore the effects that the general relativistic framework (Chapters 5 and 6), and early universe assumptions (Chapter 4) have on the 3D distribution of galaxies, information that can be extracted via summary statistics (Chapter 3). Finally, we will develop statistical techniques to improve the quality of these 3D maps when little information about the distance from the

galaxies to us is available (Chapter 7).

In what follows we present the fundamental concepts and observations that shall be assumed throughout this text. They are associated with the cosmic expansion, the redshift of galaxies, the cosmological principle, and the ΛCDM model that will be assumed in the remaining of the thesis. The background needed for a complete understanding is developed in details in the appendices. Finally, in Section 1.1 we define the overall goals of our work.

The expanding universe To reinforce how recent cosmology is as a science, by our modern standards, we made allusion to the so-called "Great debate" held 100 years ago at the Smithsonian Museum of Natural History, in which astronomers were not only debating the size of the hitherto observable universe, but also if distant nebulae were actually other galaxies as we know today, or if they were constituents of our own galaxy (Smith, 1982). Nine years after, the groundbreaking work of Edwin Hubble (1929) started to lay down the scales involved in extragalactic astronomy and cosmology.

Allied to Lemaître's theory of an expanding universe, originally proposed in 1927, but only disseminated among the scientific community after Eddington's translation in 1931 (Bagdonas et al., 2017), throughout the following years scientists started to interpret Hubble's observations as the expansion of the universe (Kragh, 1996). However, it was not until the discovery of the CMB (Figure 1.1) that the Hot Big Bang (HBB) model was established as a valid theory.

To understand Lemaître's argument, we will take a few steps backwards. The *radial* velocity $v \ll c$ of a galaxy can be determined by its redshift z through the Doppler effect,

$$z \equiv \frac{\lambda_o}{\lambda_e} - 1 = \frac{\Delta\lambda}{\lambda} \approx \frac{v}{c}, \tag{1.1}$$

with λ_e the wavelength of light waves emitted by the galaxy and λ_o the wavelength recorded by an observer (e.g. us on Earth). To infer the redshift z, astronomers observe the *spectral decomposition* of the integrated light emitted by the stars in a galaxy: the younger the galaxy, the larger is the number of blue stars contributing to the blue part of the spectrum, with redder parts being due to older stars (Sparke and Gallagher, 2007, p. 18). Also, the spectra of galaxies will have emission and absorption lines at specific frequencies varying between different molecules and atoms present in the intergalactic medium. Ultimately, a galactic spectrum will look like Figure 1.2, which is a mixture of many stellar spectra. By assuming that atomic physics is the same at the object (galaxy) of interest as it is here on Earth, roughly speaking the characteristic spectral features observed by a particular object in the sky are fitted against what we would expect to see in a laboratory, making λ_e at the galaxy's rest frame the same as here on Earth. Therefore, the difference observed in wavelength is translated to the galaxy velocity.

Lemaître's argument that the measured redshift of galaxies was not due to standard

Doppler effect, but rather to the relativistic effect of the stretching of light waves in an expanding space, was crucial for the community to start accepting the cosmic expansion. Assuming no other source of motion for the galaxies, Lemaître concluded that "*the receding velocities of extragalactic nebulae are a cosmical effect of the expansion of the universe*" (Lemaître, 1927, 1931). In this work he also derived the so-called distance-redshift relation as a consequence of the Doppler effect due to the variation of the radius of an expanding universe, which is presented from a modern point-of-view in what follows.

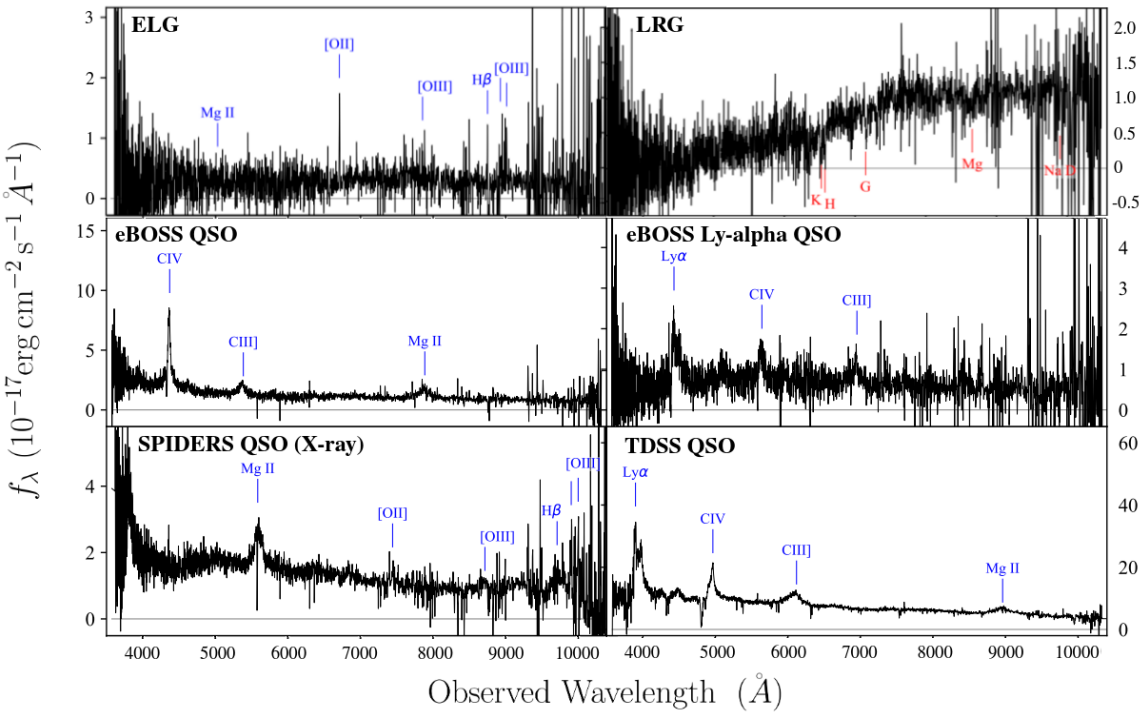


Figure 1.2: Examples of different galaxy spectra of the eBOSS survey. It is shown an emission line galaxy (*ELG*), a luminous red galaxy (*LRG*), a quasar from the core "cosmological" eBOSS sample (*eBOSS QSO*), a quasar selected at $z > 2.2$ for Ly-α forest studies (*eBOSS Ly-α QSO*), an X-ray emitting quasar selected by SPIDERS (*SPIDERS QSO*), and a TDSS-selected variable broad absorption line quasar (*TDSS QSO*). As one can see, different galaxies have different emission lines (labelled in blue), while QSOs have broad emission lines. The LRG shows absorption features (labelled in red) as a result of the metallicity of the stellar atmospheres contained in the galaxy. [Extracted from https://cds.cern.ch/record/2276002/plots].

Considering large enough scales so that the effects of gravity are negligible face to the cosmic expansion, all points (galaxies) in space are parting away from each other according to the relation

$$\boldsymbol{x}(t) = a(t)\,\boldsymbol{r}(t_0), \tag{1.2}$$

where the scale factor, $a(t)$, tells us how the distances in a model universe are evolving and it is obtained by solving, for example, Friedmann's equations (see Appendix B): different energy contents in the universe drive different time dependencies on $a(t)$. In Equation (1.2), $\boldsymbol{x}(t)$ refers

to Eulerian (physical) coordinates and $\boldsymbol{r}(t_0)$ are the Eulerian coordinates at some initial time t_0, i.e., $\boldsymbol{r}(t_0) \equiv \boldsymbol{q}$, where $\boldsymbol{q}$ are called Lagrangian coordinates that "each fluid element would attribute to itself", and thus are fixed in time. From Equation (1.2), we arrive at the velocity-distance relation of Hubble and Humason (1931):

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{v}_{\rm H}(t) = \frac{da}{dt}\boldsymbol{r}(t_0) = \frac{\dot{a}(t)}{a(t)}\boldsymbol{x}(t) \equiv H(t)\,\boldsymbol{x}(t). \tag{1.3}$$

Therefore, at recent times $t \approx t_0$ (today), $v_{\rm H} = H_0\, x$, which is the recession velocity due to the cosmic expansion, often called *Hubble flow*. The proportionality factor $H(t)$ is called Hubble parameter, and its present value H_0 is the Hubble constant.

If the galaxies' velocities are solely due to the cosmic expansion, then by joining Equations (1.3) and (1.1), we arrive at the distance-redshift relation that became known as the *Hubble's Law*[3]:

$$z_{\rm H} = \frac{H(t)}{c}\,\boldsymbol{x}(t). \tag{1.4}$$

This implies that the velocities in the Hubble flow, and therefore the redshift of galaxies, depend upon the proportionality parameter $H(t)$, which carries a time dependence that is determined by the energy content and geometry of the universe, as it is discussed in Appendix B. It also shows that the farther the galaxy, the larger its redshift will be. Finally, one comment for the inexperienced reader: because light propagates with a finite speed c, observing objects with large redshifts implies that we are observing them in the past (within our past light cone).

The cosmological principle As already mentioned, a central question to cosmology is determining the global structure of spacetime: by knowing its geometry and matter content, we can infer the causal structure of the (observable[4]) universe. From that we can either know the events which are causally connected or disconnected from specific spacetime points. This has profound consequences: from the need of adding a period of an extreme acceleration in the expansion in the far past to deal with the so-called *horizon problem*, thus having implications in the physics of the primordial universe that cosmology also aims at addressing, to the inference of what could happen to the fate of the universe given its current energy content.

Some works have shown that, due to the internal theoretical structure of general relativity, we cannot fully grasp the global structure of spacetime, no matter how much data we gather,

[3] One of the first people to introduce the term *Hubble's law* was Lemaître in a book review of 1950, suggesting also that "whereas he had more than a share in Hubble's constant, his role in the law was limited to a prediction without sufficient observational confirmation." (Kragh, 2018).

[4] We take a conservative stand of asserting cosmological inferences solely to the portions of the Universe we can observe. Attempts to extend observations beyond our cosmic horizon are rooted in non-empirical arguments (e.g. Ellis, 2011) and we shall not consider them.

leading to an empirical underdetermination of cosmological models (see for example Manchak, 2009, and references therein). It is then argued that the cosmological principle helps to partially break this inherently observational indistinguishability (Beisbart, 2009; Butterfield, 2014), but at the cost of picking a specific model on the basis of mathematical simplicity and aesthetic arguments (Clarkson et al., 2009).

To be more precise, by having two observational indistinguishable spacetimes $(\mathcal{M}, g_{\mu\nu})$ and $(\mathcal{M}', g'_{\mu\nu})$ we mean that, for every event $p \in \mathcal{M}$ and $p' \in \mathcal{M}'$, their observational past are the same. Hence, an observer (such as ourselves) looking at its past light cone cannot distinguish between $(\mathcal{M}, g_{\mu\nu})$ or $(\mathcal{M}', g'_{\mu\nu})$ on observational grounds. Because the difference between both spacetimes is contained within the cosmological model chosen, the models are said to be underdetermined.

There are different models, such as the inhomogeneous Lemaître-Tolman-Bondi (LTB) class of solutions of the Einstein's equations, or anisotropic metrics such as the Bianchi class, which can, in principle, fit the observational data. A common hypothesis is that physical laws as described in our local frame hold for all spacetime points. Even with this set of hypotheses, these models are all underdetermined according to Manchak's theorem. Within the homogeneous and isotropic class of solutions (see FLRW models in Section A.3), there are many possible models which depend upon the curvature of the universe (current constraints point to a flat geometry) and energy content (as in the ΛCDM model, with a current period of dark energy dominance). Thus, it is only with the assumption of the cosmological principle that we limit ourselves to a very stringent set of models that can be ruled out by observational data.

From a philosophical perspective, one may invoke the Copernican principle[5] to reinforce the cosmological principle, although this is an argument susceptible to a sceptical treatment (see Beisbart, 2009, Section 4.3). A more interesting view is the one exploited by Beisbart (2009) and Butterfield (2014), where the cosmological principle has a major role in breaking the underdetermination of cosmological models.

From physical principles, geometrical theorems allow us to taper the observational space, as pointed out in Butterfield (2014, Section 3.1.2): for example, isotropy everywhere implies homogeneity (Weinberg, 1972); another example is the one connecting spatially isotropic radiation to the implication of the cosmological principle (Ehlers et al., 1968; Stoeger et al., 1995).

From observational grounds, tests of the cosmological principle have a non-trivial nature: one first derives consequences of it, and then tries to find deviations from these expectations. A recurrent argument in favour of the cosmological principle are the CMB observations (Figure 1.1) that point to a high level of homogeneity (anchored by its tiny fluctuations around

[5] Recently, Camarena et al. (2021) found stringent constraints allowing for a very small deviation of the Copernican principle, within the context of a local ΛLTB spherical inhomogeneity.

the global average value of 2.7 K) and a seeming[6] isotropy; however, arguments of whether the inhomogeneities of small scales could impact the large-scale dynamics[7], and observations pointing to a possible break of isotropy, as it is the case of the so-called CMB anomalies (e.g. the cold spot and residual anisotropies), are becoming more and more frequent, forcing us to step back and rethink the validity of the cosmological principle. There are many interesting works pointing to a possible break of statistical isotropy: the detection of spatial variations of the H_0 parameter (Migkas et al., 2020; Luongo et al., 2021), the CMB anomalies and cosmic dipoles (see Perivolaropoulos and Skara, 2021, Sections III.2.4, III.2.5 and III.3, and references therein), and recent analysis performed by Appleby et al. (2021) of the SDSS-III BOSS galaxy catalogues, which points to either to systematics in the data, or for anisotropies in the matter field at low redshifts. If this is indeed the case, then we it will be necessary, for example, to model a directional dependence on cosmological parameters (for example, Heinesen, 2021, explores a model-independent luminosity-distance[8] expansion that will allow us to probe the directional dependence of cosmological parameters such as H_0).

To summarise: from the theoretical perspective, works going beyond the cosmological principle are beginning to gain more space in the cosmology literature, as the efforts are aligning to an increasing computational power (Macpherson and Heinesen, 2021). From the observational point-of-view, it demands a huge amount of data that will soon be provided by the upcoming galaxy surveys. In the meantime, while it is hard to settle down the matter of isotropy, and thus claiming the cosmological principle from observational evidences, in the remaining of this work it will be assumed from the arguments of mathematical simplicity, aesthetics and as a way to break observational indistinguishability of cosmological models.

The ΛCDM cosmological model The mainstream cosmological model, dubbed ΛCDM, is based on a set of assumptions. Regarding its cornerstone principles, we have the general theory of relativity dictating gravitational collapse, quantum field theory devising quantum fluctuations during a period of inflation, in the very early universe and the cosmological principle (statistical spatial isotropy and homogeneity) that enables the use of FLRW-like metrics. Consequently, it consists of extrapolating local physical laws to cosmological scales (Smeenk, 2013).

[6] If we observe an isotropic CMB, then we need to invoke the Copernican principle to extend this observational isotropy to the whole observable universe, as we cannot move to another position and redo a CMB observation.

[7] This is known as the "fitting problem" in cosmology (Ellis and Stoeger, 1987): assigning macroscopic variables and field equations to the spacetime microstate, that is, the problem of formulating a large-scale theory (macrostate) from a realistic lumpy (microstate) universe with local inhomogeneities.

[8] The luminosity distance is a key quantity in analyses that aim at constraining cosmological parameters with data such as quasars and supernovae, but this is completely model dependent as it requires adopting a functional form for $H(z)$ (see Appendix A.3.5). Heinesen's proposal is to use a generic expansion that does not rely on any cosmological model, including the FLRW metric assumption.

The FLRW metric, and the assumption of general relativity, set the basis for the geometry and dynamics of spacetime. However, as implied by the Einstein's equations (see Equation A.2), we still need to describe the energy content. Due to the symmetries enforced upon the FLRW metric (see Section A.3), the matter is described by a perfect fluid, which is characterised by its energy density and pressure. The ΛCDM model describes an approximately flat universe composed by radiation (photons and neutrinos), baryonic and cold dark matter (CDM), and a cosmological constant Λ[9]. The solutions of the Einstein's equations show that we can work, at least in the background level, assuming a superhot and dense stage where radiation is the dominant component, followed by a period of matter dominance, with a latter epoch whose evolution is mostly dominated by a dark energy fluid.

The dark energy fluid, described by the cosmological constant in the ΛCDM model, leads to an accelerated expansion. It started plunging itself into the mainstream model after some large-scale structure studies (e.g. Strauss and Willick, 1995; Ostriker and Steinhardt, 1995), but entered in it for good after the works of Riess et al. (1998) and Perlmutter et al. (1999), which indicate a late-time cosmic acceleration. We will not be able to do justice to the whole story behind the development and observational evidences for the ΛCDM model. We will do even less justice to alternative models, which range from attempts to modify Einstein's gravity to explain the apparent late acceleration, to breaks in the cosmological principle. In this overview, we will briefly highlight what characterises the model, and the current constraints given by the Planck collaboration (Aghanim et al., 2020). For an in-depth discussion on these topics, we refer the interested reader to Faraoni and Capozziello (2011); Smeenk (2013); Bull et al. (2016); Perivolaropoulos and Skara (2021), and references therein.

The ΛCDM model can be parametrised in terms of 6 cosmological parameters, namely the baryon and cold dark matter density, ω_b and ω_c, the angular acoustic scale θ_*, the reionisation optical depth τ, and the amplitude and scalar spectral index of primordial density fluctuations, respectively A_s and n_s. By extrapolating local laws to cosmological scales, these parameters become tightly constrained by CMB observations, and after assuming the ΛCDM model to describe the late-time evolution of the universe, it is possible to infer yet another parameter: the Hubble constant today, $H_0 = 67.36 \pm 0.54$ km/s/Mpc, which gives an estimate for the age of the universe within this particular model. Most recent constraints for these parameters are given by the Planck collaboration (Aghanim et al., 2020), and are quoted in Table 1.1.

For being in accordance with current observations, accounting for structure formation,

[9]The cosmological constant was originally introduced by Einstein as a modification of the field equations to reach a static universe solution. However, the amount of energy density in the form of a cosmological constant must be fine-tuned to the density in the form of ordinary and dark matter so that the static solution is achieved. The most minute unbalance between the cosmological constant and the other forms of matter results in a dynamical solution (Smeenk, 2013).

Table 1.1: Cosmological parameters resulting from the CMB power-spectrum analysis presented in Aghanim et al. (2020), defining the so-called base-ΛCDM model.

Parameter	TT,TE,EE+lowE+lensing (68% limits)
ω_b	0.02237 ± 0.00015
ω_c	0.1200 ± 0.0012
$100\theta_*$	1.04092 ± 0.00031
τ	0.0544 ± 0.0073
$\ln(10^{10}A_s)$	3.044 ± 0.014
n_s	0.9649 ± 0.0042

and for giving a coherent cosmic picture, it is safe to say that the ΛCDM model has been a success. Finally, its many internal tensions and challenges are an opportunity to establishing a new cosmological picture in the future, disclosing hitherto unforeseen phenomena.

Inflation and initial conditions The final ingredient missing from the picture concerns the initial seeds to hatch structure formation. The most studied way to address this issue, explaining the origin of structures (such as galaxies, and the CMB temperature fluctuations — see Figure 1.1), and answering many question raised by the time it was conceived, is the inflationary model.

The intrinsic random nature of the quantum processes that took place at the primordial universe, and that, as we believe, gave origin to the seeds which set the beginning of structure formation, leaves distinct imprints on the late-time matter field. These seeds correspond to fluctuations around what would be a homoegeneous background and they can be decomposed, as any other signal, into an infinite sum of periodic waves with specific wavelengths and amplitudes. Just like the mean and the variance are used to infer properties of a sample, in cosmology the statistical description is given by the same quantities, with the variance being given by a slightly more general concept: the power spectrum, which codifies the power and the wavenumber dependency of those waves (see Figure 3.1) by extracting information from pairs of objects (such as cold dark matter particles or galaxies, for example).

Inflation predicts, among other appealing features, that the spectrum of primordial density fluctuations is nearly scale invariant, that is, $n_s \sim 1$ (see Table 1.1 for the best-fitting value). This implies that when the perturbations are decomposed into its Fourier counterpart, every Fourier mode (or wavelength) contributes equally to the density fluctuations. Harrison (1970), Peebles and Yu (1970), and Zeldovich (1972) argued in favour of such scale-invariant spectrum "because the amplitude did not diverge at small wavelengths, which would produce too many black holes, or at large wavelengths, which would produce too much distortion in the cosmic background radiation" (Ostriker and Steinhardt, 1995).

1.1 Goals of this thesis

Inflation and primordial non-Gaussianities As already pointed out, we will be interested in the effects that early universe assumptions, such as those imposed during an inflationary stage, have on the 3D distribution of galaxies, and on other dark matter tracers, such as halos, as well (see Section 2.1). Apart from the nearly scale-invariant spectrum, another key inflationary observable concerns the levels of non-Gaussianities present in the high-order correlation functions of the primordial perturbations, manifested at late (in the LSS) and early times (in the CMB), with the prominent feature of a scale-dependent bias correction to the 2-point correlation function of LSS tracers (e.g. galaxies and dark matter halos). In Chapter 4 these concepts are developed in more details. One of our goals, still in development, is to implement the effects of this scale-dependent bias in a fast halo generator (Voivodic et al., 2019). With that we plan to investigate primordial non-Gaussianities in galaxy surveys, by looking at realistic galaxy mocks.

Relativistic effects As we will also discuss in Chapter 4, this scale-dependent bias effect is manifested at scales large enough for the Newtonian treatment to be adequate. This led us to our second goal: how does the relativistic treatment of cosmological observables impact our interpretations of the 3D distribution of galaxies? While Equation (1.1) holds for low redshifts, and small velocities, Equation (1.4) is valid if we neglect the peculiar motion of galaxies. However, the truly observed redshift depends on many different effects not accounted by the simplistic assumption that a galaxy's redshift traces directly the cosmic expansion. This argument would hold in a homogeneous universe; however, departures from this assumption due to structure formation appear in the distance-redshift relation, Equation (1.4), in the form of the so-called redshift-space distortions (RSDs). In Chapter 5, we present the corrections that will appear not only to the redshift, but also to the counts of galaxies (an important cosmological observable), from considering the fully general relativistic framework to interpret redshift measurements. This leads to the so-called relativistic effects (RE), that is a whole independent branch of observational cosmology. At first sight they might seem only a source of error, but their expansion in Legendre multipoles ℓ turned out to be a striking cosmological probe, providing a novel observational method. RE appear as corrections to the standard (Newtonian) modelling of redshift-space variables and multipoles, giving a significant contribution at ultra-large scales. Whilst seen as a source of systematic by hampering the primordial signal (Lopez-Honorez et al., 2012; Camera et al., 2015; Renk et al., 2016; Wang et al., 2020), their unique signature, with no Newtonian counterpart, is manifested through the appearance of odd multipoles ($\ell = 1, 3, \ldots$), in addition to the even ones generated by redshift-space distortions in the purely non-relativistic approximation. This feature is, by itself, a new source of observables allowing novel probes of gravity (Renk et al., 2016), such as tests of the equivalence principle on cosmological scales

(Bonvin and Fleury, 2018) and the gravitational redshift effect (McDonald, 2009; Bonvin et al., 2014; Breton et al., 2018; Saga et al., 2020). Moreover, because they bear a strong dependence on the Weyl potential, they provide an alternative test for theories of gravity (e.g., Lombriser et al., 2013; Baker and Bull, 2015), while the dependence on astrophysical parameters like the magnification and evolution biases opens a new window to a better understanding of the LSS. This study is described in Chapter 6, with a preliminary theoretical background given in Chapter 5.

Calibrating photometric redshifts Finally, the different experiments used to map the large-scale distribution of matter, described in Chapter 2, will open up a range of possibilities to explore the multiple-tracers technique, and can complement each other to overcome particular barriers. While the redshift-inference method based on the full galactic spectra (see Figure 1.2 and the discussion below it), known as spectroscopy deliver us high-precision redshifts, they are too time consuming to provide dense galaxy catalogues. An alternative way to obtain redshifts is through the photometric redshifts technique. However, as we point out in Section 2.2.2, they have a non-negligible error that may compromise different cosmological goals. For that reason, in Chapter 7 we present a method to reduce these errors by combining a photometric survey with neutral hydrogen maps obtained through the intensity mapping technique (Section 2.3), via what is called *clustering redshifts*.

1.2 Notation and conventions

Time derivatives Primes denote derivatives with respect to the conformal time τ, whereas dots denote derivatives with respect to the proper time t: $f' = \partial_\tau f$, and $\dot{f} = \partial_t f$.

Metric signature We will adopt the signature $(-,+,+,+)$.

Spacetime indices Unless otherwise stated, Latin indices will run over the spatial dimensions, with Greek indices running through the time and spatial coordinates (in our case, 4 dimensions).

Fourier transform convention

$$f(\boldsymbol{x}) = \int \frac{\mathrm{d}^d k}{(2\pi)^d}\, \mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{x}} f(\boldsymbol{k}), \quad f(\boldsymbol{k}) = \int \mathrm{d}^d x\, \mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{x}} f(\boldsymbol{x}) \tag{1.5}$$

where d is the spatial dimension.

Spherical harmonics normalisation

$$Y_{\ell m}(\theta, \varphi) = \sqrt{\frac{2\ell + 1}{4\pi}} \sqrt{\frac{(\ell - m)!}{(\ell + m)!}} \mathcal{P}_{\ell m}(\cos\theta)\, \mathrm{e}^{im\varphi} \tag{1.6}$$

where $\mathcal{P}_{\ell m}$ are the associated Legendre polynomials, connected to the Legendre polynomials $\mathcal{L}_\ell(\mu)$, when $m = 0$: $P_{\ell 0} \equiv \mathcal{L}_\ell$.

Chapter 2

Cosmological Surveys: Mapping the Large-Scale Structure

In this chapter we describe the basics of cosmological surveys. We inspect the physical objects utilised in cosmology, such as galaxies (of different types) and the integrated emission of the neutral hydrogen atom, and how they are mapped by these surveys. This motivates the set of observables that we define and deal in details in Chapters 5 and 6, which are built directly from these objects. These observables are then employed to build our past light cone. We also introduce in this chapter the key quantities and definitions needed for the work developed in Chapter 7.

2.1 General considerations

There are different ways to trace the matter distribution in the Universe. The most direct one is to collect the light emitted by extragalactic sources such as galaxies, gas or the microwave photons from the cosmic microwave background (CMB). Another way to infer the matter field is through the deflection of light produced by the phenomenon of gravitational lensing. Finally, we are settling into a new era of astronomy where it is possible to go beyond the electromagnetic spectrum and observe, instead of electromagnetic waves, gravitational waves.

So far, the best constraints we have on the mainstream, ΛCDM, cosmological model come from the exquisitely detailed CMB maps (Aghanim et al., 2020), only possible through the advance of spacecraft and detectors technologies. However, the amount of information available in the two-dimensional CMB maps is limited: if we consider an angular resolution of ~ 10 arcmin, which corresponds to angular scales characterised by the multipole $\ell \sim 2200$, then have only $\sim 10^6$ independent modes available in an area of 2×10^4 deg^2, corresponding to $f_{\rm sky} \sim 0.5$ obtained after removing a stripe containing our Galaxy.

Another avenue to map the matter distribution consists in observing galaxies, that comes

in different shapes and sizes, and building catalogues that can be subsequently used to map the three-dimensional distribution of matter. These observations will enable constraints competitive to, and eventually surpassing, those of CMB (Alvarez et al., 2014; Abazajian et al., 2015; Karagiannis et al., 2018), as the 3D maps provided by these surveys carry much more information than the CMB's 2D maps: the volume covered by current and forthcoming surveys ranges $V \sim 10 - 100\ \mathrm{Gpc}^3$, so the amount of independent Fourier modes are of the order of $10^7 - 10^8$.

However, to extract information from galaxy surveys, one must understand their basic constituents: the galaxies themselves. Making a long story short, redder galaxies have low stellar-formation rate, are older and more massive than the blue galaxies that have intense stellar formation, are younger and not too massive. Therefore, these two plain-to-see classes of galaxies will trace the matter distribution differently. For example, within the class of blue galaxies, we have those that possess starburst regions, i.e. places where massive stars are being constantly made. To form these stars, a lot of intergalactic gas is being consumed and, after these young and massive stars are made, they ionise the gas in this star-birth region, giving rise to strong emission-line features in the galactic spectrum (e.g. uppermost spectrum shown in Figure 1.2; Sparke and Gallagher, 2007). These emission-line galaxies (ELGs) are easily identified, and are usually located in regions with a high density of intergalactic gas, orbiting a more massive and older red galaxy: in a cluster of galaxies, one will usually encounter a massive red galaxy near the centre, with bluer galaxies in the outskirts of the cluster; therefore, these two classes of objects trace differently the matter distribution. Finally, two other classes of galaxies commonly used in cosmology are the luminous red galaxies (LRGs) and quasars (QSOs)[1].

By observing a galaxy cluster, we can see that it is composed of intergalactic gas, which can be traced with X-ray observations, and of the stars within the cluster's galaxies. In 1933, the astronomer Fritz Zwicky applied the virial theorem to estimate the mass of the Coma cluster (Zwicky, 1933, 2009). Using the galaxy clusters published by Hubble and Humason (1931), he was able to observe that the dispersion velocity of eight galaxies in the Coma cluster was much larger than the expected: the total amount of gas and stars within the cluster were not enough to make the Coma cluster a bounded gravitational system. This led to the conclusion that galaxy clusters have a major component that is not in the form of visible objects, which is known by dark matter (Bertone and Hooper, 2018, Chapter III).

A more recent and striking evidence for dark matter is the Bullet cluster (Clowe et al., 2006), where the distribution of stars and galaxies do not follow the distribution of gas. If gas

[1] LRGs are commonly found around redshift $z \sim 2$, corresponding to an universe of ~ 3 Gyr old within the ΛCDM model, and lacking young blue stars, indicating that they must have formed at least 1 billion years prior to the observation, i.e. around redshift $z \sim 3$ (Sparke and Gallagher, 2007, p. 239). QSOs, on the other, are active galactic nuclei accreting gas through the host galaxy's supermassive black hole. They are the most luminous objects in the Universe, fact that make them great to map high redshifts, that is, a younger universe.

was the largest part of the cluster's mass, then these objects should follow the gravitational potential of the gas. However, with the use of a weak lensing map[2], it was evident that galaxies and stars were trapped to a gravitational potential created by something that cannot be electromagnetically observed, i.e. dark matter.

Another crucial evidence for dark matter resides at galactic scales. We will not enter into the details, but it is relevant to point out that dark matter is also needed to explain the rotation curve of individual galaxies (see Chapter IV of Bertone and Hooper, 2018, and references therein). Finally, at cosmological scales, the temperature spectrum of the CMB radiation, in the context of a ΛCDM universe, points to a large fraction of dark matter.

Because baryons are not the major part of the matter content, the evolution of gravitationally bounded structures in the universe will be dictated by the dark matter field, and it is this field what we want to trace and understand. When said, in the beginning of this chapter, that we want to trace the large-scale distribution of matter, it was the dark matter we were referring to. Therefore, we want to predict how dark matter will collapse given what we know about the observable universe. Our job is then to connect the spatial distribution of galaxies with the dark matter distribution.

An approach for predicting and interpreting analytically the statistics of dark matter and galaxy clustering is the so-called *halo model*, which we give an overview in Appendix C. It assumes that all the mass in the universe is in the form of dark matter halos of different sizes, which are distinct bounded units of some mass resulting from a collapse that virialised (Cooray and Sheth, 2002). For example, a galaxy like the Milky Way can be thought to be embedded in a small dark matter halo, with the baryonic matter distributed inside this halo. Going to larger scales, a cluster of galaxies can also be seen as a large dark matter halo populated by galaxies, with each galaxy type having a different bias. Finally, halos of different masses, and other secondary properties, will trace the dark matter field differently. Therefore, we will also need to give them a bias prescription that will depend on how we define these halos.

To illustrate this discussion, in Figure 2.1 we show two slices of a past light cone, where the Cartesian coordinate z can be associated with the redshift. Each point in the figure represents a distinct dark matter halo, and the colours a different mass bin. As we can see, the less massive halos (*blue*) are easily found up to larger radial distances, therefore, up to higher redshifts. The intermediate mass halos (*red*) in the figure can still be found at larger redshifts, but with a smaller abundance, while the high mass halos (*orange*) are more scarce, but becomes more abundant at smaller radial distances. This happens since halos are formed from the gravitational collapse of dark matter particles, so massive halos are a result of a merger of less massive halos.

[2]The idea of estimating the mass of a galaxy cluster through gravitational lensing was first introduced by Zwicky (Zwicky, 1937).

Therefore, they have a larger formation timescale. Another feature is that more massive halos tend to be more clustered than the less massive ones. That is, they have a larger bias[3].

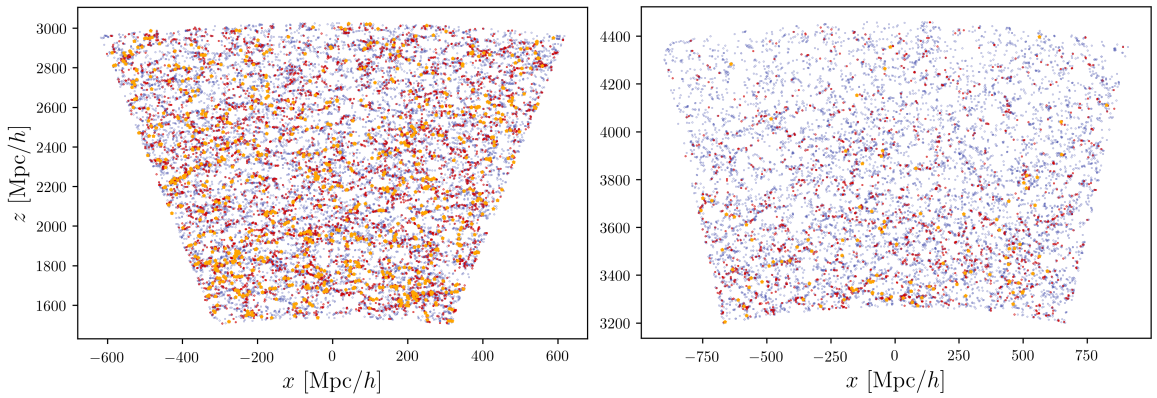


Figure 2.1: Dark matter halos of a dark matter only N-body simulation. Each dark matter particle in the simulation has a mass of $M \approx 2.6 \times 10^9\, M_\odot$. Blue points represent halos with at least 600 dark matter particles, and at most 2400 particles, i.e. with mass M_1 in the range $[1.6, 6.3) \times 10^{12}\, M_\odot$. Red points represent halos with more than 2400 particles, but less than 9600 particles, i.e. $M_2 \in [6.3, 25.4) \times 10^{12}\, M_\odot$, and the orange points are the halos with at most 19200 particles, i.e. $M_3 \in [2.5, 5.1) \times 10^{13}\, M_\odot$. Both panels show the same past light cone, but we divided it into a low (*left*) and high (*right*) redshift part for a better visualisation of the redshift evolution of each halo population.

In what follows, we refer to different types of galaxies (such as ELGs, LRGs and QSOs), or the dark matter halos they inhabit, as biased tracers of the dark matter field. With the exception of dark matter halos, different tracers can be observed by different types of surveys, which can be spectroscopic, photometric, or radio-like. There is a rich history behind the development of galaxy surveys (see for example Martínez and Saar, 2001), but we shall jump to the current state-of-the-art, mentioning the upcoming Vera Rubin/LSST, Euclid, and SKA surveys. Their main features are shown in Table 2.1.

Table 2.1: Main features of upcoming surveys. Based on Table 2 of Alvarez et al. (2014). SKA phase-2 (P2) data was extracted from Bull et al. (2015a).

	LSST	**Euclid**	**SKA (P2)**
Survey type	Photometric	Photo + Spec	21 cm
Base	Ground	Space	Ground
Redshift	$z < 3$	$z < 3$	$0.18 < z < 1.84$
Survey area (deg^2)	20 k	15 k	30 k
Number of objects	$\sim 10^9$ gals.	$\sim 10^7$ redshifts $\sim 10^9$ photo-z	$\sim 10^9$ gals.

[3]For a formal introduction of the bias and its relation to clustering, see Section 3.1, and Equation (3.22).

2.2 Galaxy redshift surveys

When we look at the night sky, we are performing a primitive type of astronomical observation: we perceive the light emitted by a celestial object on the 2-sphere and draw a 2D map. While our prior experiences allow us to infer that the Moon is closer to us than some random star, we cannot construct a 3D picture of the stars being observed.

As suggested by the name, a galaxy survey detects the integrated light emitted by galaxies; that is, the light coming from all stars within the galaxy. Either by decomposing the light into a spectrum (spectroscopy), or by selecting specific features in different band-pass filters (photometry) we can build a 3D map of the Universe with galaxy redshift surveys. The basic constituents of a galaxy redshift survey are the objects detected by whatever method the survey is adopting, at a given angular position on the sky, set by the observer, and the redshift inferred from the spectra (or pseudo-spectra) of these objects: just like the visual system is capable of creating a 3D representations of the world from our past experiences, the third dimension that extends the celestial sphere is inferred from our understanding of the Doppler effect, Equation (1.1), stellar astrophysics (to understand the galaxy spectra), and our theoretical modelling of an expanding universe that allow us to convert redshifts to distances through a redshift-distance relation, Equations (1.4) and (A.45). As we briefly outlined in Chapter 1, ascribing the redshift of galaxies to the stretch of spacetime allows us to set the observational depth of each object detected by the survey.

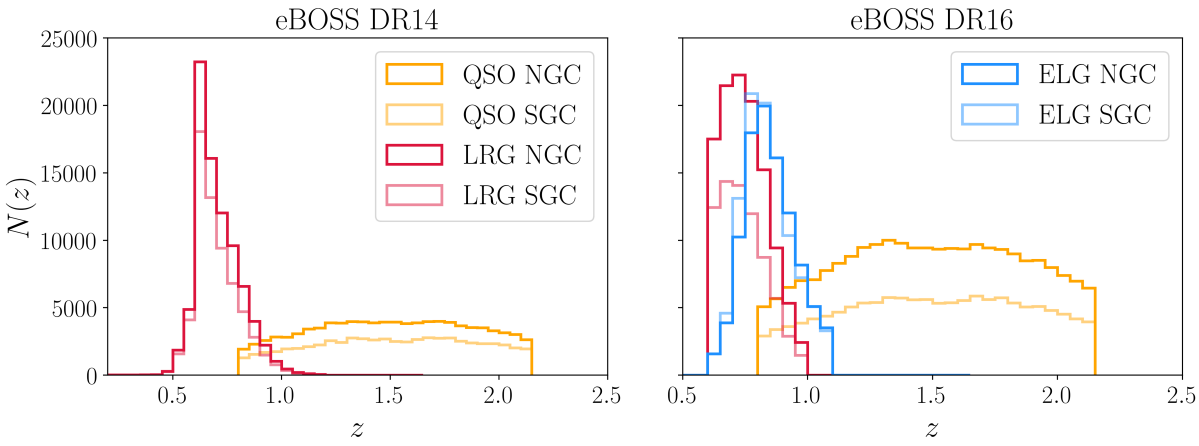


Figure 2.2: Galaxy count of the eBOSS survey for three different tracers: quasars (QSOs, *orange*), luminous red galaxies (LRG, *red*), and emission line galaxies (ELG, *blue*). NGC refers to the North Galactic Cap, and SGC to the South Galactic Cap. *Left*: objects contained in the data release 14. *Right*: final data release, DR16. As we can see, red galaxies are more abundant at lower redshifts, while quasars are observed up to larger distances. Redshift bins of $\Delta z = 0.05$ were considered. Data extracted from https://data.sdss.org/sas/dr16/eboss/lss/catalogs/.

Galaxy surveys are characterised, among many other technical properties, by the galaxy counts per unit redshift bin $\mathrm{d}z$ and solid angle $\mathrm{d}\Omega$, $n_g(z) = \mathrm{d}^3N(\hat{\boldsymbol{n}}, z)/\mathrm{d}z/\mathrm{d}\Omega$, where $\hat{\boldsymbol{n}} = (\theta, \varphi)$

is the direction of observation. In Figure 2.2, we show the galaxy count, for three different tracers, for the data releases DR14 and DR16 of the eBOSS survey. We chose redshift bins of $\Delta z = 0.05$.

Table 2.2: Expected set-up for photometric surveys. For Euclid and LSST, we extract the numbers from Ballardini et al. (2019) and references therein, apart from $z_{\rm max}$ for the LSST case. In particular, for the LSST case, we allow for three different amplitudes of photo-z error σ_0, corresponding to an optimistic scenario (0.02), the standard value assumed for luminous red galaxies (LRGs, 0.03) and the conservative expectation (0.05). $\Omega_{\rm sky}$ refers to the survey area (in deg^2), $n_{\rm sources}$ gives the number of sources per arcmin^2 within a redshift range $z_{\rm min} < z < z_{\rm max}$, and z_0, α and β are used for the overall redshift distribution of sources, Equation (2.1).

Survey	$\Omega_{\rm sky}$ (deg^2)	$n_{\rm sources}$	$z_{\rm min}$	$z_{\rm max}$	z_0	α	β	σ_0
LSST	13800	48	0.2	3	2	0.9	0.28	0.02, 0.03, 0.05
Euclid	15000	30	0	3	2	1.5	0.636	0.05

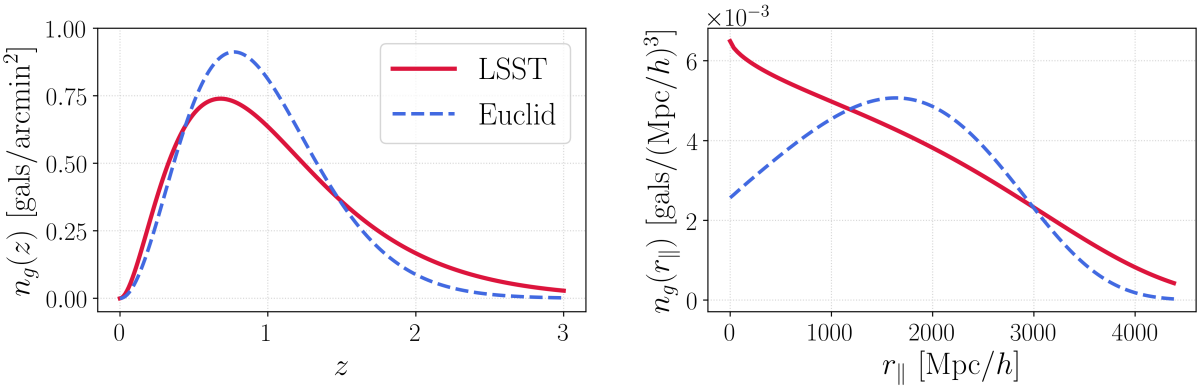


Figure 2.3: Galaxy number densities as a function of redshift (*left*) and radial comoving distance (right) for an Euclid (*blue-dashed*) and LSST-like (*red-solid*) photometric surveys. Parameters employed in the computation of these curves are shown in Table 2.2. The conversion between arcmin to Mpc/h are done with the angular diameter distance relation (Equation A.47).

Not distinguishing between the different tracers available in a survey, we assume the overall redshift distribution to be given by (Ma et al., 2006):

$$n_g(z) = A\, z^\alpha \, \exp\left[-\left(\frac{z}{z_0}\right)^\beta\right], \tag{2.1}$$

with units of galaxies/arcmin^2. This distribution is what a spectroscopic survey would obtain. For example, in Figure 2.2 each population have a different redshift distribution $n_x(z)$ (x = QSO, LRG, ELG), which is a normalised fit for the histograms shown. In Equation (2.1), the parameters α,β and z_0 are chosen for each specific survey, and are given in Table 2.2. The constant $A = \beta\, z_0^{-(\alpha+1)}/\Gamma\left[(\alpha+1)\beta^{-1}\right]$ is such that $n_g(z)$ is normalised. For the case of LSST, for example, Equation (2.1) is a general expression for all types of tracers included in the survey,

that is, without differentiating blue and red galaxies, for example. From the redshift-distance relation $\mathrm{d}r = c\,H^{-1}(z)\,\mathrm{d}z$, Equation (2.1) can further be translated to a function of the comoving radial distance $r_\parallel$ as $n_g(r_\parallel) = c^{-1}\,H(z)\,n_g(z)$, which is now in units of galaxies/(Mpc/h)/arcmin^2. We depict both $n_g(z)$ and $n_g(r_\parallel)$ in Figure 2.3.

Finally, each galaxy population in the survey will be connected to the dark matter field by some biasing scheme. Again, considering a single population of galaxies, for the photometric samples of the LSST and Euclid-like surveys, the linear biases b_1 assumed are given by (see Ballardini et al., 2019, and references therein)

$$b_{g,1}(z) \equiv b_g(z) = \begin{cases} \sqrt{1+z} & \text{(Euclid)} \\ 0.95\,D^{-1}(z) & \text{(LSST)} \end{cases}, \tag{2.2}$$

and are shown in Figure (2.4).

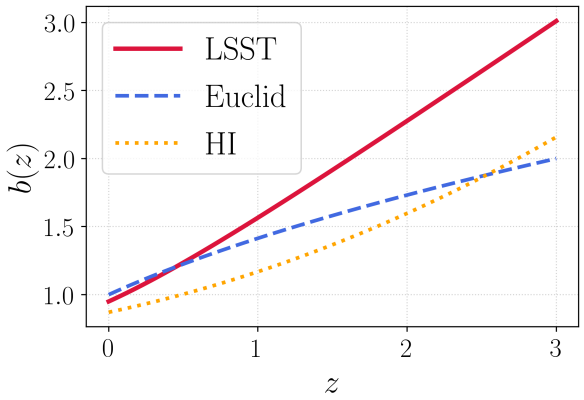


Figure 2.4: Galaxy biases as a function of redshift for an Euclid (*blue-dashed*) and LSST-like (*red-solid*) photometric surveys. HI refers to a neutral hydrogen intensity mapping survey, discussed in Section 2.3.

2.2.1 Spectroscopic surveys

Spectroscopic surveys deliver to us precise redshift measurements (with error smaller than 10^{-4}). The light emitted by a galaxy is directed from the telescope's focal plane to the spectrograph, which is designed to decompose the galaxy's light into its spectrum, with optical fibres. For example, SDSS-I/II were able to receive 320 fibres from a single plate[4]. With the procedure described in Chapter 1 (see discussion around Figure 1.2), the redshifts can be obtained with high precision. However, to collect the spectra, each plate demands an exposure time. Therefore, the acquisition of spectra for millions of galaxies is a time consuming process.

[4] https://www.sdss.org/dr12/help/glossary/

Nonetheless, interesting scientific goals can be achieved with the use of the 3-dimensional maps provided by these surveys. For example, the Doppler magnification effect κ_v, which arises from local contributions of redshift-space distortions to the total magnification $\kappa = \kappa_g + \kappa_v$ of galaxies (where κ_g is the usual convergence term coming from weak-lensing contributions), requires objects to be separated by $\Delta z \simeq 0.02$ in redshift. This means that it is important to have an accurate method such as the spectroscopic one (Bacon et al., 2015). In Coates et al. (2021) it is shown, in Figure 5, the magnitude of this correction with an ensemble of relativistic N-body simulations: it will have a major impact on large scales and at low redshifts, where the contributions from peculiar velocities are larger. Because this effect depends on the velocity field of galaxies, it can provide information on dark energy and modified gravity (Bacon et al., 2015). The search for other effects constrained via redshift-space distortions (RSD; see, for example, Chapter 6) can be limited without the accuracy of spectroscopic redshifts, although some works are developing tools to deal with this limitation (for example, Zhang et al., 2021, and references therein).

In the next section we overview an alternative route to extract 3-dimensional galaxy clustering information that is faster and, therefore, provides a larger number of objects.

2.2.2 Photometric surveys

Photometric surveys are a less expensive route to provide 3-dimensional maps of the Universe. While spectroscopic surveys offer high-precision redshifts, their downside is the amount of time needed to get the spectrum of each surveyed object in order to obtain the redshifts. This leads to a sparse sample and hence a weaker signal. To circumvent this issue, one may carry out a photometric survey, where the spectra of objects are estimated via strong spectral features (e.g. the 4000 Å break, common in the spectra of elliptical galaxies due to the presence of metals in the atmosphere of low mass stars that absorb light with smaller wavelengths — Sparke and Gallagher, 2007; Mo et al., 2010). The byproduct of these surveys is a dense catalogue, containing millions (or even billions) of objects with photometric redshifts (photo-zs), providing a larger signal-to-noise for cosmological studies.

The basic idea behind these surveys is to observe the light emitted by galaxies in different broad-band filters[5]. We can extrapolate the colours/magnitudes observed in these bands for local galaxies ($z \ll 1$), and from the relation (see Sparke and Gallagher, 2007, p. 340)

$$m_{\rm BP} = M_{\rm BP} + 5\log_{10}\left(\frac{d_{\rm L}}{10\,{\rm pc}}\right) + k_{\rm BP}(z) + e_{\rm BP}(z) \tag{2.3}$$

[5]For the use of narrow-band filters, see for example Silva (2021), Chapter 4.

we can estimate what is the apparent magnitude of a certain type of galaxy if it was located at a redshift z. In this expression, m_{BP} is the apparent magnitude observed in a band-pass (BP) filter, M_{BP} is the absolute magnitude the galaxy would have if seen from a distance of 10 pc, emitting as it would at $z = 0$, and $k(z)$ and $e(z)$ are known as the k-correction and evolutionary terms (for details, see Sparke and Gallagher, 2007, p. 341).

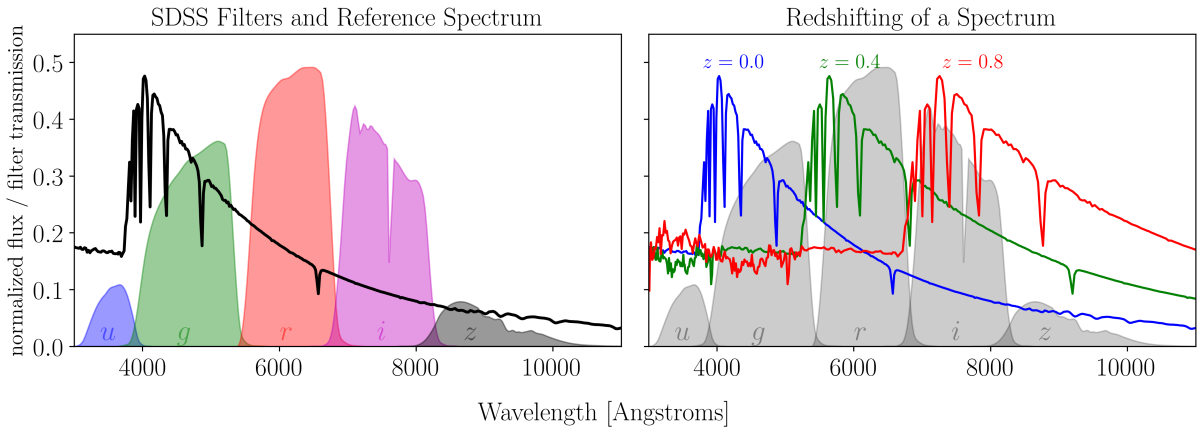


Figure 2.5: *Left*: The reference spectrum of the star Vega, with the 5 SDSS filters: *ugriz*. *Right*: the expected Vega's spectrum at redshifts $z = 0.4$ (*green*), and $z = 0.8$ (*red*). Notice that the reference spectrum is bright in the u and g filters, but dimmer in the i and z. At redshift $z = 0.8$, the opposite happens. [Extracted from https://ogrisel.github.io/scikit-learn.org/sklearn-tutorial/tutorial/astronomy/regression.html].

By looking at different types of galaxies for redshifts $z \ll 1$, we can infer what their spectra would look like at higher redshifts. As an example we show a reference spectrum of the star Vega on the right panel of Figure 2.5, with the SDSS *ugriz* filters shown in different colours. On the right, the reference spectrum is at redshift $z = 0$, and what we would expect to see if Vega was at redshift $z = 0.4$ (*green*), and $z = 0.8$ (*red*). The same method applies to galaxies (see for example Mo et al., 2010, Section 2.6.2).

The method described here is only one among many other ways to estimate the redshift of galaxies through photometry. We refer the interested reader to Silva (2021) for a complete discussion.

Despite fast and capable of surveying a large number of galaxies, photometric surveys suffer from the so-called photometric redshift (photo-z) uncertainties. These can be larger for a lower number of filters, as we will not be able to discriminate among different spectral features, as well as for broader filters, since the ability to trace different absorption and emission lines is weakened. Also, if we attempt to observe galaxies with few strong spectral features, for example star-forming galaxies with a less pronounced, or absent 4000 Å break, our redshift estimates will be degraded (Silva, 2021, p. 91). Hence, photo-z uncertainties are a key issue for cosmological studies based on photometric surveys, as the larger the redshift error, the larger is the blurring of structures along the radial direction due to the uncertainty. This can severely

bias cosmological parameter constraints, especially when the effects of certain cosmological features are degenerated with the effects of photo-z uncertainties on the summary statistics used to perform the statistical inference.

These photo-z uncertainties can be modelled trough the probability distribution $P(z_{\rm ph}|z)$ of assigning a photo-z $z_{\rm ph}$ to a galaxy whose true (spectroscopic) redshift is z. We choose a normal distribution with the standard deviation given by the redshift scatter, $\sigma_z = \sigma_0(1+z)$ (Lima and Hu, 2007):

$$P(z_{\rm ph}|z) = \frac{1}{\sigma_z\sqrt{2\pi}}\exp\left[-\frac{1}{2}\frac{(z_{\rm ph}-z)}{\sigma_z^2}\right]. \tag{2.4}$$

These uncertainties are manifested in the survey selection function $\phi(z)$, which can be cast into a specific photo-z bin i as

$$\phi_i(z) = n_g(z)\int_{z_{\rm ph}^i}^{z_{\rm ph}^{i+1}} {\rm d}z_{\rm ph} P(z_{\rm ph}|z). \tag{2.5}$$

Once again, we perform a change of variables in Equations (2.4) and (2.5) to write them in terms of the comoving distance, $\phi_i(r_\parallel)$, with the redshift scatter being straightforwardly given by $\sigma_\parallel = c\,H^{-1}(z)\,\sigma_z$. Equation (2.5) is depicted in Figure 2.6 for the survey specifications shown in Table 2.2. We consider three different scattering amplitudes σ_0 for the LSST-like survey[6], while keeping the one for Euclid fixed, as shown in Table 2.2. In this Figure, each coloured curve represents a different redshift bin ϕ_i. For example, we show 10 redshift bins, with the purple shaded region representing ϕ_2, which is centred at $z \approx 0.6$, for both surveys.

2.3 21-cm intensity mapping

Mapping the large-scale structure with the 21-cm line of neutral hydrogen is a great example of quantum mechanics applied to astronomy. Because hydrogen is the most abundant element in the universe, it is present in stars, inside every single galaxy, in the interstellar and intergalactic medium, and even when there were no galaxies formed; therefore, it is worthwhile to explore methods to detect it.

After the decoupling (CMB, $z \sim 1100$), the universe is expected to be neutral until $z \sim 7$, period at which the first stars are formed, reionising the universe. After this period ($z \lesssim 6$), called *reionisation*, it is expected that neutral hydrogen (HI) atoms can only be found within galaxies, specifically in dense gas clouds where it can be shielded from the ionising gas (Bull et al., 2015b; Durrer, 2020, Section 8.5). Even though in these regions we do not

[6]For populations a strong 4000 Å break (e.g. red galaxies), $\sigma_0 = 0.03$, while $\sigma_0 = 0.05$ is expected for galaxies with less pronounced features (Chavez, 2014). The survey's goal is set to $\sigma_0 = 0.02$ (Alonso et al., 2017).

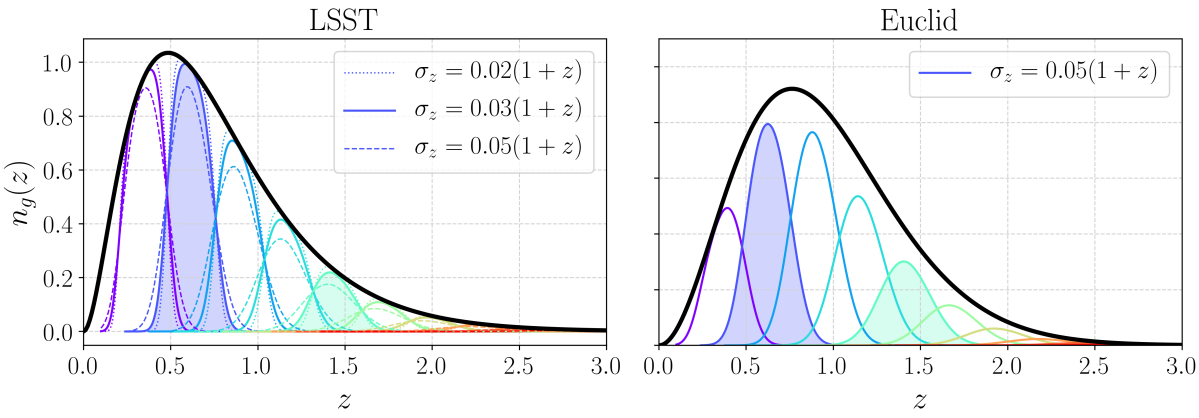


Figure 2.6: Example of redshift distributions $n_g(z)$ (*thick-black curves*), Equation (2.1), for LSST and Euclid-like surveys, and 10 photo-z bins (Equation 2.5) with different σ_0 controlling the variance of the normal distribution (Equation 2.4). Parameters utilised in the distribution can be found in Table 2.2. Each colour represents a different redshift bin i in Equation (2.5). The shaded region is an example of selecting the second redshift bin, i.e. $\phi_2(z)$, centred around $z \approx 0.6$.

expect the hydrogen to be ionised, the HI can be slightly excited by low-energy CMB photons, for example, aligning the electron and proton spins. Because this configuration is in the triplet state, there will be a natural transition to the misaligned state, resulting in the hyperfine spin-flip transition of the HI atom (for introductory discussions see, e.g., Cohen-Tannoudji et al., 2006; Griffiths, 1982). The photon's frequency emitted in this transition is probably one of the best measured values in physics: $\nu_{21} = 1,420,405,751.768 \pm 0.001$ Hz, corresponding to a photon of wavelength $\lambda_{21} \approx 21.07$ cm at the observer's rest frame (Cohen-Tannoudji et al., 2006, Chapter XII).

If the only source of the 21-cm signal is the gas contained inside galaxies that has been excited by low-energy photons, then we also expect the HI overdensity field, $\delta_{\rm HI}$, to be a biased tracer of the dark matter field, δ_m:

$$\delta_{\rm HI} \propto \delta_m. \tag{2.6}$$

While it is possible to perform galaxy redshift surveys using the 21-cm line, which detects the radio emission of individual galaxies and find their redshift by the same method of fitting the observed light frequency to the rest frame ν_{21}, i.e.

$$\nu_{\rm obs} = \frac{\nu_{21}}{1+z}, \tag{2.7}$$

we can consider an intensity mapping (IM) survey.

The IM technique[7] measures the integrated intensity of the 21-cm line of many sources,

[7]This technique can also be performed with other lines. For a review, see Kovetz et al. (2017).

over a solid angle that characterises the telescope beam. It provides us with a fast technique to survey the sky, delivering redshifts with spectroscopic precision, but at the expense of obtaining resolved galaxies, and thus having a poor angular resolution.

Because hydrogen is everywhere, one of the main (if not the paramount) challenges of 21-cm cosmology is to model the emission coming from non-cosmological sources: astrophysical emissions originated mainly from synchrotron radiation, coming from our own Galaxy, contaminates the cosmological signal in the frequency range of interest. Other foregrounds coming from outside our Galaxy and generated by different origins also contaminate the 21-cm signal. All these foregrounds usually have signals several orders of magnitude higher than the cosmological one (see Figure 3 and 7 of Alonso et al., 2014). In particular, methods to clean the foregrounds degrade the cosmological signal along the line-of-sight, corresponding to a loss of long-wavelength radial modes (Witzemann et al., 2019). Cunnington et al. (2020) showed how cleaning processes damp the HI power spectrum at large scales, biasing the measurement of primordial non-Gaussianities, which rely on such scales to be constrained. This large-scale foreground damping corresponds to a high-pass filter in Fourier space, i.e. small wavemodes (corresponding to large scales) are lost: basically, the cosmological signal is spectrally smooth at large scales, while it has a lot of structure for smaller scales (larger wavemodes); because foregrounds are also spectrally smooth at large scales, they are mixed with the cosmological signal, leading to a loss of large-scale information.

2.3.1 Cosmology with 21-cm IM

Let us consider a cloud of neutral hydrogen. If it is in thermal equilibrium, then it will radiate as a blackbody, which is characterised by the spectral brightness at a frequency ν:

$$B_\nu(\nu, T) = \frac{4\pi\hbar\nu^2}{c^2}\frac{1}{e^{2\pi\hbar\nu/k_B T} - 1}, \tag{2.8}$$

where $\hbar$ is the Planck's constant, k_B is the Boltzmann's constant, and T is the absolute temperature of the blackbody.

For the redshifts considered for IM observations, i.e. $z \lesssim 6$, the frequency range for the 21-cm line is $2\times10^8 \leq \nu_{\rm obs} \leq 1.4\times10^9$ (Hz). If we look at the exponent $x \equiv 2\pi\hbar\nu/k_B T$, showed in Figure 2.7, we can see that, for temperatures of $T \geq 2.7$ K, $x \ll 1$ is a good approximation. Expanding the exponential in Equation (2.8) around $x = 0$, we arrive at the so-called Rayleigh-Jeans approximation

$$B_\nu(\nu, T) \approx \frac{2\,k_B\,T\,\nu^2}{c^2}, \tag{2.9}$$

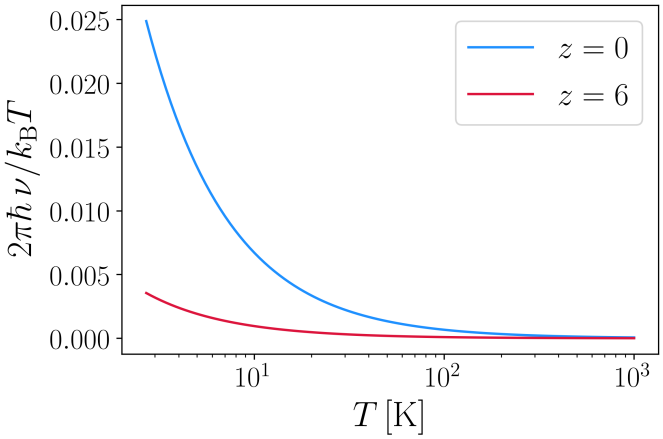


Figure 2.7: Exponent of Equation (2.8) as a function of temperature, starting at $T = 2.7$ K, and for redshifts $z = 0$ (*blue*) and $z = 6$ (*red*).

which allows us to define the *brightness temperature* $T_{\rm HI}$:

$$T_{\rm HI}(\nu) = \frac{B_\nu\, c^2}{2\, k_B\, \nu^2}. \tag{2.10}$$

Intensity mapping of the HI signal maps the fluctuations $\delta T_{\rm HI}$ of the brightness temperature $T_{\rm HI}$ around its homogeneous component $\bar{T}_{\rm HI}$,

$$\delta T_{\rm HI}(\boldsymbol{x}) \equiv \bar{T}_{\rm HI}(z)\, \delta_{\rm HI}(\boldsymbol{x}) \equiv \bar{T}_{\rm HI}(z)\, [b_{\rm HI}(z)\, \delta(\boldsymbol{x}) + \ldots]\,, \tag{2.11}$$

where the linear bias $b_{\rm HI,1}$ obtained from a halo model fit to simulations can be written as (Santos et al., 2015, 2017)

$$b_{\rm HI}(z) = \frac{b^0_{\rm HI}}{0.677105}\left(0.66655 + 0.17765\, z + 0.050223\, z^2\right), \tag{2.12}$$

with $b^0_{\rm HI} = 4.3/4.86$ (Ballardini et al., 2019).

The temperature fluctuation, Equation (2.11), is prone to different effects if compared to the galaxy signal. Firstly, the finite angular resolution of the telescope will impact its detection. The standard modelling of this beam smoothing effect in the signal is given by the inclusion of a Gaussian kernel that acts along the perpendicular direction (for example, Cunnington et al., 2020):

$$\mathcal{S}_b(k_\perp) = \exp\left[-k_\perp^2 r^2(z)\frac{\theta^2_{\rm FWHM}(z)}{16\ln 2}\right] \equiv \exp\left[-\frac{1}{2}k_\perp^2 R_b^2\right], \tag{2.13}$$

where $R_b \equiv r(z)\,\theta_{\rm FWHM}(z)/\sqrt{8\ln 2}$, $\theta_{\rm FWHM}(z) = c\,\nu^{-1}\,D^{-1}_{\rm dish}$ is the the full-width-half-maximum (FWHM) of the main angular beam (in radians) for a fixed dish diameter $D_{\rm dish}$ (in meters) at a frequency ν, associated with a certain redshift z, and $r(z)$ is the comoving distance at z. As it

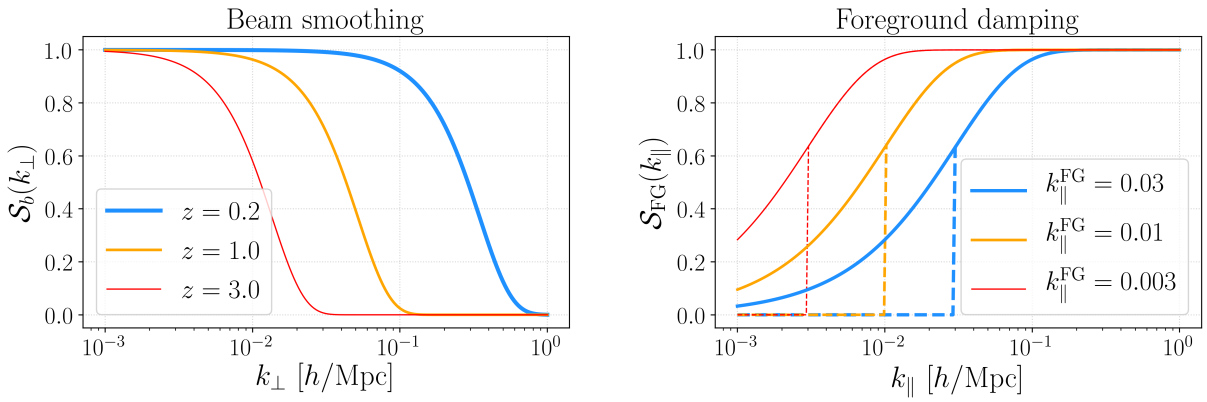


Figure 2.8: Beam smoothing (*left*) and foreground damping (*right*) functions. The beam smoothing depicted is for a fixed dish diameter $D_{\rm dish} = 15$ meters. For $\mathcal{S}_{\rm FG}$, $k_\parallel^{\rm min} = 0$ (*solid curves*), $k_\parallel^{\rm min} = k_\parallel^{\rm FG}$ (*dashed*), and $\alpha_{\rm FG} = 1$.

can be seen, the loss of information due to beam smoothing will be larger for higher redshifts and/or smaller dish diameters. The effect of beam smoothing on the perpendicular plane to the observations is shown in the left panel of Figure 2.8 for $D_{\rm dish} = 15$ meters and different redshifts.

Secondly, foregrounds originated from astrophysical emissions within and outside our Galaxy have a major impact. The cleaning of contaminated modes damps the power spectrum at very large scales, and thus its effects can be modelled generically by the function Cunnington et al. (2020, 2021):

$$\mathcal{S}_{\rm FG}(k_\parallel) = \alpha_{\rm FG}\,\Theta_{\rm FG}(k_\parallel)\left(1 - \exp\left[-\frac{k_\parallel}{k_\parallel^{\rm FG}}\right]\right), \tag{2.14}$$

where $\alpha_{\rm FG}$ describes an overall damping of the power spectrum, $k_\parallel^{\rm FG}$ is the damping scale below which wavemodes are suppressed when the foregrounds are removed (for the case of a perfect cleaning procedure, $k_\parallel^{\rm FG} = 0$), and the Heaviside step function,

$$\Theta_{\rm FG}(k_\parallel) = \begin{cases} 0, & k_\parallel < k_\parallel^{\rm min} \\ 1, & k_\parallel > k_\parallel^{\rm min} \end{cases}, \tag{2.15}$$

mimics the loss of modes unobserved by the survey, e.g. modes smaller than the minimum accessible wavemode $k_\parallel^{\rm min} = 2\pi/L_z$. The effect of foreground damping on the line-of-sight scales is shown in the right panel of Figure 2.8 in the absence of overall damping (i.e. $\alpha_{\rm FG} = 1$) and supposing $k_\parallel^{\rm min} = 0$. The presence of the Heaviside theta, i.e. of a limited number of observable scales, will cause an abrupt cut-off at the specific $k_\parallel^{\rm min}$ scale (dashed lines for the extreme case in which we lose all modes below the foreground cut-off scale, i.e. $k_{\rm min} = k_\parallel^{\rm FG}$), thus limiting the amount information that can be extracted from these surveys.

Hence, the observed temperature fluctuation will be given by

$$\delta T_{\rm HI}^{\rm obs}(\boldsymbol{k}) = \mathcal{S}_{\rm b}(k_\perp)\,\mathcal{S}_{\rm FG}(k_\parallel)\,\delta T_{\rm HI}(\boldsymbol{k}). \tag{2.16}$$

Finally, thermal fluctuations in the instrument act as source of a white noise. For an IM operating at the single-dish approach, this thermal noise, which is uncorrelated with the signal of interest, can be described by a Gaussian field with variance (Pourtsidou et al., 2017)

$$\sigma^2_{\rm pix} = T^2_{\rm sys}\,\frac{1}{\Delta\nu\, t_{\rm tot}}\,\frac{\Omega_{\rm tot}}{\Omega_{\rm pix}}\,\frac{1}{N_{\rm dishes} N_{\rm beams}}. \tag{2.17}$$

In this equation, $T_{\rm sys} = T_{\rm rx} + T_{\rm spill} + T_{\rm CMB} + T_{\rm gal}$ encapsulates the temperature contributions from the ground (spill-over $T_{\rm spill} \approx 3$ K), from the system's noise ($T_{\rm rx}$), from the CMB ($T_{\rm CMB} = 2.7$ K), and from our Galaxy (Bacon et al., 2020),

$$T_{\rm gal} = 25\,{\rm K}\left(\frac{408\,{\rm MHz}}{\nu}\right)^{2.75}. \tag{2.18}$$

The term $\Delta\nu$ is the frequency resolution, $t_{\rm tot}$ is the total time of observation, $\Omega_{\rm tot} = 4\pi f_{\rm sky}$ is the total surveyed area, $\Omega_{\rm pix}$ is the pixel solid angle, which for a Gaussian beam is given by $\Omega_{\rm pix} = 1.133\theta^2_{\rm FWHM}$ (Condon and Ransom, 2016, Section 3.3.3), and $N_{\rm dishes}$ and $N_{\rm beams}$ is the number of dishes and beams in the survey.

2.4 Multiple-tracers approach

Galaxy surveys observe different types of galaxies. As it has been discussed, the main galaxy populations utilised in cosmology are the LRGs, ELGs, and QSOs. Because they trace the dark matter field differently, it is imperative to separate this information and treat them as the different objects that they are. As we have also seen, intensity mapping provides an additional avenue to trace the dark matter field. Therefore, we are able to combine different tracers, as it has been done recently (Zhao et al., 2021), although the overlapping area containing both ELGs and LRGs in the survey was very limited. The upcoming surveys briefly outlined in Table 2.1 will have a larger area, thus a larger overlap region.

Combining different tracers constitute the so-called multiple-tracers (for short, multi-tracer or MT) technique, to circumvent the cosmic variance (CV) issue that plague the largest observable scales. As we shall reinforce throughout this thesis, a robust detection of primordial non-Gaussianities (see Chapter 4) and relativistic effects (see Chapters 5 and 6) relies on our ability to reach large scales; hence, cosmic variance will be the ultimate limitation we will face in our studies.

CV arises in all cosmological observations, affecting not only LSS studies, but CMB as well: it is related to the fact that we only observe a finite volume of the Universe and wavelengths λ of the order of the survey volume cannot be well sampled. It is also associated with the fact that, supposing we start the universe from random initial conditions, we observe only one outcome of the whole ensemble of possible realisations of those initial conditions. Hence, the statistical fluctuations of this particular realisation and the limited number of available modes at the largest scales will contribute to CV.

With the promising synergies between the main upcoming surveys, the possibility of beating CV through the MT technique[8] (Seljak, 2009; Abramo and Leonard, 2013; Abramo et al., 2015) is what makes galaxy survey cosmology an interesting and natural field to search for primordial non-Gaussianities (Yamauchi et al., 2017) and relativistic effects (Alonso and Ferreira, 2015). Also, if systematic effects between different experiments are uncorrelated, they will be naturally cancelled in cross-correlating different probes (e.g. cross-correlating $\delta_{\rm HI}$ with the galaxy field δ_g).

Table 2.3: Comparison of different surveys. High precision in redshift measurements assumes a redshift error of the order of $\sigma_z \sim O(10^{-4})$, while the poor precision refers to $\sigma_z \sim O(10^{-2})$. The filtering refers to the region in Fourier space preserved by the survey type: a photometric survey loses large radial $k_\parallel$ modes due to photometric error, corresponding to a low-pas k-filter, while intensity mapping will suffer from a loss of small $k_\parallel$-modes due to foregrounds, and of large perpendicular $k_\perp$ modes due to beam smoothing (Figure 2.8).

	Spectroscopic	**Photometric**	**Intensity mapping**
Sources	Individually resolved 3D positions	Individually resolved galaxies	Unresolved map (intensity)
Number of sources	Low	High	—
Redshift precision	High	Poor	High
Angular resolution	High	High	Poor
Filtering	—	Low-pass $k_\parallel$-filter	High-pass $k_\parallel$-filter Low-pass $k_\perp$-filter

Another possibility of employing the MT technique is the use of the cross-correlations to overcome particular barriers of a certain survey. In Table 2.3 we show how different cosmological surveys are capable of complementing each other. For example: it is a common practice to infer the redshift distribution of photometric surveys from the cross-correlation with spectroscopic samples, via a technique known as *clustering redshifts* (Newman, 2008; Ménard et al., 2013; Bates et al., 2019; Cunnington et al., 2019a,b). It employs the shared clustering signal between the two samples to recover the information of the photometric probe. Whilst photo-

[8]For a technical definition of CV, and a specialised discussion on how MT helps mitigating this issue, see Section 3.1.4.

metric surveys suffer from poor redshifts (they are often calibrated with spectroscopic samples), 21-cm surveys deliver redshifts with spectroscopic precision; on the other hand, photometry resolves individual sources, while IM provides only a continuum of the emission line coming from different directions, without resolving individual sources, suffering from poor angular resolution. Alonso et al. (2017) showed how this complementarity can improve photometric redshifts, and Alonso et al. (2021) combined the cross-correlation between CMB lensing and radio galaxy clustering to reconstruct its redshift distribution. The synergies between photometric (Section 2.2.2) and intensity mapping (Section 2.3) surveys for the clustering redshifts approach is explored in Chapter 7.

Chapter 3

Statistical Analysis of Cosmological Surveys

In this chapter we present a brief overview of the second and third-order summary statistics, power spectra and bispectra respectively, used to extract large-scale structure signals, together with the main sources of noise. We also present an overview of the common methods to estimate these statistics in simulations and real data. Details on how to estimate the multipole moments of these summary statistics and how to incorporate the survey geometry will be accounted for in Chapter 6.

3.1 The signal and the noise

The intrinsic random nature of the quantum processes that we believe gave origin to the seeds that set the beginning of structure formation in the universe demands a statistical description of the late-time matter/tracer field. If we believe that our universe can be treated by the fair sample hypothesis, meaning that if we take distant enough patches of the sky they should supposedly represent independent realisations of the same physical processes that gave rise to the observable universe, this implies that ensemble averages (i.e. averages over different realisations of the same process) can be treated as spatial averages within a certain level of confidence, since different regions correspond to different realisations under this hypothesis.

Within the inflationary picture, discussed in Chapter 4, the power spectrum of the primordial density field is given by

$$\mathcal{P}(k) \sim A_s\, k^{n_s-1}, \tag{3.1}$$

which is then evolved to late times[1]. Hence, we have predictions for the statistics of the field

[1] For more details, see the discussion around Equation (4.3).

and cosmological observations will determine parameters such as A_s and n_s, the primordial amplitude and spectral index, respectively (see Table 1.1 for the best-fitting CMB values).

There are other parameters entering our early universe physics and thus setting the initial conditions. For example, we also have the tensor-to-scalar ratio r of the primordial gravitational waves and the $f_{\rm NL}$ parameter describing departures from a perfect Gaussian field (see Chapter 4), with the latter possibly carrying a scale dependence $\sim k^{n_{\rm NL}}$ that will need an additional parameter $n_{\rm NL}$ (for example see Ballardini et al., 2019, and references therein). Finally, as we know from solving the background equations of a FLRW universe (see Appendix A), we need the parameters describing the energy content and the background evolution of the expansion. Therefore, most of the cosmological information we have comes from fitting a model to the observed data and setting constraints to these parameters.

However, cosmological observations are not different from any other sort of measurement: they are embedded in a sea of noise. In what follows, we describe what is the signal from both two and three-point statistics, and what is the primary source of noise in these measurements.

3.1.1 The signal from the second-order statistics

Second-order statistics can be described in terms of the two-point correlation function or its Fourier transform, the power spectrum. Both have advantages depending on the application, making them complementary probes of the statistical nature of the dark matter density field. On one hand, the correlation function gives the probability of finding a pair of objects separated by a certain distance $\boldsymbol{x} = \boldsymbol{x}_2 - \boldsymbol{x}_1$, with respect to a uniform field (i.e. with no cosmological structure):

$$\mathrm{d}P(\boldsymbol{x}_1, \boldsymbol{x}_2) = \bar{n}^2[1 + \xi(\boldsymbol{x})]\mathrm{d}^3x_1\mathrm{d}^3x_2, \tag{3.2}$$

where $\xi(\boldsymbol{x}) = \langle\delta(\boldsymbol{x}_1)\delta(\boldsymbol{x}_2)\rangle$ is the two-point correlation function (2PCF) of the overdensity field δ. On the other hand, the power spectrum isolates different scales, a fact that makes its covariance matrix more diagonal. The power spectrum is also extremely fast to be estimated if one uses Fast Fourier Transforms (FFTs) and is in direct contact with theory, which is easier to develop in Fourier space. However, as it shall be discussed in Chapter 6, the power spectrum is sensitive to the survey geometry, requiring a careful interpretation of what is being observed, whereas the estimated correlation function, despite being slower to compute, is a direct measurement with a cleaner interpretation.

To elucidate what the power spectrum is, we took Figure 3.1 from Guth (1997). In this picture, it is shown the density perturbations (top panel) as a function of the one-dimensional position. This signal can be decomposed into an infinite sum of sines and cosines, of different frequencies and amplitudes, as shown in the panels (a), (b) and (c). In the continuum limit, the

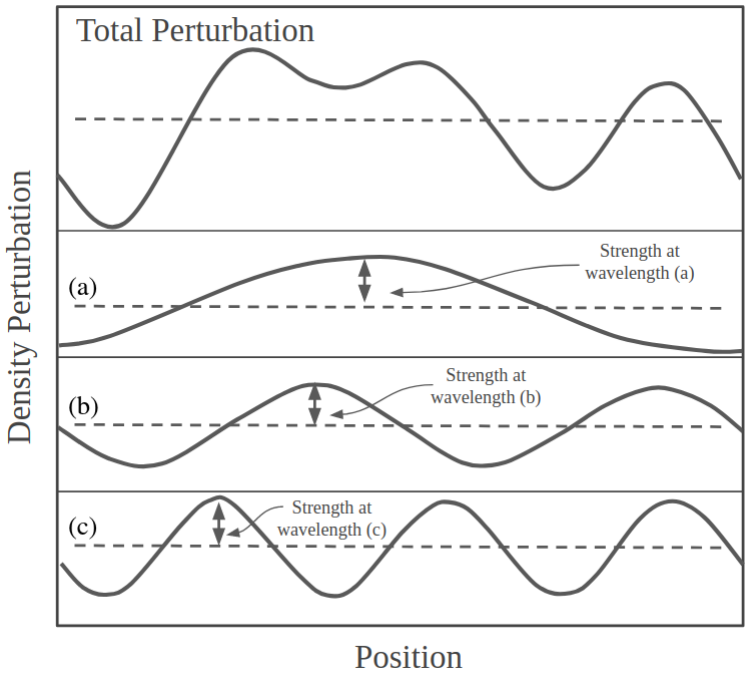


Figure 3.1: Spectral decomposition of a signal, which in this case we take as the density perturbations in the universe as a function of a one-dimensional position. The power spectrum gives the strength of the signal at each wavelength. Thus, a scale invariant power spectrum means that each wavelength contributes equally to the total perturbation. [Extracted from Guth (1997), Figure 13.2.]

sum becomes an integral and the frequency domain becomes the real line. The power spectrum, $P(k)$, represents the strength of the signal at each wavelength λ characterised by the wavenumber $k = 2\pi/\lambda$.

One can also think about the power spectrum as the correlation between two Fourier modes of the stochastic field. Then, in terms of the 2PCF, we can define the power spectrum as

$$P(\boldsymbol{k}) \equiv \int \mathrm{d}^3 x \, \mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{x}} \xi(\boldsymbol{x}), \tag{3.3}$$

or, equivalently,

$$\langle \delta(\boldsymbol{k}_1)\delta(\boldsymbol{k}_2)\rangle = (2\pi)^3 \delta^D(\boldsymbol{k}_1 + \boldsymbol{k}_2) P(\boldsymbol{k}_1), \tag{3.4}$$

where $\delta^D(\boldsymbol{k})$ is the three-dimensional Dirac delta function.

In this work we use the following Fourier conventions:

$$f(\boldsymbol{x}) = \int \frac{\mathrm{d}^d k}{(2\pi)^d} \, \mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{x}} f(\boldsymbol{k}) \text{ and } f(\boldsymbol{k}) = \int \mathrm{d}^d x \, \mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{x}} f(\boldsymbol{x}), \tag{3.5}$$

where d is the spatial dimension. Hence, the Dirac delta is written as:

$$\delta^D(\boldsymbol{k}) = \frac{1}{(2\pi)^d} \int \mathrm{d}^d \, \mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{x}}. \tag{3.6}$$

As we shall see in Chapter 4, primordial non-Gaussianities manifest themselves in the CMB and in the matter distribution of the universe. Regarding their effects on the large-scale structure (LSS), the power spectrum $P_\alpha(\boldsymbol{k}) = b_\alpha^2(z,k)\, P(z,\boldsymbol{k})$ of LSS tracers,

$$\langle \delta_\alpha(z,\boldsymbol{k})\, \delta_\alpha(z,\boldsymbol{k}')\rangle = (2\pi)^3 \delta^D(\boldsymbol{k}+\boldsymbol{k}')\, b_\alpha^2(z,\boldsymbol{k}) P(z,\boldsymbol{k}), \tag{3.7}$$

is a sensitive probe to detect the additional scale-dependence of k^{-2} on the bias $b_\alpha(\boldsymbol{k})$ (Dalal et al., 2008) of those tracers, induced by the local type (Komatsu and Spergel, 2001) of PNG. In Equation (3.7), $P_m(z,k) \equiv P(z,k) = \mathcal{M}^2(k,z) P_\Phi(k)$ is the matter power spectrum at some redshift z, and $\mathcal{M}(k,z) = 2\,k^2\, T(k)\, D(z)\, (\Omega_m\, H_0^2)^{-1}$, with $T(k)$ and $D(z)$ being, respectively, the matter transfer function and the growth factor (Sefusatti and Komatsu, 2007). Hence, non-Gaussian cosmological information can also be extracted from the power spectrum.

3.1.2 The signal from the third-order statistics

Together with the mean, the two-point correlation completely describes the statistics of Gaussian fluctuations. However, even if the universe's initial conditions came from a field whose statistics are those of a quantised harmonic oscillator, i.e. with a Gaussian spectrum, interactions with the gravity sector and the late-time structure formation generate departures from a completely Gaussian scenario. Therefore, the 2-point statistics does not give us access to the whole information content of the underlying density field. For that reason, in this thesis we will deal if the most obvious high-order statistics: the Fourier counterpart of the three-point correlation function (3PCF), namely the bispectrum.

The 3PCF measures the excess probability of finding a triplet of galaxies. It is defined as

$$\zeta(\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_3) \equiv \langle \delta(\boldsymbol{x}_1)\delta(\boldsymbol{x}_2)\delta(\boldsymbol{x}_3)\rangle. \tag{3.8}$$

In complete analogy with the power spectrum, we can define the bispectrum as

$$B(\boldsymbol{k}_1,\boldsymbol{k}_2,\boldsymbol{k}_3) = \int \mathrm{d}^3x_1\, \mathrm{d}^3x_2\, \mathrm{d}^3x_3\; \mathrm{e}^{-i\boldsymbol{k}_1\cdot\boldsymbol{x}_1}\, \mathrm{e}^{-i\boldsymbol{k}_2\cdot\boldsymbol{x}_2}\, \mathrm{e}^{-i\boldsymbol{k}_3\cdot\boldsymbol{x}_3}\, \zeta(\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_3), \tag{3.9}$$

or, equivalently, as

$$\langle \delta(\boldsymbol{k}_1)\, \delta(\boldsymbol{k}_2)\, \delta(\boldsymbol{k}_3)\rangle = (2\pi)^3 \delta^D(\boldsymbol{k}_1+\boldsymbol{k}_2+\boldsymbol{k}_3)\, B(\boldsymbol{k}_1,\boldsymbol{k}_2,\boldsymbol{k}_3). \tag{3.10}$$

In the same way that the power spectrum is a correlation between two Fourier modes k, the bispectrum correlates the triplet $\Delta = \{\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3\}$. Each possible triangular configuration formed by the three wavemodes can be parametrised by different amplitude parameters. Thus, just as the power spectrum gives the amplitude of the different two-point correlation between k_1 and k_2, the bispectrum gives us the amplitude of the different triangle contributions to the complete signal. The shapes commonly expected from inflationary models are depicted in Figure 3.2.

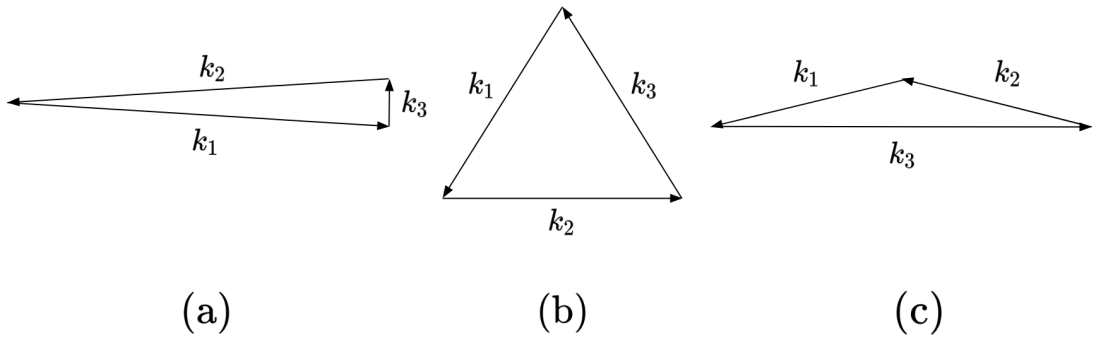


Figure 3.2: (a) Squeezed configuration ($k_3 \ll k_1 \sim -k_2$); (b) Equilateral configuration ($k_1 \simeq k_2 \simeq k_3$) and (c) Flattened configuration ($k_3 \simeq k_1 + k_2$).

Attempts to detected PNG from the CMB bispectrum have been done by the WMAP satellite (Senatore et al., 2010), with the latest constraints, coming from the Planck collaboration (Akrami et al., 2020), indicating Gaussianity of the primordial perturbations: $f_{\rm NL}^{\rm loc} = -0.9 \pm 5.1$, $f_{\rm NL}^{\rm eq} = -26 \pm 47$ and $f_{\rm NL}^{\rm ort} = -38 \pm 24$, representing the amplitude of the squeezed, equilateral and orthogonal (a combination of equilateral and squeezed shapes) configurations, respectively. Yet, such limits are not satisfactory. For the case of local-type PNG, Castorina et al. (2019) gave the latest constraints on these type of PNG with the DR14 of eBOSS quasars: $-51 < f_{\rm NL}^{\rm loc} < 21$. They also made a forecast showing that the full eBOSS dataset is capable of reaching similar constraints: $\sigma(f_{\rm NL}^{\rm loc}) \simeq 5 - 8$. This means that future surveys can help us push the limits on PNG constraints.

Neglecting shot-noise, which will be discussed in what follows, the (theoretical) galaxy bispectrum $B_{\alpha\beta\gamma}(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3)$ is defined as:

$$\langle \delta_\alpha(\boldsymbol{k}_1)\, \delta_\beta(\boldsymbol{k}_2)\, \delta_\gamma(\boldsymbol{k}_3) \rangle = (2\pi)^3 \delta^D(\boldsymbol{k}_1 + \boldsymbol{k}_2 + \boldsymbol{k}_3) B_{\alpha\beta\gamma}(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3), \tag{3.11}$$

where Greek letters refer to different tracers (e.g. blue galaxies, red galaxies and quasars).

At large scales, we can consider the bias expansion (Fry and Gaztañaga, 1993)

$$\delta_\alpha(\boldsymbol{x}) = \alpha_1 \delta(\boldsymbol{x}) + \frac{\alpha_2}{2} \delta^2(\boldsymbol{x}) + \ldots, \tag{3.12}$$

where we call $\alpha_1 \equiv b_\alpha^{(1)}$ and $\alpha_2 \equiv b_\alpha^{(2)}$ the linear and second order local-in-matter density bias, respectively. Considering also the the evolved matter density contrast in perturbation theory up to second order, $\delta(\boldsymbol{k}) = \delta^{(1)}(\boldsymbol{k}) + \delta^{(2)}(\boldsymbol{k}) + \ldots$, the cross bispectrum of galaxies reads

$$B_{\alpha\beta\gamma}(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3) = \alpha_1\beta_1\gamma_1 B_m^{\text{tree}} + \alpha_1\beta_1\gamma_2 P(k_1)P(k_2) + \alpha_1\beta_2\gamma_1 P(k_1)P(k_3) + \alpha_2\beta_1\gamma_1 P(k_2)P(k_3), \tag{3.13}$$

with $\alpha_1 \equiv b_\alpha^{(1)}, \alpha_2 \equiv b_\alpha^{(2)}$, and the same holding for β and γ, $B_m^{\text{tree}} = B_0 + B_{\text{grav}}$ is the tree-level matter bispectrum with a possible primordial contribution, $B_0 = \mathcal{M}(k_1, z)\mathcal{M}(k_2, z)\mathcal{M}(k_3, z)\, B_\Phi$, evolved up to redshift z, and B_{grav} is the tree-level bispectrum induced by gravity,

$$B_{\text{grav}} = 2\left[F_2(\boldsymbol{k}_1, \boldsymbol{k}_2)P_\Phi(k_1)P_\Phi(k_2) + 2 \text{ perms.}\right], \tag{3.14}$$

where F_2 is the second-order kernel obtained from perturbation theory, in an Einstein-de Sitter (EdS) universe (Jain and Bertschinger, 1994; Matarrese et al., 1997; Bernardeau et al., 2002):

$$F_2(\boldsymbol{k}_1, \boldsymbol{k}_2) = \frac{5}{7} + \frac{1}{2}\frac{\boldsymbol{k}_1 \cdot \boldsymbol{k}_2}{k_1 k_2}\left(\frac{k_1}{k_2} + \frac{k_2}{k_1}\right) + \frac{2}{7}\frac{(\boldsymbol{k}_1 \cdot \boldsymbol{k}_2)^2}{k_1^2 k_2^2}. \tag{3.15}$$

More generally, this kernel can be written as (Catelan et al., 1995):

$$J(\boldsymbol{k}_1, \boldsymbol{k}_2, \Omega_{m,0}) = 1 - B(\Omega_{m,0}) + \frac{1}{2}\frac{\boldsymbol{k}_1 \cdot \boldsymbol{k}_2}{k_1 k_2}\left(\frac{k_1}{k_2} + \frac{k_2}{k_1}\right) + B(\Omega_{m,0})\frac{(\boldsymbol{k}_1 \cdot \boldsymbol{k}_2)^2}{k_1^2 k_2^2}, \tag{3.16}$$

where $\Omega_{m,0}$ is the matter density today, and $B(\Omega_{m,0}) = 2/7$ for the Einstein-de Sitter case. As pointed out in Matarrese et al. (1997), Equation (3.16) is weakly dependent on cosmology and hence it is the common practice to work with the EdS universe assumption. The EdS approximation is also employed in the time evolution of some quantities, such as the bias expansion for the redshift-space power spectrum, and it is known to work well within the precision of current and next generation surveys, yielding reasonably accurate results for a vast range of redshifts (Donath and Senatore, 2020).

3.1.3 Shot noise

The shot noise is not just a fluctuation around the expected value because galaxies are discrete objects we choose to model with a Poisson statistics; it has a physical origin. In this section we give a brief physical explanation for the shot noise, based on discussions with Marko Simonović, which we deeply thank for bringing this subject up.

The shot noise emerges from the intrinsic stochasticity of the discrete samples of objects, in the sense that, for example, galaxies and dark matter halos depend upon the merging of

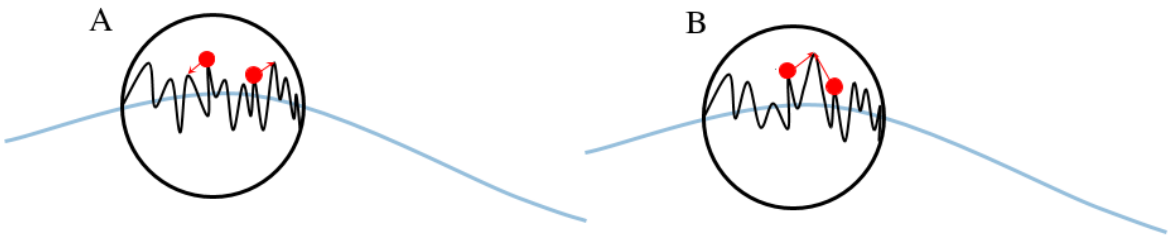


Figure 3.3: Small-scale fluctuations (*black curves*) that we average over a region of radius R (*black circle*) sitting on top of a large-scale fluctuation (*blue curves*). (*Left*) Case A: two objects (*red dots*) sitting on top of overdensities do not merge, because they are attracted to opposite regions due to small-scale fluctuations. (*Right*) Case B: the two objects merge as the small scale fluctuations end up favouring a common overdensity attracting both objects. In this case, the total number of objects is 1 unit less than in case A.

structures to exist. As already mentioned, the halo/galaxy overdensity can be written as

$$\delta_\alpha(\boldsymbol{x}) = b_1\delta(\boldsymbol{x}) + b_2\delta^2(\boldsymbol{x}) + \ldots + \epsilon(\boldsymbol{x}), \tag{3.17}$$

where we included a stochastic $\epsilon(\boldsymbol{x})$ term that comes from smoothing the density field within small scales.

There are two possible cases: (A) two objects merge or (B) they do not merge. This is completely random, with the possibility of A or B happening depending on small scale modes which we average over in the absence of a better understanding. In the first case (A), the number of these objects is larger (by one unit in this idealised situation) than in the second case (B), as shown in Figure 3.3 . If we compute the Fourier transform of the difference between the overdensity in scenario A, $\delta^A_{\alpha,s}(\boldsymbol{x})$, and scenario B, $\delta^B_{\alpha,s}(\boldsymbol{x})$, of this tracer, where s refers to the small scale modes, then we obtain

$$\Delta\delta_\alpha(k) = \int_R \mathrm{d}^3x\, \mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{x}}\,(\delta^A_{\alpha,s} - \delta^B_{\alpha,s}) = \text{shot noise}, \tag{3.18}$$

as they do not cancel since the number density is not conserved. For the case of dark matter (CDM) particles, the overdensity is the same, in both scenarios, as matter *is* conserved. Hence, there is no shot noise for CDM.

From that, it is also clear that difference tracers may carry shot noise, as opposed to the standard treatment where they are assumed to be uncorrelated. This has been explored in Hamaus et al. (2010) and the cross-correlation between different tracers was found to reduce the shot noise in the power-spectrum estimates.

The shot noise in Equation (3.18), which represents the stochasticity of LSS tracers relative to the dark matter field (Hamaus et al., 2010), can be modelled when one thinks of it as

the consequence of a discrete set (galaxies and halos) sampling a continuous field. Therefore, there are two stochastic processes determining how galaxies will be distributed in the sky: the continuous random process characterising the mass distribution $\rho_m(\boldsymbol{x})$, coming from theories such as inflation, and the random point process generating the discrete distribution over the continuous random field $\rho_g(\boldsymbol{x}) = b_1\,\rho_m(\boldsymbol{x})$, as depicted in Figure 3.3. Here, the proportionality constant b_1 is the linear bias connecting the dark matter and tracer fields, which represents an enhancement in the probability of forming tracers due to the large-scale modulation (light-blue curves in Figure 3.3; see Appendix C.1.3 for discussion). The distribution of the latter is usually taken to be Poissonian $P(N|\mu)$ with mean μ.

To illustrate why the Poisson distribution is a reasonable assumption, we show in Figure 3.4 the distribution of all satellite galaxies that populate 350 halos, obtained from the Illustris TNG300 simulation[2] (Nelson et al., 2021), contained inside a mass bin $6.31 \times 10^{13} \leq M_h < 1.58 \times 10^{14}$ $[M_\odot/h]$ (blue histogram). The mean mass of all 350 halos is $\log_{10} \bar{M}_h = 13.98$. At first sight, the Poisson distribution seems unsuitable to describe the occurrence of N galaxies for a given mean μ. However, if we refine the halo mass bin by splitting it, for example, into 10 sub-bins $M_h^{(i)}$ containing N_i halos, one can obtain the joint (dashed) curve by stacking 10 distinct Poisson samples $P(N|\mu_i)$, where μ_i is the mean number of satellites inside each one of the sub-bins (this process is illustrated in Figure 3.5). As one can see, the observed distribution of the broader mass bin is well described by the stacking of Poisson samples.

The discrete count of tracers at a position $\boldsymbol{x}_i$ can then be written as

$$n_\alpha(\boldsymbol{x}_i) = \bar{n}_\alpha + \delta n_\alpha^{(G)}(\boldsymbol{x}_i) + \delta n_\alpha^{(P)}(\boldsymbol{x}_i), \tag{3.19}$$

where $\bar{n}_\alpha$ is the mean number density inside a volume element, $\delta n_\alpha^{(G)}(\boldsymbol{x}_i)$ is the continuous Gaussian contribution to the number density of tracers, and $\delta n_\alpha^{(P)}(\boldsymbol{x}_i)$ is the Poissonian fluctuation that emerges from the stochastic process discussed above.

By computing the two-point correlation function of the overdensity field

$$\delta_\alpha(\boldsymbol{x}_i) \equiv \frac{n_\alpha(\boldsymbol{x}_i)}{\bar{n}_\alpha} - 1 = \frac{\delta n_\alpha^{(G)}(\boldsymbol{x}_i)}{\bar{n}_\alpha} + \frac{\delta n_\alpha^{(P)}(\boldsymbol{x}_i)}{\bar{n}_\alpha}, \tag{3.20}$$

we obtain

$$\langle \delta_\alpha(\boldsymbol{x}_i)\,\delta_\beta(\boldsymbol{x}_j)\rangle = \frac{1}{\bar{n}_\alpha\,\bar{n}_\beta}\langle \delta n_\alpha^{(G)}(\boldsymbol{x}_i)\,\delta n_\beta^{(G)}(\boldsymbol{x}_j)\rangle + \frac{1}{\bar{n}_\alpha\,\bar{n}_\beta}\langle \delta n_\alpha^{(P)}(\boldsymbol{x}_i)\,\delta n_\beta^{(P)}(\boldsymbol{x}_j)\rangle \tag{3.21}$$

after assuming the Gaussian and Poissonian processes to be uncorrelated. The first correlator is

[2] https://www.tng-project.org/data/

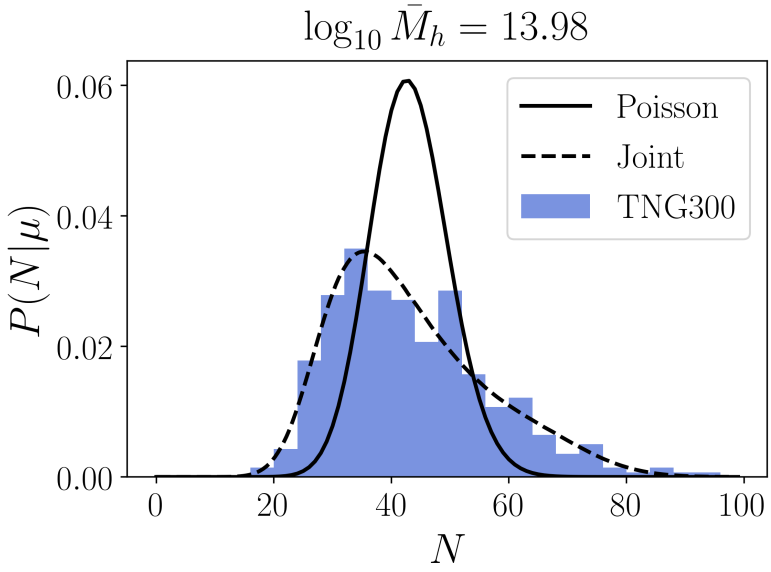


Figure 3.4: Distribution $P(N|\mu)$ of obtaining N satellite galaxies for 350 Illustris TNG300 dark matter halos (Nelson et al., 2021). The blue region shows the distribution of N occurrences for all the satellites inside the mass bin $M_h \in [6.31 \times 10^{13}, 1.58 \times 10^{14})$ $[M_\odot/h]$. By breaking M_h into 10 sub-bins, $M_h^{(i)}$, containing N_i halos, each one will have a corresponding mean number of satellites μ_i. From that, it is possible to drawn a Poisson sample $P(N|\mu_i)$ for each sub-bin. The stacking of these 10 Poissonian samples yields the joint curve (*dashed*). The solid curve represents the Poisson distribution $P(N|\mu)$, where μ is the mean inside the broader mass bin M_h.

related to the dark matter field via some biasing scheme,

$$\frac{1}{\bar{n}_\alpha \bar{n}_\beta} \langle \delta n_\alpha^{(G)}(\boldsymbol{x}_i)\, \delta n_\beta^{(G)}(\boldsymbol{x}_j) \rangle = b_\alpha b_\beta \langle \delta(\boldsymbol{x}_i)\, \delta(\boldsymbol{x}_j) \rangle \equiv b_\alpha b_\beta\, \xi(\boldsymbol{x}_{ij}), \tag{3.22}$$

where $\boldsymbol{x}_{ij} \equiv \boldsymbol{x}_i - \boldsymbol{x}_j$. For a Poisson distribution,

$$\langle \delta n_\alpha^{(P)}(\boldsymbol{x}_i)\, \delta n_\beta^{(P)}(\boldsymbol{x}_j) \rangle = \bar{n}_\alpha\, \delta^D(\boldsymbol{x}_{ij})\, \delta^{\mathrm{K}}_{\alpha\beta}, \tag{3.23}$$

where $\delta^{\mathrm{K}}_{\alpha\beta}$ is the Kronecker delta. Hence,

$$\langle \delta_\alpha(\boldsymbol{x}_i)\, \delta_\beta(\boldsymbol{x}_j) \rangle \equiv \xi_{\alpha\beta}(\boldsymbol{x}_{ij}) = b_\alpha\, b_\beta\, \xi(\boldsymbol{x}_{ij}) + \frac{1}{\bar{n}_\alpha} \delta^D(\boldsymbol{x}_j - \boldsymbol{x}_i)\, \delta^K_{\alpha\beta}. \tag{3.24}$$

Equation (3.24) above has a correction of $\bar{n}^{-1}$ to the continuous density field $\xi_{\alpha\beta}(\boldsymbol{x}_{12})$ coming from the discrete nature of the tracers. This correction is known as the Poisson shot noise.

Taking the Fourier transform of Equation (3.24), we obtain the galaxy power spectrum:

$$P_{\alpha\beta}(\boldsymbol{k}) = b_\alpha b_\beta P(\boldsymbol{k}) + \frac{1}{\bar{n}_\alpha} \delta^K_{\alpha\beta}. \tag{3.25}$$

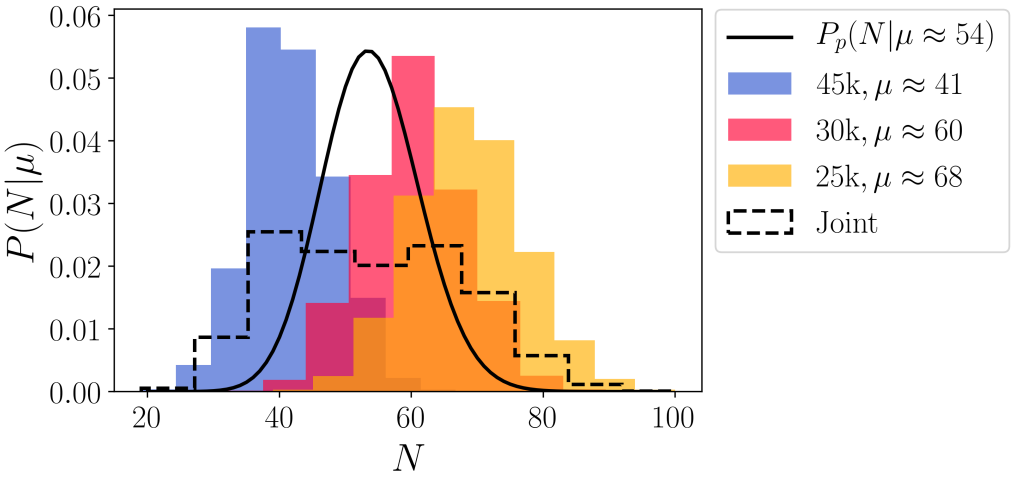


Figure 3.5: For illustrative purposes, we draw three Poisson samples from $N_1 = 45000$ (*blue*), $N_2 = 30000$ (*magenta*) and $N_3 = 25000$ (*orange*) realisations with the expected number of events given by $\mu_1 \approx 41$, $\mu_2 \approx 60$ and $\mu_1 \approx 68$, respectively. Each of these samples showcase 3 sub-bins of a broader bin (in the case of Figure 3.4, corresponding to halo masses). The dashed histogram shows the stacking of these three Poisson samples into a single one. As we can see, the combination of Poisson distributions is not Poissonian. The solid curve corresponds to a Poisson distribution with mean $\mu = \sum_{i=1}^{3} N_i \mu_i / \sum_{i=1}^{3} N_i \approx 54$, which clearly does not represent the joint distribution.

3.1.4 Cosmic variance

The second source of noise around the cosmological signal described in Sections 3.1.1 and 3.1.2 is the well known cosmic variance, affecting the largest scales available. It is related to the fact that we only observe a finite volume of the universe and wavelengths λ of the order of the survey volume cannot be well sampled. As already discussed, and as shall be reinforced throughout this text, a robust PNG and RE detection relies on our ability to reach large scales; hence, cosmic variance will be the ultimate limitation we will face in our studies.

As derived in Section D.2, the Fisher matrix for a survey containing multiple types of LSS tracers is given by Equation (D.37), which we write below for the sake of discussion:

$$\mathcal{F}_{\mu\nu} = \frac{1}{4} \frac{\delta_{\mu\nu} \mathcal{P}_\mu \mathcal{P} (1 + \mathcal{P}) + \mathcal{P}_\mu \mathcal{P}_\nu (1 - \mathcal{P})}{(1 + \mathcal{P})^2}. \tag{3.26}$$

This equation reduces to the so-called FKP Fisher matrix in the simplified case where there is only one tracer (e.g. only one type of galaxy). To elucidate this expression, we highlight the following points:

- The Fisher matrix is the ideal case set by the Cramér-Rao lower bound (CRLB), which states that the variance of some estimator $\hat{\theta}$ for a parameter θ, whose true value is $\bar{\theta}$, is

$$\mathrm{var}[\hat{\theta}] \geq \left\langle \frac{\partial^2 \ln p(\boldsymbol{d}|\theta)}{\partial \theta^2} \right\rangle^{-1} \equiv \mathcal{F}. \tag{3.27}$$

In this expression, the brackets denote an ensemble average over the parameter in the frequentist sense, $p(\boldsymbol{d}|\theta)$ is the probability of the data given a certain value of θ, and $\mathcal{F}$ is the Fisher matrix. When the CRLB is saturated, then the variance is the same as the Fisher matrix, that is, $\mathcal{F}$ imposes the minimum error one can attain in the estimation of θ via $\hat{\theta}$. When the estimator saturates the CRLB, then it is said to be optimal if its expected value $\langle\hat{\theta}\rangle$ corresponds to the true value $\bar{\theta}$ (see, e.g., Chapter 3 of Kay, 1993);

- $\mathcal{P}_\mu = \bar{n}_\mu\, b_\mu^2\, P$ is the clustering strength of each tracer μ in units of shot noise $1/\bar{n}$, b_μ is the tracer's linear bias, P is the matter power spectrum, and $\mathcal{P} = \sum_\mu \mathcal{P}_\mu$ is the total clustering strength of all tracers combined. Notice that the higher its value, the larger the power spectrum signal;

- For a single tracer, $\mathcal{P}_\mu = \mathcal{P}$, and Equation (3.26) becomes

$$\mathcal{F}_\mu = \frac{1}{2}\left(\frac{\mathcal{P}_\mu}{1+\mathcal{P}_\mu}\right)^2, \tag{3.28}$$

which, in the case of infinite signal $\mathcal{P}_\mu \to \infty$, yields

$$\mathcal{F}_\mu \to \frac{1}{2}. \tag{3.29}$$

This is the formal way to comprehend cosmic variance: the Fisher matrix is bounded, implying that there will always be a limited amount of information, translated by the ever-present error in our measurements, no matter how large the signal is.

As first shown in Seljak (2009), a way to overcome this limitation is by employing the combined analysis of multiple species of LSS tracers and, as shown in Abramo et al. (2015), who first derived Equation (3.26), the combination of tracers allows us to reach, in theory, an infinite amount of information: if $\mathcal{P}_\mu \to \infty$, then $\mathcal{F}_{\mu\nu} \to \infty$ and, therefore, the power-spectrum variance vanishes.

3.2 An overview of estimators

As discussed above, the power spectrum and bispectrum allows an efficient way to compress data from galaxy surveys in order to extract cosmological information through constraints on model parameters. There is a multitude of ways to estimate these summary statistics from galaxy surveys, and here we make an overview of such methods.

3.2.1 Direct methods

Direct methods of estimating the power spectrum and bispectrum are suitable for simulations defined in a cubic box with constant mean number density of objects. However, if we are dealing with real data, where the mean number density of objects is not a constant, but rather a function of position in real space, i.e. $\bar{n} = \bar{n}(\boldsymbol{x})$, than one can use the traditional approach discussed in Section 3.2.2 to incorporate the selection function information in the density field, allowing for a better accuracy performance.

In this section we review the works of Jeong (2010) and Watkinson et al. (2017), where they employed a guide to estimate polyspectra by means of the Fast Fourier Transform (FFT). Throughout this thesis, we call them FFT polyspectra estimators.

3.2.1.1 The FFT Polyspectra Estimator

Consider the discrete Fourier transform

$$\delta(\boldsymbol{x}) = \frac{1}{V}\sum \delta(\boldsymbol{k})\, \mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{x}} \quad \text{and} \quad \delta(\boldsymbol{k}) = \frac{V}{N_{\rm pix}}\sum \delta(\boldsymbol{x})\, \mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{x}}, \tag{3.30}$$

where $\delta(\boldsymbol{x})$ is the density field one wishes to analyse, $\delta(\boldsymbol{k})$ is its discrete Fourier transform and $N_{\rm pix} = N^3$ is the total number of pixels (cells) within the volume V. The polyspectra $\mathscr{P}$, defined as

$$(2\pi)^3 \mathscr{P}(\boldsymbol{k}_1, \ldots, \boldsymbol{k}_p)\, \delta^D(\boldsymbol{k}_1 + \ldots + \boldsymbol{k}_p) = \Big\langle \delta(\boldsymbol{k}_1) \ldots \delta(\boldsymbol{k}_p) \Big\rangle, \tag{3.31}$$

must be translated to its discrete version since the grids of simulations contain pixelized data. The connection between the Kronecker (discrete) δ^K and the Dirac (continuum) δ^D is given by:

$$\delta^D(\boldsymbol{k}) \to \frac{1}{k_{\rm F}^3}\delta^K(\boldsymbol{m}) = \frac{V}{(2\pi)^3}\delta^K(\boldsymbol{m}), \tag{3.32}$$

where $\boldsymbol{m}$ is the dimensionless pixel coordinate (m_x, m_y, m_y), i.e. maps the grid points, such that $\boldsymbol{k} = \frac{2\pi}{L}\boldsymbol{m} \equiv k_{\rm F}\boldsymbol{m}$, and $\boldsymbol{L} = (L_x, L_y, L_z)$ is the length of the simulated box sides. For the case of an unnormalised FFT routine, the FFT of the input density contrast, $\Delta(k_{\rm F}\boldsymbol{m})$, is related to the Fourier density contrast as

$$\Delta(k_{\rm F}\boldsymbol{m}) = \frac{N_{\rm pix}}{V}\langle\delta(\boldsymbol{k})\rangle_{\boldsymbol{m}} \tag{3.33}$$

where the $\Delta(k_{\rm F}\boldsymbol{m})$ is the Fast Fourier Transform (FFT) of the input density contrast, which can be easily obtained through numerical libraries such as the `numpy.fft`, in Python, or `FFTW`, in C, and $\langle\delta(\boldsymbol{k})\rangle_{\boldsymbol{m}}$ is the theoretical FFT, Equation (3.30), averaged over a discrete bin $\boldsymbol{m}$. Therefore,

the discrete version of Equation (3.31) is:

$$\mathscr{P}(\boldsymbol{k}_1,\ldots,\boldsymbol{k}_p)\delta^K(\boldsymbol{m}_1+\ldots+\boldsymbol{m}_p) \approx \frac{V^{p-1}}{N^p_{\rm pix}}\Big\langle \Delta(k_{\rm F}\boldsymbol{m}_1)\ldots\Delta(k_{\rm F}\boldsymbol{m}_p)\Big\rangle. \tag{3.34}$$

The equality of Equation (3.31) becomes an approximation in Equation (3.34) since the average is performed within a bin-width of, at least, $k_{\rm F}$ (Watkinson et al., 2017). By enforcing the Kronecker delta on the right-hand side of Equation (3.34),

$$\begin{aligned}\mathscr{P}(\boldsymbol{k}_1,\ldots,\boldsymbol{k}_p) &\approx \frac{V^{p-1}}{N^p_{\rm pix}}\left\langle \prod_{i=1}^{p}\delta^K(\boldsymbol{m}_1+\ldots+\boldsymbol{m}_p)\Delta(k_{\rm F}\boldsymbol{m}_i)\right\rangle; \\ &\approx \frac{V^{p-1}}{N^p_{\rm pix}}\frac{\sum_{\boldsymbol{m}_1\simeq\boldsymbol{n}_1}\cdots\sum_{\boldsymbol{m}_p\simeq\boldsymbol{n}_p}\prod_{i=1}^{p}\delta^K(\boldsymbol{m}_1+\ldots+\boldsymbol{m}_p)\Delta(\boldsymbol{m}_i)}{\sum_{\boldsymbol{m}_1\simeq\boldsymbol{n}_1}\cdots\sum_{\boldsymbol{m}_p\simeq\boldsymbol{n}_p}\prod_{i=1}^{p}\delta^K(\boldsymbol{m}_1+\ldots+\boldsymbol{m}_p)},\end{aligned} \tag{3.35}$$

where $\sum_{\boldsymbol{m}_i\simeq\boldsymbol{n}_i}$ formally means the sum over all $\boldsymbol{m}_i$ vectors that fall within some arbitrary bin of width $\boldsymbol{s}$ and $\boldsymbol{n}=\left(\frac{N_x x}{L_x},\frac{N_y y}{L_y},\frac{N_z z}{L_z}\right)$, with N_x, N_y, N_z the number of pixels on each side of the box. By noticing that

$$\delta^K(\boldsymbol{m}_1+\ldots+\boldsymbol{m}_p)=\frac{1}{N_{\rm pix}}\sum_{\boldsymbol{n}}^{N_{\rm pix}}\prod_{i=1}^{p}{\rm e}^{2\pi\, i\,\boldsymbol{n}\cdot\boldsymbol{m}_i/N_{\rm side}}, \tag{3.36}$$

with $N_{\rm side}$ being the number of pixels on each side of the cube, the polyspectrum estimator is written as

$$\mathscr{P}(\boldsymbol{k}_1,\ldots,\boldsymbol{k}_p) \approx \frac{V^{p-1}}{N^p_{\rm pix}}\frac{\sum_{\boldsymbol{n}}^{N_{\rm pix}}\prod_{i=1}^{p}\delta(\boldsymbol{n},\boldsymbol{k}_i)}{\sum_{\boldsymbol{n}}^{N_{\rm pix}}\prod_{i=1}^{p}I(\boldsymbol{n},\boldsymbol{k}_i)}, \tag{3.37}$$

where

$$\delta(\boldsymbol{n},\boldsymbol{k}_i)=\sum_{\boldsymbol{m}_i\in\boldsymbol{k}_i\pm s/2}\Delta(\boldsymbol{m}_i){\rm e}^{2\pi\, i\,\boldsymbol{n}\cdot\boldsymbol{m}_i/N_{\rm side}}, \tag{3.38}$$

and

$$I(\boldsymbol{n},\boldsymbol{k}_i)=\sum_{\boldsymbol{m}_i\in\boldsymbol{k}_i\pm s/2}{\rm e}^{2\pi\, i\,\boldsymbol{n}\cdot\boldsymbol{m}_i/N_{\rm side}}. \tag{3.39}$$

Notice that, for the case of $p=3$, $I(\boldsymbol{n},\boldsymbol{k}_i)$ is the total number of independent triangles with sides (k_1,k_2,k_3) (Jeong, 2010). The procedure to obtain $\mathscr{P}$ is illustrated in Figure 3.6 for the case of the bispectrum, i.e. $p=3$.

Compared to the direct sampling method (see Jeong, 2010, p. 268), which computes the bispectrum as

$$B(\boldsymbol{k}_1,\boldsymbol{k}_2,\boldsymbol{k}_3)=\frac{1}{(2\pi)^3}\frac{1}{N_{\rm tri}}\sum_{m\in{\rm Tri}_{123}}\Delta(k_{\rm F}\boldsymbol{m}_1)\Delta(k_{\rm F}\boldsymbol{m}_2)\Delta(k_{\rm F}\boldsymbol{m}_3), \tag{3.40}$$

where $N_{\rm tri}$ is the total number of triangles and ${\rm Tri}_{123}$ is the set of vectors $\{\boldsymbol{m}_1,\boldsymbol{m}_2,\boldsymbol{m}_3\}$ that form the triangles with $\boldsymbol{k}_1+\boldsymbol{k}_2+\boldsymbol{k}_3=0$, the main disadvantage of the direct measurement,

Equation (3.40), is that one must account for all possible triangles available. This process poses a computational challenge to estimate the bispectrum, which is of order $N^3_{\rm pix}$ (all possible combinations of pixels forming the triangles of interest).

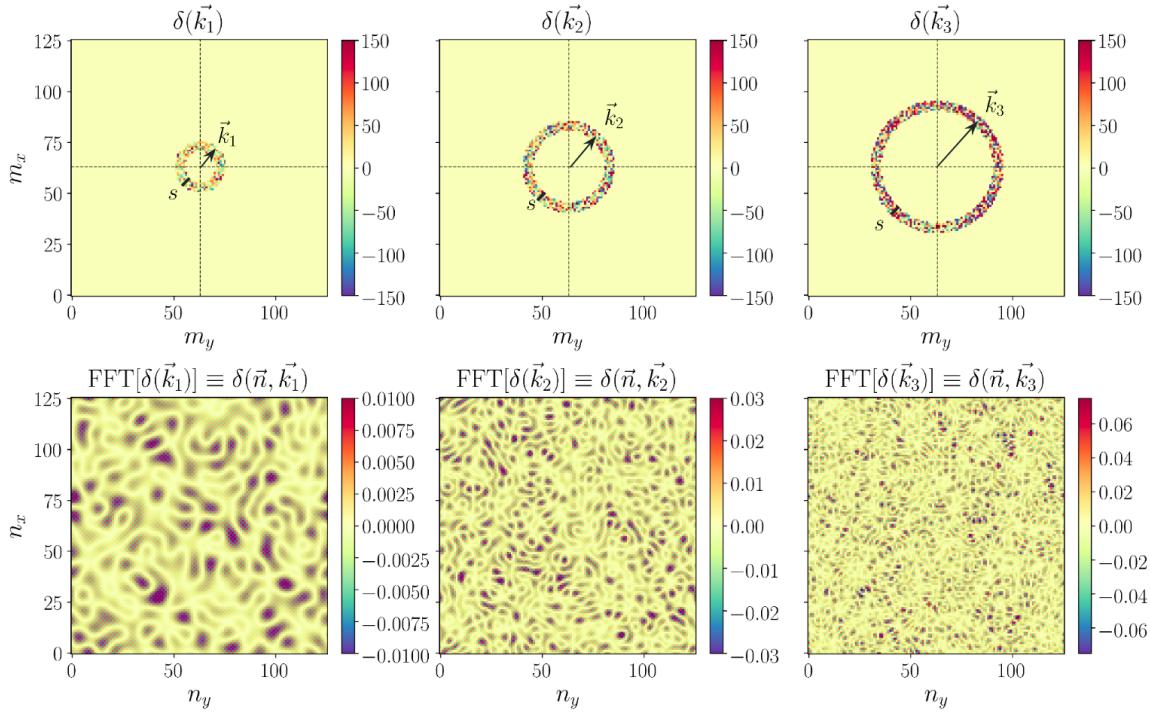


Figure 3.6: The process of obtaining $\delta(\boldsymbol{n}, \boldsymbol{k}_i)$. We begin by getting the normalisation factors $I(\boldsymbol{n}, \boldsymbol{k}_i)$ which are one whenever a pixel matches the condition $\boldsymbol{m}_i \in \boldsymbol{k}_i \pm s/2$. Then, $\delta(\boldsymbol{n}, k_i) = I(\boldsymbol{n}, k_i)\,\delta_{\rm g}$. (*Top*) Point selection for which the FFT of the galaxies density field $\delta_{\rm g}$ falls into a bin of width s, separated from the origin by $|\boldsymbol{k}_i|$. (*Bottom*) Inverse transform of the objects shown in the top figure. This process introduces the $\boldsymbol{n}$ in $\delta(\boldsymbol{k}_i)$.

In the case of the FFT estimator, there is no need to select the possible configurations; instead, the spherical shells containing $\delta(k_i)$ (top panels of Figure 3.6), when Fourier transformed to real space, take care of all the possible triangles defined by $\boldsymbol{k}_1, \boldsymbol{k}_2$ and $\boldsymbol{k}_3$; that is, in the bottom panels showed in Figure 3.6, only the points of the density contrast field forming the triangles that satisfy $\boldsymbol{k}_1 + \boldsymbol{k}_2 + \boldsymbol{k}_3 = 0$ are selected. This reduces the number of operations to a fairly small number of operations: since the grid and the thickness s of the spherical shells are finite, the number of shells is limited.

It is important to stress that, in order to apply such estimator, we work within a cosmological box, which is obtained by transforming the truly observed galaxy position (z, θ, φ), where z is its measured redshift, to a set of Cartesian coordinates (x, y, z) that can only be obtained once we go from redshift to comoving coordinates (as in Equation A.45).

3.2.2 Traditional methods

As defined in Tegmark et al. (1998), traditional methods obtain the power spectrum by means of band-powers estimates that are computed "by multiplying the observed density field by some weight function, Fourier transforming it, taking the squared modulus of the result and averaging over shells in k-space". Perhaps the most common method to estimate the power spectrum is the one developed by (Feldman et al., 1994), also known as the FPK method. It has the advantage of also working for the bispectrum (Scoccimarro, 2000; Martínez and Saar, 2001; Gil-Marín et al., 2015; Gil-Marín et al., 2015), and in this case we will call it the Generalised FKP (GFKP) method. It was also generalised for the anisotropic analysis of clustering, subject of Chapter 6, and the corresponding estimator, which we dub YBS after the works of Yamamoto et al. (2000, 2006); Bianchi et al. (2015) and Scoccimarro (2015), will be derived in the aforementioned chapter.

3.2.2.1 Estimating the power spectrum

As already mentioned, one of the advantages of using the power spectrum instead of the two-point correlation function ξ is due to the fact that Fourier modes $\boldsymbol{k}$ are decoupled, at least at first order. It is also much simpler and fast to compute higher-order statistics, such as the bispectrum, because we can rely on Fast Fourier Transform (FFT) algorithms.

Here we present the most common method employed in galaxy redshift surveys to estimate the power spectrum from a pre-compressed data, namely the FKP method (Feldman et al., 1994). We follow the derivation of Jeong (2010), which is much more complete than the one of the original FKP paper. The reason why this method is applied to a pre-compressed data relies in the fact that it deals with $N_X(\boldsymbol{x}_{ijk})$ that can either be the count-in-cells of a random $X = r$ or the data (halo/galaxy) $X = \{h, g\}$ catalogue. Here, $\boldsymbol{x}_{ijk}$ is the position of each cell in a three-dimensional grid that can be obtained by a mass assignment scheme, e.g. Nearest Grid Point (NGP), Cloud In Cell (CIC), or Triangular Shaped Cloud (TSC), as shown in Figure 3.7.

We begin by defining the weighted galaxy fluctuation, or the overdensity field, as

$$F(\boldsymbol{x}) = \frac{w(\boldsymbol{x})}{\mathcal{N}} \left[n_h(\boldsymbol{x}) - \alpha\, n_r(\boldsymbol{x}) \right], \tag{3.41}$$

where $n_h(\boldsymbol{x}) = \sum_{i=1}^{N_h} \delta^D(\boldsymbol{x} - \boldsymbol{x}_i)$ is the number density that will be written as a grid, after a mass assignment scheme is chosen, and $\mathcal{N}$ is a normalisation factor. Therefore, in practice $n_h(\boldsymbol{x}) = N_h(\boldsymbol{x}_{ijk})$ is the count-in-cells grid and $n_r(\boldsymbol{x})$ is the corresponding quantity for the random catalogue, which is obtained by randomly sampling α^{-1} times more objects within the survey volume, with the same selection function as the real data.

It is more instructive to write $F(\boldsymbol{x}) \propto w(\boldsymbol{x}) \left[n(\boldsymbol{x}) - \bar{n}(\boldsymbol{x}) \right] = w(\boldsymbol{x})\, \bar{n}(\boldsymbol{x})\, \delta(\boldsymbol{x}) = W(\boldsymbol{x}) \delta(\boldsymbol{x})$, where we call $W(\boldsymbol{x})$ the window function. Then, in Fourier space $F(\boldsymbol{k})$ is the convolution of the

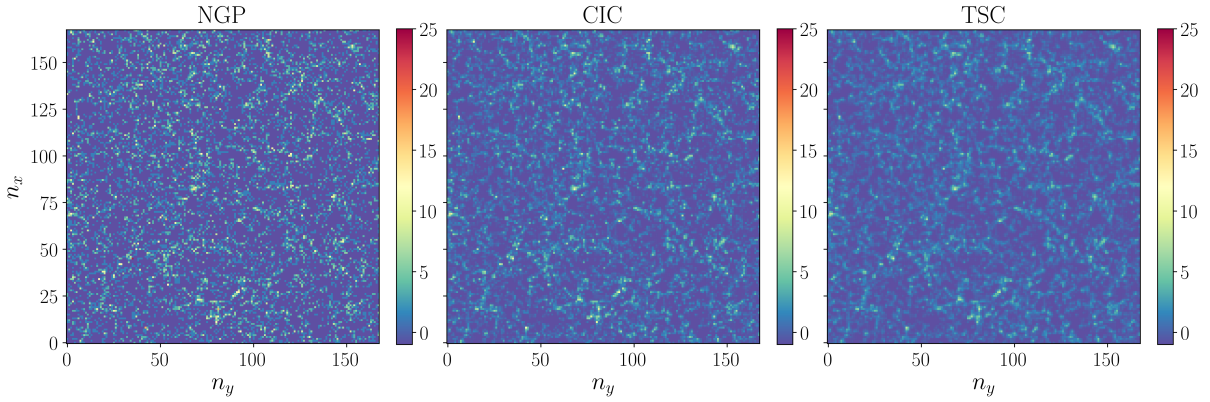


Figure 3.7: Different assignment schemes utilised to interpolate the data into a regular grid: Nearest Grid Point (NGP, *left*), Cloud In Cell (CIC, *middle*) and Triangular Shaped Cloud (TSC, *right*). We used the publicly available `Pylians` code to generate the grids.

window with the density contrast: $F(\boldsymbol{k}) = (2\pi)^{-3}\int \mathrm{d}^3q\, W(\boldsymbol{k}-\boldsymbol{q})\,\delta(\boldsymbol{q})$, and one can show that $\langle F(\boldsymbol{k})F(-\boldsymbol{k})\rangle = (2\pi)^{-3}\int \mathrm{d}^3q\, |W(\boldsymbol{k}-\boldsymbol{q})|^2\, P(\boldsymbol{q}) + \int \mathrm{d}^3x\, w^2(\boldsymbol{x})\,\bar{n}(\boldsymbol{x})$. Therefore, in Feldman et al. (1994) it is considered the overdensity field divided by the magnitude of the window function, $W^2 \equiv \int \mathrm{d}^3x\, W^2(\boldsymbol{x})$, which we called $\mathcal{N}$ (Jeong, 2010).

A little bit of intuition behind this definition: the observed halo (or galaxy) number density is modelled as $n(\boldsymbol{x}) = \bar{n}(\boldsymbol{x})[1+\delta(\boldsymbol{x})]$, where $\bar{n}$ defines the background (no clustering) mean number density. Deviations from the random (Poissonian) counting, which encodes cosmological effects, are given by the intrinsic fluctuation field $\delta(\boldsymbol{x})$. Thus, by subtracting a synthetic Poissonian catalogue (with no intrinsic clustering) from the real data yields the true cosmological fluctuation field that we aim.

Consider the Fourier transform of the overdensity field:

$$F(\boldsymbol{k}) = \frac{1}{\mathcal{N}}\int \mathrm{d}^3x\; w(\boldsymbol{x})\,\mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{x}}\,[n_h(\boldsymbol{x}) - \alpha\, n_r(\boldsymbol{x})]. \tag{3.42}$$

Then,

$$\langle F(\boldsymbol{k})F^*(\boldsymbol{k})\rangle = \frac{1}{\mathcal{N}^2}\int \mathrm{d}^3x \int \mathrm{d}^3x'\; w(\boldsymbol{x})\,w(\boldsymbol{x}')\;\mathrm{e}^{-i\boldsymbol{k}\cdot(\boldsymbol{x}-\boldsymbol{x}')}\times \\ \langle [n_h(\boldsymbol{x}) - \alpha\, n_r(\boldsymbol{x})][n_h(\boldsymbol{x}') - \alpha\, n_r(\boldsymbol{x}')]\rangle. \tag{3.43}$$

Since n_h and n_r are independent, $\langle n_h n_r\rangle = \langle n_h\rangle\langle n_r\rangle = \bar{n}_h\bar{n}_r$, and $\bar{n}_h = \alpha\bar{n}_r$, we have that

$$\langle n_h(\boldsymbol{x})n_h(\boldsymbol{x}')\rangle = \bar{n}_h(\boldsymbol{x})\,\bar{n}_h(\boldsymbol{x}')[1+\xi(\boldsymbol{x}-\boldsymbol{x}')] + \bar{n}_h(\boldsymbol{x})\,\delta^D(\boldsymbol{x}-\boldsymbol{x}'), \tag{3.44}$$

$$\langle n_r(\boldsymbol{x})\,n_r(\boldsymbol{x}')\rangle = \alpha^{-2}\,\bar{n}_h(\boldsymbol{x})\,\bar{n}_h(\boldsymbol{x}') + \alpha^{-1}\,\bar{n}_h(\boldsymbol{x})\,\delta^D(\boldsymbol{x}-\boldsymbol{x}'), \tag{3.45}$$

$$\langle n_h(\boldsymbol{x})\,n_r(\boldsymbol{x}')\rangle = \alpha^{-1}\,\bar{n}_h(\boldsymbol{x})\,\bar{n}_h(\boldsymbol{x}'), \tag{3.46}$$

and then

$$\begin{aligned}\langle|F(\boldsymbol{k})|^2\rangle &= \frac{1}{\mathcal{N}^2}\int \mathrm{d}^3x\,\mathrm{d}^3x'\ w\,w'\ \mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{\Delta x}}\times \\ &\qquad [\bar{n}\,\bar{n}'(1+\xi)+\bar{n}\,\delta^D+\bar{n}\,\bar{n}'+\alpha\,\bar{n}\,\delta^D-2\bar{n}\,\bar{n}'], \\ &= \frac{1}{\mathcal{N}^2}\int \mathrm{d}^3x\,\mathrm{d}^3x'\ w\,w'\mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{\Delta x}}\bar{n}\,\bar{n}'\,\xi + \\ &\qquad \frac{1}{\mathcal{N}^2}\int \mathrm{d}^3x\,\mathrm{d}^3x'\ w\,w'\mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{\Delta x}}\,(1+\alpha)\,\bar{n}\,\delta^D(\boldsymbol{x}-\boldsymbol{x}'), \\ &= \frac{1}{\mathcal{N}^2}\int \mathrm{d}^3x\,\mathrm{d}^3x'\ w\,w'\mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{\Delta x}}\bar{n}\,\bar{n}'\,\xi\ +\ \frac{(1+\alpha)}{\mathcal{N}^2}\int \mathrm{d}^3x\ w^2\,\bar{n}. \end{aligned} \tag{3.47}$$

The last equality came from the Dirac delta integration, yielding $\boldsymbol{x}=\boldsymbol{x}'$ and hence $\mathrm{e}^{-i\boldsymbol{k}\cdot(\boldsymbol{x}-\boldsymbol{x}')}=1$.

Concentrating on the first term:

$$\begin{aligned}\mathcal{I} &\equiv \frac{1}{\mathcal{N}^2}\int \mathrm{d}^3x\,\mathrm{d}^3x'\ w(\boldsymbol{x})\,w(\boldsymbol{x}')\,\bar{n}(\boldsymbol{x})\,\bar{n}(\boldsymbol{x}')\,\xi(\boldsymbol{x}-\boldsymbol{x}')\,\mathrm{e}^{-i\boldsymbol{k}\cdot(\boldsymbol{x}-\boldsymbol{x}')} \\ &= \frac{1}{\mathcal{N}^2}\int \frac{\mathrm{d}^3k'}{(2\pi)^3}\ \mathrm{e}^{i\boldsymbol{k}'\cdot(\boldsymbol{x}-\boldsymbol{x}')}\,P(\boldsymbol{k}')\int \mathrm{d}^3x\,\mathrm{d}^3x'\ w\,w'\,\bar{n}\,\bar{n}'\,\mathrm{e}^{-i\boldsymbol{k}\cdot(\boldsymbol{x}-\boldsymbol{x}')} \\ &= \frac{1}{\mathcal{N}^2}\int \frac{\mathrm{d}^3k'}{(2\pi)^3}\ P(\boldsymbol{k}')\left[\int \mathrm{d}^3x\ w\,\bar{n}\,\mathrm{e}^{-i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{x}}\ \int \mathrm{d}^3x'\ w'\,\bar{n}'\,\mathrm{e}^{i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{x}'}\right]\end{aligned}$$

$$\mathcal{I} = \frac{1}{\mathcal{N}^2}\int \frac{\mathrm{d}^3k'}{(2\pi)^3}\ P(\boldsymbol{k}')|W(\boldsymbol{k}-\boldsymbol{k}')|^2. \tag{3.48}$$

Therefore,

$$\langle|F(\boldsymbol{k})|^2\rangle = \frac{1}{\mathcal{N}^2}\int \frac{\mathrm{d}^3k'}{(2\pi)^3}\ P(\boldsymbol{k}')\,|W(\boldsymbol{k}-\boldsymbol{k}')|^2 + P_{\mathrm{shot}}, \tag{3.49}$$

where

$$P_{\mathrm{shot}} = \frac{(1+\alpha)}{\mathcal{N}^2}\int \mathrm{d}^3x\ w^2(\boldsymbol{x})\,\bar{n}(\boldsymbol{x}) \tag{3.50}$$

In the work of FKP,

$$\mathcal{N}^2 = \int \mathrm{d}^3x\,W^2(\boldsymbol{x}) \tag{3.51}$$

is the magnitude of the window function. To arrive at FKP's final estimator, we must assume $W(\boldsymbol{k}-\boldsymbol{k}')\sim\delta^D(\boldsymbol{k}-\boldsymbol{k}')$. This is true for $|\boldsymbol{k}|\gg 1/V^{1/3}$, with V the survey volume. However, at large scales, this assumption breaks down as $|\boldsymbol{k}|\sim 1/V^{1/3}$. However, when such constrain holds true, and the window can be approximated by a Dirac delta function, then

$$\langle|F(\boldsymbol{k})|^2\rangle \sim P(\boldsymbol{k}) + P_{\mathrm{shot}}. \tag{3.52}$$

Therefore, the final estimate will be given by the mean over k-shells:

$$\hat{P}(k) = \frac{1}{V_k} \int_{V_k} \mathrm{d}^3 k \left[\langle |F(\boldsymbol{k})|^2 \rangle - P_{\mathrm{shot}} \right]. \tag{3.53}$$

The weighting scheme that optimises the power-spectrum estimation is given by Feldman et al. (1994):

$$w(\boldsymbol{x}) = \frac{1}{\bar{n}(\boldsymbol{x}) P_0(k) + 1}, \tag{3.54}$$

where $P_0(k)$ is an *ansatz* corresponding to the expected power-spectrum amplitude at the region where the optimisation is desired (e.g. usually one takes the amplitude at the BAO scale for studies whose prime focus is the BAO signal).

There are two main limitations to this estimator: firstly, the FKP weight does not account for different biased tracers, i.e. it assumes the bias is one. This is accounted for in the work of Percival et al. (2004)[3], who generalised the FKP weights for multiple tracers, resulting in a minimum-variance estimate of the matter power spectrum. Besides, the FKP estimator assumes the *global* plane-parallel (PP) approximation, i.e. all points in the data and random catalogues have the same line-of-sight (LOS). Hence, this estimator is not optimal for surveys where the PP approximation breaks down. To overcome this limitation, one can resort Yamamoto estimator Yamamoto et al. (2006), further improved by Bianchi et al. (2015) and Scoccimarro (2015), to deal with a varying line-of-sight direction.

3.2.2.2 Estimating the bispectrum

The idea of this method is to generalise the FKP estimator of the previous section to get an estimator for the bispectrum. The three-point correlation function of the overdensity field $F(\boldsymbol{k})$ is given by:

$$\langle F(\boldsymbol{k}_1) F(\boldsymbol{k}_2) F(\boldsymbol{k}_3) \rangle = V \left[B_\mu(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3) + \frac{1}{\bar{n}} [P_\mu(k_1) + P_\mu(k_2) + P_\mu(k_3)] + \frac{1}{\bar{n}^2} \right] \tag{3.55}$$

In the case of a real survey, where the selection function varies with the position in real space, $\bar{n} = \bar{n}(\boldsymbol{x})$, by performing the same steps of the FKP case, for $F(\boldsymbol{x}) = \mathcal{N}^{-1} w(\boldsymbol{x}) [n(\boldsymbol{x}) - \alpha n_r(\boldsymbol{x})]$, we arrive at the generalised FKP ("GFKP") estimator

$$\langle F(\boldsymbol{k}_1) F(\boldsymbol{k}_2) F(\boldsymbol{k}_3) \rangle = \frac{I_{33}}{I_{22}^{3/2}} \left[B_\mu(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3) + \frac{I_{32}}{I_{33}} [P_\mu(k_1) + P_\mu(k_2) + P_\mu(k_3)] + \frac{I_{31}}{I_{33}} \right], \tag{3.56}$$

[3] Known as PVP estimator.

where

$$I_{ij} \equiv \int \mathrm{d}^3x \; w^i(\boldsymbol{x})\, \bar{n}^j(\boldsymbol{x}). \tag{3.57}$$

This has been first derived in Matarrese et al. (1997); then Scoccimarro (2000) showed this to be an optimal estimator for a single tracer, just as in the power spectrum case of Feldman et al. (1994), and it has been applied to the BOSS survey in Gil-Marín et al. (2015); Gil-Marín et al. (2015).

3.2.3 Quadratic compression

The quadratic compression aims at finding estimators $\hat{p}_\mu$ for the band-powers (bins in Fourier space) of the power spectrum, such that the CRLB is saturated: $\mathrm{Cov}(\hat{p}_\mu, \hat{p}_\nu) = F^{-1}_{\mu\nu}$. In analogy in the FKP method (see Martínez and Saar, 2001, for a thorough discussion), we define the quadratic form

$$\hat{q}_\mu = \sum_{ij} E^{ij}_\mu \, \delta_i \, \delta_j - \Delta_\mu, \tag{3.58}$$

where E^{ij}_μ is some weighting function (e.g. FKP) we must find in order to have an optimal estimator, and Δ_μ is the estimator bias we must properly subtract in order to have an accurate estimator (e.g. the shot-noise term in the FKP estimator). As we shall clarify further, the power spectrum can thus defined as the linear transformation

$$\hat{p}_\mu = \sum_\alpha F^{-1}_{\mu\alpha} \hat{q}_\alpha. \tag{3.59}$$

The covariance of the estimator defined in Equation (3.59) then reads:

$$\begin{aligned}\mathrm{Cov}(\hat{p}_\mu, \hat{p}_\nu) &= \mathrm{Cov}\left(\sum_\alpha F^{-1}_{\mu\alpha}\hat{q}_\alpha, \sum_\beta F^{-1}_{\nu\beta}\hat{q}_\beta\right) = \sum_{\alpha\beta} F^{-1}_{\mu\alpha}F^{-1}_{\nu\beta}\langle\hat{q}_\alpha\hat{q}_\beta\rangle - \sum_\alpha F^{-1}_{\mu\alpha}\langle\hat{q}_\alpha\rangle \sum_\beta F^{-1}_{\nu\beta}\langle\hat{q}_\beta\rangle \\ &= \sum_{\alpha\beta} F^{-1}_{\mu\alpha}F^{-1}_{\nu\beta}\left(\langle\hat{q}_\alpha\hat{q}_\beta\rangle - \langle\hat{q}_\alpha\rangle\langle\hat{q}_\beta\rangle\right).\end{aligned} \tag{3.60}$$

Because Δ_α is not a random variable, the expected values of the quadratic form is given by

$$\langle\hat{q}_\alpha\rangle = E^{ij}_\alpha\langle\delta_i\delta_j\rangle - \Delta_\alpha, \tag{3.61}$$

implying that

$$\langle\hat{q}_\alpha\rangle\langle\hat{q}_\beta\rangle = E^{ij}_\alpha E^{mn}_\beta\langle\delta_i\delta_j\rangle\langle\delta_m\delta_n\rangle - E^{ij}_\alpha\langle\delta_i\delta_j\rangle\Delta_\beta - E^{mn}_\beta\langle\delta_m\delta_n\rangle\Delta_\alpha, \tag{3.62}$$

and

$$\langle \hat{q}_\alpha \hat{q}_\beta \rangle = E^{ij}_\alpha E^{mn}_\beta \langle \delta_i \delta_j \delta_m \delta_n \rangle - E^{ij}_\alpha \langle \delta_i \delta_j \rangle \Delta_\beta - E^{mn}_\beta \langle \delta_m \delta_n \rangle \Delta_\alpha. \tag{3.63}$$

Therefore,

$$\langle \hat{q}_\alpha \hat{q}_\beta \rangle - \langle \hat{q}_\alpha \rangle \langle \hat{q}_\beta \rangle = E^{ij}_\alpha E^{mn}_\beta (\langle \delta_i \delta_j \delta_m \delta_n \rangle - \langle \delta_i \delta_j \rangle \langle \delta_m \delta_n \rangle). \tag{3.64}$$

Now a crucial assumption is made about the data δ: they are *Gaussianly* distributed. Then, we can reduce the four-point function in the covariance to a product of covariance matrices as

$$\langle \hat{q}_\alpha \hat{q}_\beta \rangle - \langle \hat{q}_\alpha \rangle \langle \hat{q}_\beta \rangle = E^{ij}_\alpha E^{mn}_\beta (C_{im} C_{jn} + C_{jm} C_{in}). \tag{3.65}$$

From the cyclic property of the trace of matrices, we can write $\mathrm{Cov}(\hat{p}_\mu, \hat{p}_\nu) = \sum_{\alpha\beta} F^{-1}_{\mu\alpha} F^{-1}_{\nu\beta} \mathrm{Cov}(\hat{q}_\alpha, \hat{q}_\beta)$, and thus is straightforward to see that, by imposing $\mathrm{Cov}(\hat{q}_\alpha, \hat{q}_\beta) = F_{\alpha\beta}$, then $\mathrm{Cov}(\hat{p}_\mu, \hat{p}_\nu) = F^{-1}_{\mu\nu}$, saturating the CRLB. From this condition,

$$\mathrm{Cov}(\hat{q}_\alpha, \hat{q}_\beta) = \sum_{ij} \sum_{mn} E^{ij}_\alpha E^{mn}_\beta (C_{im} C_{jn} + C_{jm} C_{in}) \stackrel{!}{=} F_{\alpha\beta}, \tag{3.66}$$

we derive the weights E:

$$E_\alpha = \frac{1}{2} C^{-1} \frac{\partial C}{\partial p_\alpha} C^{-1} = \frac{1}{2} \sum_{k\ell} C^{-1}_{ik} \frac{\partial C_{k\ell}}{\partial p_\alpha} C^{-1}_{\ell j}. \tag{3.67}$$

The bias term can be found by imposing $\langle \hat{p}_\mu \rangle = \bar{p}_\mu$, that is, the expected value of the estimator is equal to the theoretical expectation. It is given by

$$\Delta_\mu = \frac{1}{2} \sum_{k\ell} C^{-1}_{\ell k} \frac{\partial C_{k\ell}}{\partial p_\mu} - \sum_\beta F_{\mu\beta} \bar{p}_\beta. \tag{3.68}$$

Therefore, we see that in order to obtain optimal estimators for the power spectrum via $\hat{p}_\mu = \sum_\alpha F^{-1}_{\mu\alpha} \hat{q}_\alpha$, one needs the Fisher matrix. In Appendix D.2 we present the derivation of Abramo et al. (2015), henceforth ASL, in all of its details, leading to the multi-tracer Fisher matrix that allows us to combine different tracers in order to improve the measurements of the individual power spectra.

3.2.4 Cubic compression

In Heavens (1998), an optimal bispectrum estimator designed for the CMB was found. During the process, the Gaussian hypothesis was used to write down the covariance for the cubic form

$$\hat{y}_\alpha = E^{ijk}_\alpha x_i x_j x_k, \tag{3.69}$$

where x_i is the CMB temperature fluctuation $\Delta T/T$ in some pixel i. While the Gaussian assumption is true for the CMB, as the vast number of experiments shows us, it is a widespread belief that the same does not hold even for the largest scales of the LSS (Szapudi, 2008). Also, as mentioned in Heavens (1998), the breakdown of Gaussianity might bias the estimates obtained from the estimator in Equation (3.69).

The approach taken in Heavens (1998) follows, nonetheless, as a natural generalisation of the quadratic procedures developed for the power spectrum, and it will be used in Chapter 7 as well. Therefore, in this section we make a complete derivation of the results presented in Heavens (1998).

In Heavens (1998), α is the shorthand notation for a specific triangular configuration. Thus, in order to not confuse the indices associated with specific LSS tracers, we will change the notation such that, in our case, α will be the triplet $\Delta = \{k_1, k_2, k_3\}$, renaming the cubic form $\hat{y}_\alpha \to \hat{\Gamma}_\Delta$. We stress the fact that i, j and k refer to the N pixels under analysis, and instead of CMB fluctuations one could, in principle, think about the galaxy number count fluctuation at a certain pixel i in real space, δ_i. Also, recall that the real space 3PCF is connected to bispectrum as

$$\zeta_{ijk} \equiv \langle \delta_i \delta_j \delta_k \rangle = \sum_\Delta B_\Delta R^\Delta_{ijk}, \tag{3.70}$$

with R connecting real space with Fourier space.

By inspection of Equation (3.69), and because we want that to be an estimator for the bispectrum, one infers that E^{ijk}_Δ translates from real space to Fourier space, giving the additional constrain $\mathrm{Tr}[E^\Delta_{ijk} R^\Delta_{ijk}] = 1$.

Under the *Gaussian* approximation[4], the covariance of the cubic form $\hat{\Gamma}$ is given by

$$\mathrm{Cov}(\hat{\Gamma}_\Delta, \hat{\Gamma}_{\Delta'}) = \xi_{ii'}(6\,\xi_{jj'}\xi_{kk'} + 9\,\xi_{jk}\xi_{j'k'}) E^\Delta_{ijk} E^{\Delta'}_{i'j'k'}, \tag{3.71}$$

and the variance is

$$V_{\Delta\Delta} = \xi_{ii'}(6\,\xi_{jj'}\xi_{kk'} + 9\,\xi_{jk}\xi_{j'k'}) E^\Delta_{ijk} E^\Delta_{i'j'k'}. \tag{3.72}$$

To derive Equation (3.71) we employed Wick's theorem to compute the six-point correlation function (6PCF)

$$\langle \delta_i \delta_j \delta_k \delta_{i'} \delta_{j'} \delta_{k'} \rangle = \langle \delta_i \delta_j \delta_k \rangle \langle \delta_{i'} \delta_{j'} \delta_{k'} \rangle + \langle \delta_i \delta_j \rangle \langle \delta_k \delta_{i'} \rangle \langle \delta_{j'} \delta_{k'} \rangle + 14 \text{ perms.}, \tag{3.73}$$

taking $\langle \delta\delta\delta \rangle \approx 0$ for any combination of indices due to the Gaussian assumption. For the disconnected piece of the covariance, given in terms of 2PCFs, we have a total of 15 terms that

[4]We emphasise the fact that, once again, this is performed for a Gaussianly distributed data, a reasonable approximation for the CMB that renders the validity of applying this estimator for CMB studies.

can be separated according to the mixing of "primed" and "unprimed" indices (see Table 3.1).

Table 3.1: Permutations for the disconnected pieces of the 6PCF. We have six permutations where all 2PCFs mix primed and unprimed pixels (in red and inside brackets) and the remaining 9 permutations consist of only one 2PCF mixing both terms.

$ij\ \ ki'\ \ j'k'$	$ik\ \ jk'\ \ i'j'$	$[ij'\ \ ji'\ \ kk']$
$ij\ \ kj'\ \ i'k'$	$ii'\ \ jk\ \ j'k'$	$[ij'\ \ jk'\ \ ki']$
$ij\ \ kk'\ \ i'j'$	$[ii'\ \ jj'\ \ kk']$	$ik'\ \ jk\ \ i'j'$
$ik\ \ ji'\ \ j'k'$	$[ii'\ \ jk'\ \ kj']$	$[ik'\ \ ji'\ \ kj']$
$ik\ \ jj'\ \ i'k'$	$ij'\ \ jk\ \ i'k'$	$[ik'\ \ jj'\ \ ki']$

We minimise the variance in Equation (3.72) with respect to the weights as

$$\xi_{ii'}\,(6\,\xi_{jj'}\,\xi_{kk'} + 9\,\xi_{jk}\,\xi_{j'k'})\,E^{\Delta}_{i'j'k'} = \lambda\,R^{\Delta}_{ijk}, \tag{3.74}$$

where λ is the Lagrange multiplier ensuring the trace constraint $\mathrm{Tr}[E^{ijk}_{\Delta}R^{\Delta}_{ijk}] = 1$. Multiplying both sides by $\xi^{-1}_{j''j}\xi^{-1}_{i''i}$:

$$\lambda\,R^{\Delta}_{ijk}\,\xi^{-1}_{i''i}\,\xi^{-1}_{j''j} = \delta^{K}_{i''i'}\,(6\,\delta^{K}_{j'j''}\,\xi_{kk'} + 9\,\delta^{K}_{kj''}\,\xi_{j'k'})\,E^{\Delta}_{i'j'k'}, \tag{3.75}$$

where $\delta^{K}_{kj''}$ is the Kronecker delta. The left-hand side has the implicit sum over i and j, meaning that i'', j'' and k are free indices. Therefore, we take the trace with respect to j'' and k,

$$\lambda\,R^{\Delta}_{ijk}\,\xi^{-1}_{kj}\,\xi^{-1}_{i''i} = 6\,\xi_{kk'}\,\delta^{K}_{kj'}\,E^{i''j'k'}_{\Delta} + 9\,\delta^{K}_{kk}\,\xi_{j'k'}\,E^{\Delta}_{i''j'k'}, \tag{3.76}$$

and arrive at

$$\lambda\,R^{\Delta}_{ijk}\,\xi^{-1}_{kj}\,\xi^{-1}_{i''i} = 3(2+3N)\,\xi_{kk'}\,E^{\Delta}_{i''kk'}, \tag{3.77}$$

by noticing that δ^{K}_{kk} is the total number of pixels N. The above equation yields the relation

$$\xi_{kk'}\,E^{\Delta}_{i''kk'} = \frac{\lambda}{3(2+3N)}\,R^{\Delta}_{ijk}\,\xi^{-1}_{kj}\,\xi^{-1}_{i''i}, \tag{3.78}$$

which we can use in Equation (3.76) to arrive at the final weights

$$E^{\Delta}_{i'j'k'} = \frac{1}{6}\xi^{-1}_{ii'}\left[\xi^{-1}_{jj'}\,\xi^{-1}_{kk'} - \left(\frac{3}{2+3N}\right)\xi^{-1}_{jk}\,\xi^{-1}_{j'k'}\right]R^{\Delta}_{ijk}. \tag{3.79}$$

With Equation (3.79) and the expression derived in the following section for the Fisher matrix, we can write the bispectrum estimator as

$$\hat{B} = \sum_{\Delta'} F^{-1}_{\Delta\Delta'}\,\hat{\Gamma}_{\Delta'}. \tag{3.80}$$

Covariance of the cubic form Now we know that if $\hat{\Gamma}_\Delta$ saturates the CRLB, then $V_{\Delta\Delta'} = F_{\Delta\Delta'}$. From that we can find the inverse covariance and check if indeed $CC^{-1} = \mathbb{1}$ is satisfied.

The Fisher matrix is given by

$$F_{\Delta\Delta'} = \frac{1}{2}\mathrm{Tr}\left[C^{-1}\frac{\partial C}{\partial B_\Delta}C^{-1}\frac{\partial C}{\partial \Delta'} + 2C^{-1}\frac{\partial \zeta}{\partial B_\Delta}\frac{\partial \zeta}{\partial B_{\Delta'}}\right]. \tag{3.81}$$

Since the covariance of the data does not depend upon the signal, due to the Gaussian hypothesis, the Fisher matrix gets simplified to

$$F_{\Delta\Delta'} = \mathrm{Tr}\left[C^{-1}\frac{\partial \zeta}{\partial B_\Delta}\frac{\partial \zeta}{\partial B_{\Delta'}}\right] = C^{-1}_{ijki'j'k'}R_{ijk}R_{i'j'k'}. \tag{3.82}$$

Imposing the CRLB:

$$\begin{aligned} C^{-1}_{ijki'j'k'}R_{ijk}R_{i'j'k'} \overset{!}{=} \xi_{ii'}\,(6\,\xi_{jj'}\xi_{kk'} + 9\,\xi_{jk}\,\xi_{j'k'})\frac{1}{6}\,\xi^{-1}_{ii''}\left[\xi^{-1}_{jj''}\,\xi^{-1}_{kk''} - \left(\frac{3}{2+3N}\right)\xi^{-1}_{jk}\,\xi^{-1}_{j''k''}\right]R_{i''j''k''} \\ \frac{1}{6}\,\xi^{-1}_{i'i'''}\left[\xi^{-1}_{j'j'''}\,\xi^{-1}_{k'k'''} - \left(\frac{3}{2+3N}\right)\xi^{-1}_{j'k'}\,\xi^{-1}_{j'''k'''}\right]R_{i'''j'''k'''} \end{aligned} \tag{3.83}$$

Since $\xi_{ii'}$, $\xi^{-1}_{ii''} = \delta_{i'i''}$,

$$\begin{aligned} C^{-1}_{ijki'j'k'}RR' = \frac{1}{36}\xi^{-1}_{i''i'''}\,(6\,\xi_{jj'}\,\xi_{kk'} + 9\,\xi_{jk}\,\xi_{j'k'})\left[\xi^{-1}_{jj''}\,\xi^{-1}_{kk''} - \left(\frac{3}{2+3N}\right)\xi^{-1}_{jk}\,\xi^{-1}_{j''k''}\right] \\ \left[\xi^{-1}_{j'j'''}\,\xi^{-1}_{k'k'''} - \left(\frac{3}{2+3N}\right)\xi^{-1}_{j'k'}\,\xi^{-1}_{j'''k'''}\right]R''\,R''' \end{aligned} \tag{3.84}$$

where we suppressed the indices i', j' and k', for example, to R' for simplicity. By expanding all the terms, we can find many contractions of the form $\xi_{jj'}\xi^{-1}_{jj''}$, and also terms that lead to $\delta_{j'j'} = N$, for instance. After playing around with some dummy indices and rearranging the terms, we arrive at

$$C^{-1}_{ijki'j'k'} = \frac{1}{6}\,\xi^{-1}_{ii'}\left[\xi^{-1}_{jj'}\,\xi^{-1}_{kk'} - \left(\frac{3}{2+3N}\right)\xi^{-1}_{jk}\,\xi^{-1}_{j'k'}\right]. \tag{3.85}$$

It is straightforward to verify that indeed $C_{ijki'j'k'}\,C^{-1}_{i'j'k'ijk} = \delta^K_{ijkijk} = N^3$.

We intend to generalise the estimator of Equation (3.80) to be applicable to multiple LSS tracers. If this can be achieved, one should bear in mind that such estimator should reduce to Equation (3.56) in the single-tracer limit.

Chapter 4

Primordial Features in the Large-Scale Structure

In this chapter we address the issue of observing the effects of primordial non-Gaussianities (PNG) in both the power spectrum and bispectrum of a set of mock galaxy catalogues. This work is at an initial stage, and hence here we will motivate, discuss the effects of the PNG type we want to consider in the analysis and the future prospects to conclude. This is being carried in collaboration with Rodrigo Voivodic.

4.1 Introductory remarks

Did the Universe have a beginning? Is it eternal, or only one tiny part of the whole? These are profound questions that, for a long time, were thought to be out of the scope of science. Moreover, around 100 years ago few people thought that our galaxy was only one in a myriad of other galaxies; the development of telescopes allowed us to see the oodles of galaxies outside, shifting our view of the cosmos. The scientific treatment of such fundamental questions, developed throughout the past century, started to engage the philosophical query of whether our universe is only one patch in a multitude of other patches (in the jargon, a part of a multiverse). In this work, we are not interested in such idea, but rather in the scientific model that coined it: cosmic inflation.

Tracing the energy scales of an expanding universe, we reach temperatures and densities that are not probed by particle accelerators. It is estimated that, from $t \sim 10^{-11}$ to $\sim 10^{-5}$ seconds, the universe was dominated by the so-called plasma of quarks and gluons, by the time when the electromagnetic and weak forces had decoupled into two different interactions. Prior to that, at around $t \sim 10^{-35}$ seconds or so, temperatures were high enough ($T \sim 10^{29}\,^{\circ}$C) so that the electroweak and the strong forces were unified during the period of the hypothetical Grand Unified Theories (GUTs). Around this point, but prior to what we believe to be an era of

quantum gravity, we can hypothesise a period in which the universe was filled by one (or many other) scalar field. The dynamics of a spacetime filled with such a "fluid" whose energy density overcomes the potential one is similar to a universe dominated by a cosmological constant. This leads us to an era of exponential expansion frequently described by what is called the inflationary scenario (Guth, 1981).

As briefly outlined in Chapter 1, inflation gives the initial conditions for the process of structure formation, and the prediction of a nearly scale-invariant spectrum of the primordial density fluctuations. It also solves issues regarding the almost flat geometry inferred by CMB observations, and the so-called horizon problem, the latter related with the CMB's high level of homogeneity. These predictions and the elegant exit of existing problems with the standard Hot Big Bang theory (which does not account for an early period of inflation) made inflation the most studied model to describe the early universe. Within this context, the scalar field that dominates the universe, driving the exponential expansion, is dubbed the inflaton field.

The simplest class of inflationary models, containing only one slowly rolling inflaton field with canonical kinetic term in the action, leads to approximately scale invariant power spectrum and an almost null bispectrum (Fourier transform of the 3-point correlation function [3PCF]), with the amplitude depending upon the parameters related to the potential of inflaton (Maldacena, 2003), also known by slow-roll parameters. The small bispectrum amplitude emerges from the coupling of the inflaton field with the gravity sector, establishing a necessary level of non-Gaussianity for all inflationary models, but which is too small to be detected. However, intrinsic non-linear gravitational evolution of the perturbations leads to a non-vanishing bispectrum at late times (Equation 3.14), making constraints on primordial non-Gaussianities (PNG) from this specific statistics an observational challenge.

On one hand, from the power spectrum point-of-view, a scale-dependent bias (Dalal et al., 2008) emerge from quadratic corrections to the gravitational potential coming from PNG of the *local* type. This specific type of non-Gaussianity is parametrised by the non-linear coupling constant[1] $f_{\rm NL}^{\rm loc}$ (Komatsu and Spergel, 2001). This quantifies the contributions coming from squeezed triangles (on the left (a) of Figure 3.2) to the bispectrum amplitude. Current constraints from the Planck collaboration (Akrami et al., 2020) set the levels of non-Gaussianity to

$$f_{\rm NL}^{\rm loc} = -0.9 \pm 5.1. \tag{4.1}$$

On the other hand, more complex inflationary models (with more scalar fields and/or non-trivial kinetic terms in the action, etc Komatsu et al., 2009) and possible alternatives to

[1]The parameter $f_{\rm NL}$ describes the amplitude of modulations induced by the primordial mode coupling (de Putter, 2018).

inflation[2] have a non-vanishing bispectrum that can be parametrised by different $f_{\rm NL}^{\triangle}$, manifested in distinct triangular shapes of the bispectrum and 3PCF, where $\triangle$ refers to some specific triangular shape. For the equilateral ($k_1 = k_2 = k_3$; see middle triangle (b) of Figure 3.2) and orthogonal (combination of squeezed and equilateral shapes), Akrami et al. (2020) constraints give us, respectively,

$$f_{\rm NL}^{\rm eq} = -26 \pm 47 \quad \text{and} \quad f_{\rm NL}^{\rm orth} = -38 \pm 24. \tag{4.2}$$

Owing to its rich information content on the particle physics and interactions of the early universe, therefore at high energies unattainable with particle colliders, detecting PNG is one of the major goals of observational cosmology and can provide insights into high-energy physics (Lyth and Riotto, 1999; Lyth, 2008; Baumann and McAllister, 2015). For single-field inflation, the consistency condition yields $f_{\rm NL} = -5(n_s - 1)/12 \sim 10^{-2}$ (Maldacena, 2003; de Putter, 2018). However, multi-field models predict $f_{\rm NL}^{\triangle} \sim 1$ (for an example see Abazajian et al., 2015; Alvarez et al., 2014; de Putter et al., 2017), making the detection threshold of $\sigma(f_{\rm NL}) \sim \mathcal{O}(10^{-2})$ the ultimate goal for future surveys.

4.2 Theoretical framework

Perturbations in the inflaton field during a possible inflationary stage drive perturbations to the gravitational potential ϕ that source the CMB anisotropies and the large-scale structures. If these perturbations were perfectly Gaussian, then they would be completely described by their two-point statistics, given by the primordial power spectrum

$$\mathcal{P}_\Phi(k) \sim A_s \left(\frac{k}{k_\star}\right)^{n_s - 1}, \tag{4.3}$$

where A_s is the primordial amplitude, $k_\star$ is a pivot scale and n_s is the scalar index of the perturbations. The latest CMB constraints gives $A_s \sim 2.1 \times 10^{-9}$, $k_\star = 0.05$ Mpc^{-1} and $n_s \sim 0.96$ (Aghanim et al., 2020).

Even though there are secondary sources of non-Gaussianities, such as the non-linear evolution driven by gravitational collapse, Equation (3.14), or the non-linear effects of clustering encoded in the second-order bias, Equation (3.13), here we focus on the primordial contributions coming from fundamental interactions during an inflationary stage.

[2] By far inflation is not the only viable theory for the primordial universe. Many theories, such as the conformal-cyclic cosmology (Penrose, 2006, 2018), string-gas cosmology (Brandenberger, 2008), and ekpyrosis (Khoury et al., 2001) provide alternative approaches to studying the early universe.

4.2.1 Primordial non-Gaussianity of the local type

The local type of PNG consists in the simplest non-linear (NL) quadratic correction to the Bardeen potential Φ from the Taylor expansion in configuration space $\boldsymbol{x}$, hence the name *local*:

$$\Phi(\boldsymbol{x}) = \phi(\boldsymbol{x}) + f_{\rm NL}(\phi^2(\boldsymbol{x}) + \langle\phi^2(\boldsymbol{x})\rangle) \equiv \phi(\boldsymbol{x}) + \Phi_{\rm NL}(\boldsymbol{x}). \tag{4.4}$$

The last term in the first equality is added to yield $\langle\Phi\rangle = 0$ and ϕ is the Bardeen potential in the absence of non-Gaussianties (Komatsu and Spergel, 2001).

In Fourier space, the quadratic correction $\Phi_{\rm NL}$ becomes a convolution

$$\Phi_{\rm NL}(\boldsymbol{x}) = \frac{f_{\rm NL}}{(2\pi)^3}\int \mathrm{d}^3k'\ \phi(\boldsymbol{k}-\boldsymbol{k}')\,\phi(\boldsymbol{k}'), \tag{4.5}$$

which leads to a non-vanishing three-point correlation function:

$$\langle\Phi(\boldsymbol{k}_1)\Phi(\boldsymbol{k}_2)\Phi(\boldsymbol{k}_3)\rangle = 2f_{\rm NL}(2\pi)^3\delta^D(\boldsymbol{k}_1+\boldsymbol{k}_2+\boldsymbol{k}_3)\left[P_\Phi(\boldsymbol{k}_1)P_\Phi(\boldsymbol{k}_2) + \ 2\ \text{perms}\right]. \tag{4.6}$$

Most bispectra used to parametrise PNG are separable, that is:

$$B_\Phi(k_1,k_2,k_3) \propto f_{\rm NL}\mathcal{S}(k_1,k_2,k_3), \tag{4.7}$$

where the *shape* $\mathcal{S}$ codifies the bispectrum dependence on the vector modes $\boldsymbol{k}_i$ (as in Figure 3.2), $f_{\rm NL}$ gives the amplitude of the signal at that shape, and $k_3 = |k_1 - k_2|$ is needed to ensure translational invariance of the statistic. From the assumption of statistical homogeneity and isotropy, the bispectrum and shape depends only on the magnitude of the wavenumbers k_1, k_2 and k_3, that is, on the size of the triangles. Nonetheless, the break of isotropy coming from, for example, redshift-space distortions introduces a dependence on the orientation of the triangles (e.g., see Nan et al., 2018).

Plugging the scale-invariant power spectrum obtained from inflation, $P_\Phi(k) = \mathcal{P}_\Phi k^{-3}$, into Equation (4.6) and comparing the result with the definition of the bispectrum, Equation (3.11), we can write

$$B_\Phi(k_1,k_2,k_3) \approx 2f_{\rm NL}\frac{\mathcal{P}_\Phi^2}{(k_1k_2k_3)^2}\left(\frac{k_1^2}{k_2k_3}+\frac{k_2^2}{k_1k_3}+\frac{k_3^2}{k_1k_2}\right), \tag{4.8}$$

and then the shape of non-Gaussianities $S(k_1,k_2,k_3)$ can be defined as

$$\mathcal{S}(k_1,k_2,k_3) \equiv \frac{(k_1k_2k_3)^2}{N}B_\Phi(k_1,k_2,k_3), \tag{4.9}$$

where N is a normalisation introduced to give $\mathcal{S}(k,k,k) = 1$ when we consider a bispectrum of

the local form, Equation (4.6), i.e. $N = 6 f_{NL} \mathcal{P}_\Phi^2$.

Therefore, we define the local shape as

$$\mathcal{S}^{\rm loc}(k_1, k_2, k_3) = \frac{1}{3}\left(\frac{k_1^2}{k_2 k_3} + \frac{k_2^2}{k_1 k_3} + \frac{k_3^2}{k_1 k_2}\right), \tag{4.10}$$

which is straightforward to see that it peaks at the squeezed configuration, i.e. $k_i \ll k_j \sim -k_\ell$.

4.2.1.1 Conventions

The introduction of quadratic corrections in the primordial potential leaves room two possible conventions: the Bardeen potential can be normalised at decoupling (CMB convention) or at low redshifts (LSS) (Camera et al., 2015). In the first convention, it is common in the literature to write

$$\Phi_{\rm CMB}(\boldsymbol{x}) = \phi_{\rm CMB} + f_{\rm NL}^{\rm CMB}(\phi_{\rm CMB}^2 - \langle \phi_{\rm CMB}^2 \rangle), \tag{4.11}$$

where we introduced the non-local parameter evaluated at decoupling[3], $f_{\rm NL}^{\rm CMB}$. The second possibility is to write, instead,

$$\Phi_{\rm LSS}(\boldsymbol{x}) = \phi_{\rm LSS} + f_{\rm NL}^{\rm LSS}(\phi_{\rm LSS}^2 - \langle \phi_{\rm LSS}^2 \rangle), \tag{4.12}$$

with $f_{\rm NL}^{\rm LSS}$ the $f_{\rm NL}$ value at low redshifts. This means that different probes (CMB experiments vs. LSS surveys) yield different values for the primordial non-Gaussian shape parametrised by $f_{\rm NL}^{\rm loc}$, simply because the Gaussian and non-Gaussian potentials decay as $g(z) = D(z)(1+z)$, on linear scales, with $D(z)$ the growth function. We reinforce that Equations (4.11) and (4.12) above are defined in configuration space, and omit the variable dependence in what follows for simplicity.

Connecting the potentials for both LSS and CMB, i.e.,

$$\phi_{\rm CMB} = \frac{g(z_{\rm dec})}{g(z)}\phi_{\rm LSS}, \tag{4.13}$$

and

$$\Phi_{\rm CMB} = \frac{g(z_{\rm dec})}{g(z)}\Phi_{\rm LSS}, \tag{4.14}$$

we can rewrite Equation (4.11) in terms of Equations (4.13) and (4.14) as

$$\Phi_{\rm LSS} = \phi_{\rm LSS} + f_{\rm NL}^{\rm CMB}\frac{g(z_{\rm dec})}{g(z)}(\phi_{\rm LSS}^2 - \langle \phi_{\rm LSS}^2 \rangle), \tag{4.15}$$

[3] Some authors call $z_{\rm dec} = \infty$ (for example see Pillepich et al., 2010).

and connect $f_{\rm NL}^{\rm CMB}$ with $f_{\rm NL}^{\rm LSS}$ via Equation (4.12) to arrive at

$$f_{\rm NL}^{\rm LSS} \equiv f_{\rm NL}^{\rm CMB} \frac{g(z_{\rm dec})}{g(z)}. \tag{4.16}$$

This will be important not only to compare LSS and CMB constraints on PNG, but also to generate non-Gaussian simulations (see Section 4.3) with reasonable $f_{\rm NL}$ values coming from the CMB constraints of Akrami et al. (2020).

4.2.2 The effect of PNG on halos

In addition to the characteristic bispectrum shown in Equation (4.8), Dalal et al. (2008) found that both the mass and correlation functions of massive halos are significantly altered by this quadratic correction. By taking Equation (4.4) to Fourier space, and relating $\phi(\boldsymbol{k})$ to $\delta(\boldsymbol{k})$ via the Poisson equation,

$$\phi = \frac{3}{2}\frac{\Omega_m}{a}\frac{H_0^2}{k^2}\delta, \tag{4.17}$$

we arrive at their main result:

$$\delta_{\rm NG} = \delta_{\rm G}(1 + 2 f_{\rm NL}\phi), \tag{4.18}$$

where we recall that ϕ is the primordial Gaussian potential. Hence, there is an overall bias $b_\alpha(z) \to b_\alpha(z) + \Delta b_\alpha(k, z, f_{\rm NL})$ which acquires a k^{-2} scale-dependence of the form:

$$\Delta b_\alpha(k, z, f_{\rm NL}) = f_{\rm NL}^{\rm loc}(b_\alpha - 1)\frac{3\,\delta_c\,\Omega_{m,0}\,H_0^2}{c^2\,k^2\,T(k)\,D(z)}, \tag{4.19}$$

where $\delta_c = 1.68$ is the critical density linearly extrapolated for a spherical collapse in an Einstein-de Sitter universe, $T(k)$ is the matter transfer function normalised at large scales, $T(k \to 0) = 1$, and $D(z)$ is the growth factor.

Slosar et al. (2008) pointed out that for the case of a mass function which is not universal, depending properties such as the merging history of the halos, the scale dependent bias becomes

$$\Delta b_\alpha(k, z, f_{\rm NL}) = f_{\rm NL}^{\rm loc}(b_\alpha - p)\frac{3\,\delta_c\,\Omega_{m,0}\,H_0^2}{c^2\,k^2\,T(k)\,D(z)}, \tag{4.20}$$

where $p = 1$ is the case of an universal mass function (Dalal et al., 2008), $p = 1.6$ stands for objects that populate only recently merged halos (Slosar et al., 2008). With this caveat, they observed that constraints for $f_{\rm NL}$ were degraded in comparison with the case of Equation (4.19). A more recent analysis was done by Castorina et al. (2019) using the eBOSS DR14, yielding tighter constraints of $-11 < f_{\rm NL} < 29$ ($p = 1$) and $-19 < f_{\rm NL} < 59$ ($p = 1.6$). More recently, Barreira et al. (2020) found $p \approx 0.55$ for galaxies selected by their stellar mass (see also Barreira, 2020), stressing the impact of the parameter p for $f_{\rm NL}$ constraints.

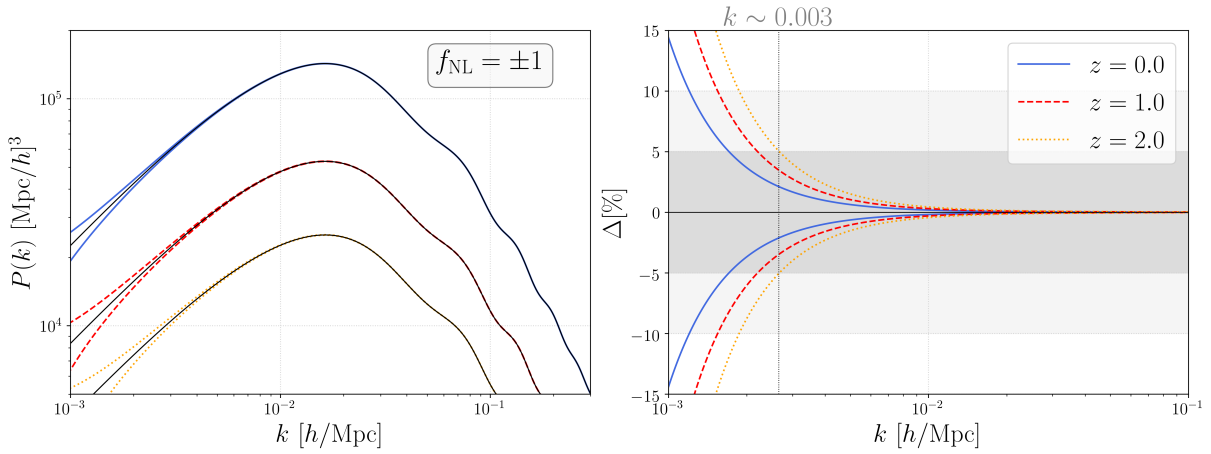


Figure 4.1: (*Left panel*) Corrections induced by local-type PNG in the power spectrum of dark matter halos of mass $M = 10^{15}\,[M_\odot/h]$, assuming no merger history (simplified assumption that can be easily broken) at $z = 0$ (*blue*), $z = 1$ (*red*) and $z = 2$ (*orange*). Linear bias is given by the fit of Tinker et al. (2010). Positive values of $f_{\rm NL}^{\rm loc} = f_{\rm NL}$ correspond to the curves above the fiducial value (*solid black*) $f_{\rm NL} = 0$, whereas $f_{\rm NL} < 0$ corresponds to a loss of power at large scales. (*Right panel*) Percent difference between the different scenarios with respect to the fiducial case. Dotted vertical line (*black*) marks the 5% difference at redshift $z = 2$, corresponding to $k \sim 0.003$ h/Mpc and $L_{\rm box} \sim 2$ Gpc/h.

Table 4.1: Characteristics of the simulations utilised in this work. These simulations will be generated at $z = 2$, and k corresponds to the largest scale achievable in the simulations.

$f_{\rm NL}$	$k\,(h/{\rm Mpc})$	$L\,({\rm Mpc}/h)$
± 1	0.003	2000
± 3	0.005	1256
4.2	0.006	1100
-6	0.007	900

In Figure 4.1 we show the PNG corrections to the linear bias of a dark matter halo of mass $M = 10^{15}\,[M_\odot/h]$, assuming $p = 1$, for the threshold $f_{\rm NL} = \pm 1$ at different redshifts. As we can see, the signal increases with redshift and because the halo bias increase with mass, we will opt to analyse simulations at higher redshift, focusing on more massive halos.

4.3 Mock catalogues

We are adapting the code `ExSHalos`, originally introduced in Voivodic et al. (2019), to generate our mocks with a certain level of non-Gaussianity of the local type. This code is built upon the Excursion-Set idea to generate fast halo catalogues for a given cosmology.

In this version, we redefine the gravitational potential as given by Equation (4.12). The input parameter is the PNG amplitude at recombination, that we assume to be the primordial one, which is then translated to the final redshift $\bar{z}$ of the mock, where the gravitational potential

is computed to generate the density field at $\bar{z}$. In a nutshell, the algorithm is:

$$\delta_{\rm G}(\boldsymbol{k}) \rightarrow \phi_{\rm G}(\boldsymbol{x}) \rightarrow \Phi_{\rm NG}(\boldsymbol{k}) \rightarrow \delta_{\rm NG}(\boldsymbol{x}). \tag{4.21}$$

To avoid the necessity of huge boxes, therefore saving computational time, the catalogues are generated at a redshift of $z = 2$: it is not only reasonable for future surveys, but also allows us to probe $f_{\rm NL} = \pm 1$ around $k = 0.003$ h/Mpc with estimators whose precision is better than 5% (see right panel of Figure 4.1).

We will explore amplitudes allowed by the current CMB constraints: $f_{\rm NL} = \pm 1$, $f_{\rm NL} = \pm 3$, $f_{\rm NL} = -6$ and $f_{\rm NL} = 4.2$, with the last two being the Akrami et al. (2020) lower and upper bounds, respectively. For that, we will make use of halo catalogues generated at redshift $z = 2$ for boxes of sides shown in Table 4.1. For different $f_{\rm NL}$ values, we will consider different box sizes to minimise the computational time required to achieve a reasonable mass resolution. Values of k in Table 4.1 are for the largest scale achievable with the corresponding box sizes, where we expect to have a 5% difference in amplitude, as shown in Figure 4.1.

Note on the amplitude From measurements of the CMB power spectrum, we know that $\Phi \sim 10^{-5}$ at decoupling. Hence, the Gaussian piece of

$$\Phi \sim \Phi_G + f_{\rm NL}\Phi_G^2, \tag{4.22}$$

which is the leading term in the expansion, must have the same amplitude. Therefore, in order to observe significant effects in the primordial perturbation, $f_{\rm NL}$ must be larger than what is allowed by current constraints. So any non-Gaussian correction represents sub-percent corrections to the Gaussian assumption for the early fields.

Chapter 5

Relativistic Analysis of the Large-Scale Structure

This chapter aims at situating the unfamiliar reader with the general relativistic analysis of cosmological observables. The differences between the fully relativistic treatment of observables from the usual approach for theoretical quantities, such as the matter density contrast, give rise to the so-called relativistic effects. Such effects are important for upcoming surveys which will probe very large scales; if not properly taken into account, they can be erroneously taken as a primordial non-Gaussian detection. But they also provide, by themselves, an interesting route to learn about gravity and how it operates on cosmological scales. Therefore, the reader will find in this chapter the basics of this theory, focusing on the power spectrum and its multipoles.

5.1 Theoretical background and rationale

As discussed in Chapter 2, galaxy redshift surveys start from the redshift of objects, such as galaxies, to give us a 3-dimensional map of the Universe. If there were no perturbations to the Hubble flow, this map would be perfect in the sense that they would give us the correct radial positions for the surveyed objects; if, on top of that, there were no gravitational lenses, the angular positions provided by them would also be a faithful image of the actual position of the galaxies.

However, deviations from a perfectly Friedmann-Lamaître-Robertson-Walker (FLRW) metric, which describes a homogeneous and isotropic universe, exist in the real Universe: we know that dark and baryonic matter clumps together to form the structures we observe, such as dark matter halos and galaxies, leading to a non-homogeneous distribution of matter (the scales R at which homogeneity may be achieved is currently unknown, but within the ΛCDM model it is expected for inhomogeneities to be relevant at $R \lesssim 100$ Mpc$/h$). The presence of these structures is also responsible for gravitational lensing that changes not only the position, but the

shape and luminosity of galaxies (Bonvin et al., 2006; Bonvin and Durrer, 2011).

Finally, we also know that galaxies are falling into the potential of the cluster they reside on, and this effect introduces deviations from a perfect Hubble flow in their motion (i.e. motions that are not solely due to the cosmic expansion): gravitational redshift gives rise to an additional perturbation to the "real" position, and many other effects complicate the mapping between the positions of objects in a perfectly homogeneous and isotropic universe with what is observed in reality.

Forthcoming surveys will cover a vast volume of the observable universe, allowing us to observe increasingly larger scales, including those that are comparable with the Hubble radius. On such scales, the Newtonian treatment is no longer accurate, as the effects of causality imposed by the limit of the speed of light are not taken into account in the Newtonian theory, which ignores the existence of a cosmological horizon. In addition, the inclusion of relativistic species, such as neutrinos, can only be properly modelled with the relativistic treatment[1]. Therefore, in order to extract the maximum amount of information from these surveys, our theory must be accurate enough on the large (and also small) scales.

Differences between the purely relativistic and the Newtonian treatment obtained in theoretical expectations for observable quantities give rise to so-called relativistic effects, which, if not properly modelled, can bias the detection of primordial non-Gaussianities (Yoo, 2010; Umeh et al., 2017; Wang et al., 2020).

We will always keep in mind what is actually observed by galaxy surveys: spectroscopic and photometric surveys give us a catalogue with the observed positions, in the redshift space, $\boldsymbol{s} = (z_{\rm obs}, \theta, \varphi)$, with the redshift estimated via the spectroscopic (then one assumes zero error in the redshift measure) or photometric technique (and, in this case, it is necessary to account in the measurements for the distribution describing the probability function of each redshift, which leads to a non-negligible error); but, in both cases, we will always have at hands a catalogue with the three-dimensional position $\boldsymbol{s}$ of objects.

To relate the observed position of galaxies to the one in an unperturbed (FLRW) universe, we use the covariant definition of redshift:

$$z = \frac{(g_{\mu\nu}k^\mu u^\nu)_{\rm s}}{(g_{\mu\nu}k^\mu u^\nu)_{\rm o}} - 1. \tag{5.1}$$

We label quantities computed at the source of emission of photons by s and the ones computed at the observer by o. This covariance definition of redshift allows us to account for all pos-

[1]There are many alternatives in the literature to take these effects into account, especially with the treatment of relativistic species in Newtonian simulations through a "dictionary" which connects the two approaches (see, for example, Fidler et al., 2017; Adamek and Fidler, 2019). Another approach is to perform simulations whose equations of motion are inherently relativistic (Adamek et al., 2016a).

sible effects such as volume, redshift and angle variations due to all sorts of general relativistic effects; hence, it incorporates additional corrections to the standard Newtonian redshift distortions, where only the peculiar velocity of the sources is responsible for the corrections along the line-of-sight (LOS) $\hat{\boldsymbol{n}}$ (as derived by Kaiser, 1984): $\lambda_{\rm o}^{\rm pec} \approx \lambda_s \left(1 + \frac{v_\parallel}{c}\right)$.

It is a common assumption that $v_\parallel \ll c$, where $v_\parallel = \boldsymbol{v} \cdot \hat{\boldsymbol{n}}$ is the peculiar velocity projection along the LOS. Furthermore, the Hubble flow is responsible for a redshift in the observed wavelength of the photons $\lambda_{\rm o}$ of the form $\lambda_{\rm o} = \lambda_{\rm o}^{\rm pec} \frac{a(t_r)}{a(t_e)}$, with t_r and t_e the time of reception and emission, respectively. Hence, peculiar velocities shift the observed redshift $z_{\rm obs}$ with respect to the one that comes solely from the cosmic expansion $z_{\rm H}$[2]:

$$\frac{\lambda_{\rm o}}{\lambda_s} \equiv 1 + z_{\rm obs} = \left(1 + \frac{v_\parallel}{c}\right)(1 + z_{\rm H}) = (1 + z_{\rm H}) + (1 + z_{\rm H})\frac{v_\parallel}{c}. \tag{5.2}$$

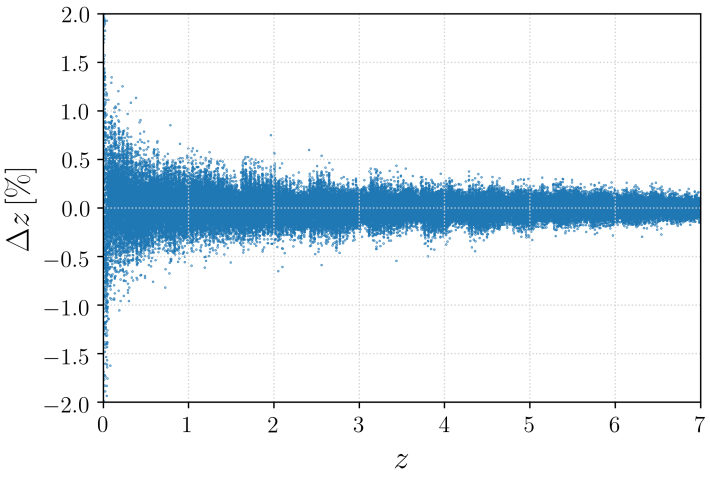


Figure 5.1: Relative difference between the redshift that dark matter halos of a N-body simulation would have if it was only sourced by the cosmic expansion, and the theoretically observed redshift they have in that simulation. This includes the perturbations coming from the peculiar velocities of the halos and the gravitational redshift, among other effects. These dark matter halos are described in details in Section 6.2, and the effects beyond the Hubble flow in Chapter 5.

For illustrative purposes, and to get a sense of the magnitude of the redshift distortions, in Figure 5.1 we show the difference between the "truly" observed redshifts $z_{\rm obs}$ (i.e., in redshift space) of dark matter halos obtained from a N-body simulation and the redshifts $z_{\rm H}$ measured for the halos in real space, that is, in the absence of redshift distortions.

Computing the overdensity of dark matter tracers with (5.2), we obtain the usual Kaiser

[2]There might be additional terms in the velocity induced by the observer's proper motion. However, by assuming negligible intrinsic dipole, this motion is known through the CMB dipole such that its contribution is easily included in the calculations and can be further subtracted (Matsubara, 2000).

(1984) expression:

$$\delta_s(\boldsymbol{s}) \approx \delta_r(\boldsymbol{r}) - \frac{1}{\mathcal{H}}\partial_r(\boldsymbol{v}\cdot\hat{\boldsymbol{n}}) - \frac{\boldsymbol{v}\cdot\hat{\boldsymbol{n}}}{\mathcal{H}r}\left(2 + \frac{\partial \ln \bar{\phi}}{\partial \ln r}\right), \tag{5.3}$$

where only first order quantities in perturbation theory are considered, and where $\frac{\partial \ln \bar{\phi}}{\partial \ln r}$ accounts for the survey selection function ϕ.

Within linear theory, we can relate quantities in configuration and Fourier space as (for example, see Amendola and Tsujikawa, 2010, Section 4.8)

$$\begin{aligned}
\delta^{(r)}(\boldsymbol{r}) &= b\int \frac{\mathrm{d}^3k}{(2\pi)^3}\,\delta^{(r)}(\boldsymbol{k})\,\mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{r}},\\
v_\parallel = \boldsymbol{v}\cdot\hat{\boldsymbol{r}} &= if\mathcal{H}\int \frac{\mathrm{d}^3k}{(2\pi)^3}\,\delta^{(r)}(\boldsymbol{k})\,\frac{\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{r}}}{k}\,\mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{r}},\\
\mathcal{H}^{-1}\partial_r v_\parallel &= -f\int \frac{\mathrm{d}^3k}{(2\pi)^3}\,\delta^{(r)}(\boldsymbol{k})\,(\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{r}})^2\,\mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{r}},\\
\alpha v_\parallel &= if\mathcal{H}\int \frac{\mathrm{d}^3k}{(2\pi)^3}\,\delta^{(r)}(\boldsymbol{k})\,\frac{\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{r}}}{k}\,\alpha\,\mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{r}},
\end{aligned} \tag{5.4}$$

where $\alpha \equiv \partial_{\ln r}\ln(r^2\phi_\beta)$, for a dark matter tracer β with linear bias b and selection function ϕ. In this set of equations, f is the growth rate, parametrised by $f(z) \sim \Omega_m(z)^\gamma$, with γ being the growth index, and $\delta^{(r)}(\boldsymbol{k})$ is the matter density contrast in real space (with no redshift distortions, only the cosmic expansion). With Equations (5.3) and (5.4), we already find the main feature we shall analyse in Chapter 6: the presence of imaginary terms due to the coupling of $v_\parallel$ with terms of the form $(\mathcal{H}r)^{-1}$. As it can be seen from the Newtonian density contrast,

$$\delta^{(s)}(\boldsymbol{k}) = \delta^{(r)}(\boldsymbol{k})\left[b + f(\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{r}})^2 - i\frac{f}{kr}\alpha(\boldsymbol{r})(\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{r}})\right], \tag{5.5}$$

the imaginary term increases for closer objects (small r) and larger scales (small k).

If we consider, instead, the covariant definition of redshift, Equation (5.1), additional contributions appear due to the inhomogeneities that exist between the observer and the source, which causes deviations in the photons path, altering not only the observed angular position, but also magnifying galaxies due to gravitational lensing effect.

The most immediate relativistic correction is the gravitational redshift (blueshift) of the photons climbing up (falling down) the gravitational potential Φ as a result of the loss (gain) of energy. The resulting effect is a global redshift which, when included through the term $-\Delta\Phi/c$, contributes to the observed redshift from the source as

$$c\frac{\Delta\lambda}{\lambda} = aH\Delta x_\parallel + \Delta v_\parallel - \frac{\Delta\Phi}{c}. \tag{5.6}$$

By employing the same steps used to get Equation (5.5), with the potential term in Fourier space given by

$$\Phi(\boldsymbol{k}) = -\frac{3}{2}\Omega_m \left(\frac{aH}{kc}\right)^2 \delta(\boldsymbol{k}), \tag{5.7}$$

we arrive at

$$\delta^{(s)}(\boldsymbol{k}) = \delta^{(r)}(\boldsymbol{k})\left[b + f(\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{r}})^2 + i\left(\frac{3\Omega_m}{2}\frac{\mathcal{H}}{kc} - \frac{f}{k}\frac{\alpha(\boldsymbol{r})}{r}\right)(\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{r}})\right]. \tag{5.8}$$

Neglecting variations of the selection function and assuming a single LOS for the whole survey, $\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{r}} = \mu$, the auto-spectrum of a dark matter tracer α is given by

$$\langle\delta^s_\alpha(\boldsymbol{k})\delta^s_\alpha(\boldsymbol{k}')\rangle = (2\pi)^3\delta^D(\boldsymbol{k}+\boldsymbol{k}')\left[(b_\alpha + f\mu^2)^2 + \left(\frac{3}{2}\Omega_m\frac{aH}{kc}\mu\right)^2\right]P(k). \tag{5.9}$$

The second term inside the brackets is the relativistic correction to Kaiser's expression (first term within the brackets) due to the gravitational redshift suffered by the photons. To get an idea of this contribution to the spectrum, for an Euclid-like survey the amplitude is in the order of $\sim 10^{-2}$, while for a VRO/LSST-like survey we expect $\sim 10^{-3}$ (see specifications in Table 5.1). Compared to Kaiser's term, which is of order 1 for both cases, we see that this contribution is minor.

However, the cross-spectrum between two different tracers α and β,

$$\langle\delta^s_\alpha(\boldsymbol{k})\delta^s_\beta(\boldsymbol{k}')\rangle = (2\pi)^3\delta^D(\boldsymbol{k}+\boldsymbol{k}')\left[(b_\alpha + f\mu^2)(b_\beta + f\mu^2) + \left(\frac{3}{2}\Omega_m\frac{\mu}{k_c}\right)^2 - i\frac{3}{2}\Omega_m\frac{\mu}{k_c}(b_\alpha - b_\beta)\right]P(k), \tag{5.10}$$

acquires an intrinsic imaginary term which gives origin to a dipole, as we shall discuss further. This is a more promising observable, as it scales like $1/k_c$, where $k_c = ck/aH$ (McDonald, 2009).

Taking into account all corrections proportional to $\boldsymbol{v}\cdot\hat{\boldsymbol{n}}$ and considering theories where Euler's equation $\dot{\boldsymbol{v}} + \mathcal{H}\boldsymbol{v} + \nabla\Phi$ is valid, we get the Doppler term[3], which goes in addition to the standard redshift space distortions (RSD) $\mathcal{H}^{-1}\partial_r(\boldsymbol{v}\cdot\boldsymbol{n})$:

$$\delta^{(s)}_\alpha(\boldsymbol{r}) \equiv b_\alpha\delta^{(r)}(\boldsymbol{r}) - \frac{1}{\mathcal{H}}\partial_r(\boldsymbol{v}\cdot\boldsymbol{n}) + A_\alpha\boldsymbol{v}\cdot\boldsymbol{n}, \tag{5.11}$$

[3]The complete and general expression, without assuming the validity of Euler's equation, has the following additional terms: gravitational redshift $\mathcal{H}^{-1}\partial_r\Phi$ at order $\mathcal{H}k^{-1}$); light-cone effect $\boldsymbol{v}\cdot\boldsymbol{n}$ (*light-cone effect*) is associated with the fact that, because we observe objects in the past light cone, the observed positions are different from those positions that galaxies would have on a constant time hypersurface; acceleration term $\dot{\boldsymbol{v}}\cdot\boldsymbol{n}$ describes the acceleration of the source with respect to the observer; we also have non-local terms that express the distortions due to the gravitational lens in the observed volume, but we will not deal with integral terms at this stage (for example see Breton et al., 2018; Tansella et al., 2018; Bonvin, 2014; Bonvin et al., 2014).

where

$$A_\alpha \equiv \frac{5s_\alpha - 2}{\mathcal{H}r} + \frac{\partial(a^3\bar{n}_\alpha)}{\partial \ln a} - \frac{\mathcal{H}'}{\mathcal{H}^2} - 5s_\alpha. \tag{5.12}$$

Table 5.1: Specifications for the Euclid and LSST surveys, with which we derive the expected amplitude of the gravitational redshift signal. The values were extracted from Ballardini et al. (2019).

	Area	z range	$k_{\rm min}$
Euclid	15000 deg^2	$0 < z < 2.5$	0.001 Mpc$/h$
LSST	13800 deg^2	$0.2 < z < 1.2$	0.002 Mpc$/h$

Notice that $-5s_\alpha(\mathcal{H}r)^{-1}$ is already accounted by Kaiser (1984), with

$$s_\alpha \propto \partial_{\ln r} \ln(r^2\phi_\alpha) \tag{5.13}$$

being what we call the *magnification bias*. It depends on the survey flux threshold and is associated with the fact that, due to lensing, galaxies that would not have made the magnitude cut are included in the sample as they are magnified, with the opposite reasoning also applied to demagnified galaxies that are excluded from the sample because of lensing as well. Finally,

$$\frac{\partial(a^3\bar{n}_\alpha)}{\partial \ln a} = -(1+z)\frac{\partial \ln \bar{N}_\alpha}{\partial z} \equiv b_e \tag{5.14}$$

is the so-called *evolution bias*, where $\bar{N}_\alpha = a^3\bar{n}_\alpha$ is the *physical* number density. This parameter quantifies the intrinsic evolution of objects throughout the cosmic time. For example, galaxies and halos merge to form more massive objects, which implies in a varying number density $\bar{n}$. This will be shown in details in Section 6.2.2.3.

With the exception of the true density perturbation δ_α, all other terms in Equation (5.11) are due to distortions in the coordinate system by relating quantities in a homogeneous and isotropic universe with those that are really observed. Therefore, we arrive at the main theoretical equation of this part of the thesis:

$$\delta_\alpha^{(s)}(\boldsymbol{k}) = \delta^{(r)}(\boldsymbol{k})\left[b_\alpha + f\mu_{\boldsymbol{k}}^2 + if(\mathcal{H}k^{-1})A_\alpha\mu_{\boldsymbol{k}}\right], \tag{5.15}$$

with $\mu_{\boldsymbol{k}} \equiv (\hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{r}})$ to keep the explicit dependence on the LOS.

5.2 Multipoles of the power spectrum

Now we investigate the consequences of Equation (5.15).

5.2.1 Flat-sky approximation

When the line of sight is the same for all objects, i.e. $\hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{r}} = \mu$ is a constant (flat-sky approximation), the cross-spectrum $P^{(s)}_{\alpha\beta}(\boldsymbol{k}) = \langle \delta_\alpha(\boldsymbol{k}) \delta^*_\beta(\boldsymbol{k}) \rangle$ of two tracers α and β becomes:

$$P^{(s)}_{\alpha\beta}(\boldsymbol{k}) = P^{(r)}(k) \Bigg[(b_\alpha + f\mu^2)(b_\beta + f\mu^2) + A_\alpha A_\beta f^2 \mu^2 \frac{\mathcal{H}^2}{k^2} + if\mu \left[(b_\beta + f\mu^2) A_\alpha - (b_\alpha + f\mu^2) A_\beta \right] \frac{\mathcal{H}}{k} \Bigg]. \tag{5.16}$$

As we can immediately see, isotropy is broken by the choice of a LOS and $P^{(s)}_{\alpha\beta}(\boldsymbol{k}) = P^{(s)}(k, \mu)$ can be expanded in a series of Legendre polynomials:

$$P^{(s)}(k, \mu) = \sum_{\ell=0}^{\infty} P^{(s)}_\ell(k) \mathcal{L}_\ell(\mu), \tag{5.17}$$

where the multipoles are

$$P^{(s)}_\ell(k) = \frac{2\ell + 1}{2} \int_{-1}^{1} P^{(s)}(k, \mu) \mathcal{L}_\ell(\mu) \, d\mu \tag{5.18}$$

$$= P^{(r)}(k) \frac{2\ell + 1}{2} \int_{-1}^{1} \Bigg[(b_\alpha + f\mu^2)(b_\beta + f\mu^2) + A_\alpha A_\beta f^2 \mu^2 \frac{\mathcal{H}^2}{k^2} + if\mu \left[(b_\beta + f\mu^2) A_\alpha - (b_\alpha + f\mu^2) A_\beta \right] \frac{\mathcal{H}}{k} \Bigg] \mathcal{L}_\ell(\mu) \, d\mu \tag{5.19}$$

$$\equiv P^{(r)}(k) \, c_\ell. \tag{5.20}$$

Neglecting quadratic terms $O(\mathcal{H}/k)^2$, the expansion coefficients are given by:

$$c_0(f, b) = b_\alpha b_\beta + \frac{1}{3} f (b_\alpha + b_\beta) + \frac{1}{5} f^2 \tag{5.21}$$

$$c_1(k, f, b, A) = \frac{1}{5} if \frac{\mathcal{H}}{k} \left[A_\alpha (3f + 5b_\beta) - A_\beta (3f + 5b_\alpha) \right] \tag{5.22}$$

$$c_2(f, b) = \frac{2}{3} f (b_\alpha + b_\beta) + \frac{4}{7} f^2 \tag{5.23}$$

$$c_3(k, f, A) = \frac{2}{5} if^2 \frac{\mathcal{H}}{k} (A_\alpha - A_\beta) \tag{5.24}$$

$$c_4(f) = \frac{8}{35} f^2 \tag{5.25}$$

In this case, the monopole (5.21), quadrupole (5.23) and hexadecapole (5.25) are the same as in the Newtonian case, whereas a dipole scaling as $\mathcal{H}k^{-1}$ emerges in the cross-correlation,

$$P^{\alpha\beta}_1(k) = iP^{(r)}(k) \frac{f}{5} \left[A_\alpha (3f + 5b_\beta) - A_\beta (3f + 5b_\alpha) \right] \frac{\mathcal{H}}{k}, \tag{5.26}$$

vanishing for the auto-spectrum just as in the case of the gravitational redshift distortions.

Nonetheless, if we consider corrections of second-order $O(\mathcal{H}/k)^2$, the monopole and quadrupole acquire a k^{-2} dependence which is responsible for the constant concern of a possible contamination with the scale-dependent bias that comes from primordial non-Gaussianities (PNG), parametrised by $\sim f_{\rm NL}k^{-2}$ as discussed in Section 4.2.2. Wang et al. (2020) showed how relativistic effects contaminate the primordial non-Gaussian signal at low redshift for the monopole, but especially for the quadrupole at higher redshifts, as shown in Figure 5.2.

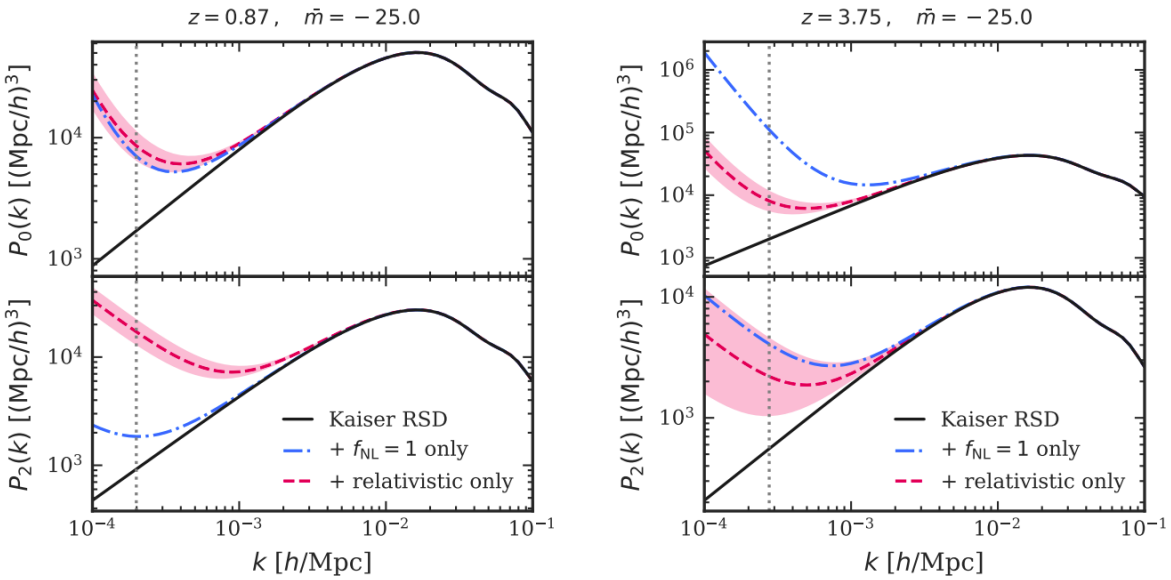


Figure 5.2: Power spectrum monopole $P_0(k)$ (*top panels*) and quadrupole $P_2(k)$ (*lower panels*) at two redshifts $z = 0.87$ (*left*) and $z = 3.75$ (*right*) for a magnitude threshold $\bar{m} = -25$. Multipoles modelled with only the Kaiser term are shown in black solid lines, the effect of local type of primordial non-Gaussianities with a fiducial $f_{\rm NL} = 1$ are shown in dash-dotted blue lines, and the relativistic effects up to second order are shown by the dashed red lines. The shaded red regions shows the impact of a varying luminosity function for the eBOSS quasars within the 95% credible interval, which directly impacts the evolution (5.14) and magnification biases (5.13). The vertical dotted lines mark the horizon scale $k = \mathcal{H}$. [Extracted from Wang et al. (2020).]

In redshift space, we can summarise the power-spectrum multipoles as

$$P_0^{(s)}(k) = P_0^{\rm N}(k) + \Delta P_0^{\rm RE}(k) + \Delta P_0^{\rm PNG}(k), \tag{5.27}$$

$$\mathrm{Im}[P_1^{(s)}(k)] = \mathrm{Im}[P_1^{\rm RE}(k)] + \mathrm{Im}[\Delta P_1^{\rm PNG}(k)], \tag{5.28}$$

$$P_2^{(s)}(k) = P_2^{\rm N}(k) + \Delta P_2^{\rm RE}(k) + \Delta P_2^{\rm PNG}(k), \tag{5.29}$$

$$\mathrm{Im}[P_3^{(s)}(k)] = P_3^{\rm RE}(k), \tag{5.30}$$

$$P_4^{(s)}(k) = P_4^{\rm N}(k), \tag{5.31}$$

where the indices N, RE and PNG refer to the Newtonian (Kaiser) multipoles and the corrections coming from relativistic effects and the primordial non-Gaussian correction to the bias, which

are given by (see, for example, Wang et al., 2020, for the even multipoles):

$$\Delta P_0^{\rm RE}(k) = \left[\frac{1}{3}A_\alpha A_\beta f^2 \left(\frac{\mathcal{H}}{k}\right)^2\right] P_m^{(r)}(k), \tag{5.32}$$

$$\Delta P_0^{\rm PNG}(k) = \left[\Delta b_\alpha \left(\frac{f}{3} + b_\beta\right) + \Delta b_\beta \left(\frac{f}{3} + b_\alpha\right) + \Delta b_\alpha \Delta b_\beta\right] P_m^{(r)}(k), \tag{5.33}$$

$$\mathrm{Im}[\Delta P_1^{\rm PNG}(k)] = f\frac{\mathcal{H}}{k}(A_\alpha \Delta b_\beta - A_\beta \Delta b_\alpha) P_m^{(r)}(k), \tag{5.34}$$

$$\Delta P_2^{\rm RE}(k) = \left[\frac{2}{3}f^2 \left(\frac{\mathcal{H}}{k}\right)^2 A_\alpha A_\beta\right] P_m^{(r)}(k), \tag{5.35}$$

$$\Delta P_2^{\rm PNG}(k) = \frac{2}{3} f(\Delta b_\alpha + \Delta b_\beta). \tag{5.36}$$

As we can see, if there are corrections due to PNG, it will impose a correction to the relativistic dipole, but scaling as k^{-3}, which makes its detection unpractical. Likewise, deviations are small for both relativistic and PNG corrections in the even multipoles, becoming only relevant at ultra-large scales. Finally, we can see how both effects are entangled, meaning that the detection of one depends on the knowledge of the other one.

5.2.2 Wide-angle corrections

If we loosen the flat-sky approximation, the cross-spectrum must be carefully handled:

$$\begin{aligned} P_{\alpha\beta}^{(s)}(\boldsymbol{k}_1, \boldsymbol{k}_2) = P^{(r)}(k_1) \Bigg[& (b_\alpha + f\mu_{\boldsymbol{k}_1}^2)(b_\beta + f\mu_{\boldsymbol{k}_2}^2) + A_\alpha A_\beta f^2 \mu_{\boldsymbol{k}_1}\mu_{\boldsymbol{k}_2}\frac{\mathcal{H}^2}{k^2} + \\ & + if\left[(b_\beta + f\mu_{\boldsymbol{k}_2}^2)A_\alpha \mu_{\boldsymbol{k}_1} - (b_\alpha + f\mu_{\boldsymbol{k}_1}^2)A_\beta \mu_{\boldsymbol{k}_2}\right]\frac{\mathcal{H}}{k}\Bigg]. \end{aligned} \tag{5.37}$$

The first term $P^{(r)}(k)$ comes from the fact that we assume a homogeneous and isotropic matter power spectrum in real space, and the different LOS are included in the term $\mu_{\boldsymbol{k}_i} = \hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{r}}_i$.

At the moment we will not deal with wide-angles, but they will be of utmost importance for further analysis on larger simulations

5.3 Summary

Our main goal regarding relativistic effects will be to measure the dipole induced by the imaginary part of the cross-spectrum; namely:

$$\mathrm{Im}[P_1^{\alpha\beta}(k)] = P^{(r)}(k)\frac{f}{5}\left[A_\alpha(3f + 5b_\beta) - A_\beta(3f + 5b_\alpha)\right]\frac{\mathcal{H}}{k}. \tag{5.38}$$

The analysis held in the scope of relativistic a N-body simulation, covering a pencil-beam region of sky, is given in Chapter 6.

Future analysis on full-sky simulations will demand a proper treatment of wide-angle effects. For short, they will have two main effects:

- The power spectrum $P^{(s)}_{\alpha\beta}(\boldsymbol{k}_1, \boldsymbol{k}_2)$ will no longer be diagonal. Homogeneity and isotropy can be recovered for the flat-sky limit ($\mu_{\boldsymbol{k}_1} = \mu_{\boldsymbol{k}_2} = \mu$);

- The imaginary part of $P^{(s)}_{\alpha\beta}(\boldsymbol{k}_1, \boldsymbol{k}_2)$,

$$\mathrm{Im}[P^{(s)}_{\alpha\beta}(\boldsymbol{k}_1, \boldsymbol{k}_2)] = f\left[(b_\beta + f\mu^2_{\boldsymbol{k}_2})A_\alpha\mu_{\boldsymbol{k}_1} - (b_\alpha + f\mu^2_{\boldsymbol{k}_1})A_\beta\mu_{\boldsymbol{k}_2}\right]\frac{\mathcal{H}}{k} \tag{5.39}$$

  does not vanish even for the autocorrelation, when $\alpha = \beta$. In this particular case:

$$\mathrm{Im}[P^{(s)}_{\alpha\alpha}(\boldsymbol{k}_1, \boldsymbol{k}_2)] = A_\alpha b_\alpha f\left[(1 + \beta\mu^2_{\boldsymbol{k}_2})\mu_{\boldsymbol{k}_1} - (1 + \beta\mu^2_{\boldsymbol{k}_1})\mu_{\boldsymbol{k}_2}\right]\frac{\mathcal{H}}{k}, \tag{5.40}$$

  where $\beta = f/b_\alpha$. Therefore, an intrinsically geometric dipole will emerge, needing thus a proper modelling for a robust extraction of the relativistic dipole.

Finally, a joint analysis with primordial non-Gaussianities and relativistic effects is aimed in a future work, as both of these effects play a significant role at the largest scales available in upcoming galaxy surveys.

Chapter 6

Observing Relativistic Features in the Large-Scale Structure

In this chapter we address the issue of observing in realistic N-body simulations relativistic effect. By realistic we mean a relativistic simulation based on the weak field approximation, thus incorporating relevant features such as a cosmological horizon, frame dragging and gravitational waves, which has, among its possible outputs, past light cones constructed "on the fly", i.e. the dark matter particles are incorporated in the past light cone directly, with no discreteness effect whatsoever. Fourier analysis in this case is more complicated than the standard power spectrum and bispectrum estimation of a N-body box due to several reasons which will be explained in what follows, and thus is closely related to what is done in actual galaxy surveys. This chapter is built upon the original work of Guandalin et al. (2020), with some parts being the collaborators' contributions.

6.1 Relativistic effects in the power spectrum

Contrary to the simplistic view of N-body simulations, which give us the 3-dimensional positions of objects at a fixed time slice, the true observed quantity in a galaxy survey is the number of dark matter tracers (e.g. galaxies or halos) $N(z, \hat{\boldsymbol{n}})$ in a pixel given by a solid angle $\mathrm{d}\Omega$ around a direction $\hat{\boldsymbol{n}} = (\theta, \varphi)$, defined with respect to the observer's line of sight (LOS), and at a redshift bin $[z, z + \mathrm{d}z]$ (Bonvin and Durrer, 2011; Bonvin, 2014). The number overdensity of some tracer α can thus be defined as

$$\delta_\alpha^{(s)}(\boldsymbol{s}) \equiv \frac{N_\alpha(z, \hat{\boldsymbol{n}}) - \bar{N}_\alpha(z)}{\bar{N}_\alpha(z)} = \frac{n_\alpha(z, \hat{\boldsymbol{n}}) - \bar{n}_\alpha(z)}{\bar{n}(z)} + \frac{\delta V(z, \hat{\boldsymbol{n}})}{\bar{V}(z)}, \tag{6.1}$$

where the equality is obtained by relating the number counts with the number density as $n(z, \hat{\boldsymbol{n}}) \equiv N(z, \hat{\boldsymbol{n}})/V(z, \hat{\boldsymbol{n}})$. In the above equation, $\bar{N}_\alpha(z)$ is the selection function of the tracer α, obtained

by angular averaging over the tracer number count.

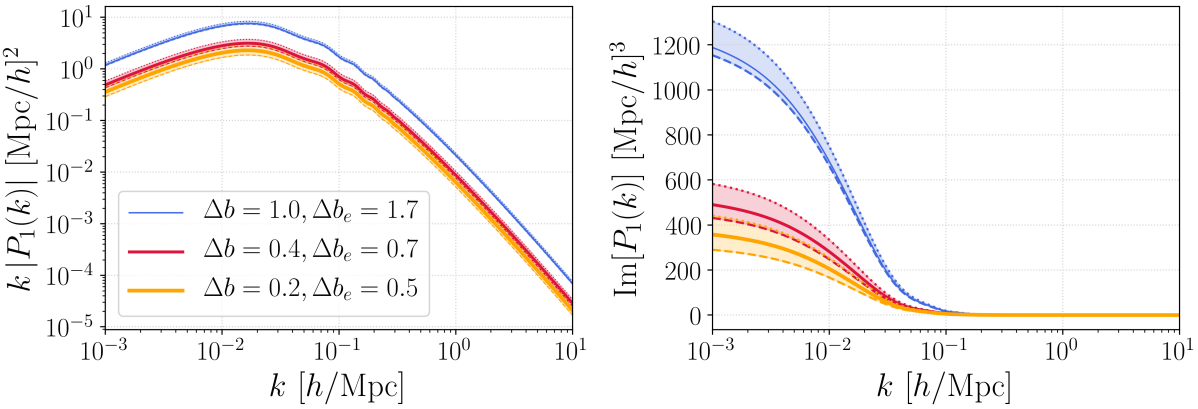


Figure 6.1: Theoretical prediction for the cross power spectrum dipole of different tracers at redshift $z = 1.9$, with the difference in linear and evolution bias shown in the legend. Solid lines represent the case where there is no magnification bias $s_\alpha = 0$, whereas shaded regions represent the effect of different magnification biases among the tracers. Dotted lines shows the limiting case where s_α is smaller than s_β by 40%, whilst the dashed ones shows the opposite case, with s_α larger than s_β by a factor of 40%.

As discussed in Chapter 5, the quantities defined in Equation (6.1) are in redshift space, meaning that they are characterised by the observed (comoving) coordinates $\boldsymbol{s} = (s, \theta, \varphi)$, with the radial comoving coordinate s being connected to the observed redshift by some cosmological model[1]. The standard treatment (Kaiser, 1984), relating the number of sources in a perfect FLRW universe with the truly observed density field via the conservation of number counts, gives rise to the so-called *redshift-space distortions*. This allows us to relate the theoretical predictions in a homogeneous universe with the observed quantities with the addition of departures from the perfect FLRW metric.

Even though the corrections first derived in Kaiser (1984) describe satisfactorily observations limited to sub-horizon scales, where the Newtonian treatment is well suited, this is not a truly observed quantity, as it is gauge-dependent. Furthermore, as we saw in Chapter 5, future galaxy surveys and cosmological observations that rely on the largest (near-horizon) scales demand a proper treatment of the LSS clustering. At smaller scales, the improved sensitivity will also hold the potential for a detection of subleading corrections (for an example see Saga et al., 2020).

In this Chapter, we explore the detection of the induced relativistic dipole in the cross-

[1]The radial comoving coordinate in redshift space s, obtained from the *observed* redshift, should not be confused with the magnification bias s_α of some tracer α, which will carry a Greek index throughout this work. We also draw the reader's attention to the radial comoving coordinate denoted by r in real space, obtained from the unperturbed (Hubble flow) redshift of a perfect FLRW universe.

spectrum of cold dark matter tracers, Equation (5.38):

$$P_1^{\alpha\beta}(k) = i\frac{f}{5}\frac{\mathcal{H}}{k}\Big[A_\alpha(3f + 5b_\beta) - A_\beta(3f + 5b_\alpha)\Big]P^{(r)}(k). \tag{6.2}$$

Because there is no Newtonian counterpart for this dipole, with the exception of wide-angle effects (that we do not account for in this work) and an induced dipole coming from the coupling of the window function, which we discuss further, they represent a smoking gun for relativistic effects in the galaxy clustering. We also call the reader's attention to the fact that this dipole term is antisymmetric, meaning that $\langle\delta_\alpha(\boldsymbol{k})\delta^*_\beta(\boldsymbol{k})\rangle = -\langle\delta_\beta(\boldsymbol{k})\delta^*_\alpha(\boldsymbol{k})\rangle$. In Figures 6.1 and 6.2 we illustrate the dipole term in both the Fourier and configuration spaces, respectively, for three linear and evolution bias differences (different colours) at a fixed redshift $z = 2$.

In what follows we explore the detection of Equation (5.38) in a relativistic simulation of a light cone, described in Section 6.2. Since we will be dealing with dark matter halos, the magnification bias s_α in the Doppler term vanishes simply because our samples are not limited in magnitude. Therefore, in addition to the linear bias of the halos, the remaining parameter entering the theoretical predictions is the evolution bias (Equation 5.14).

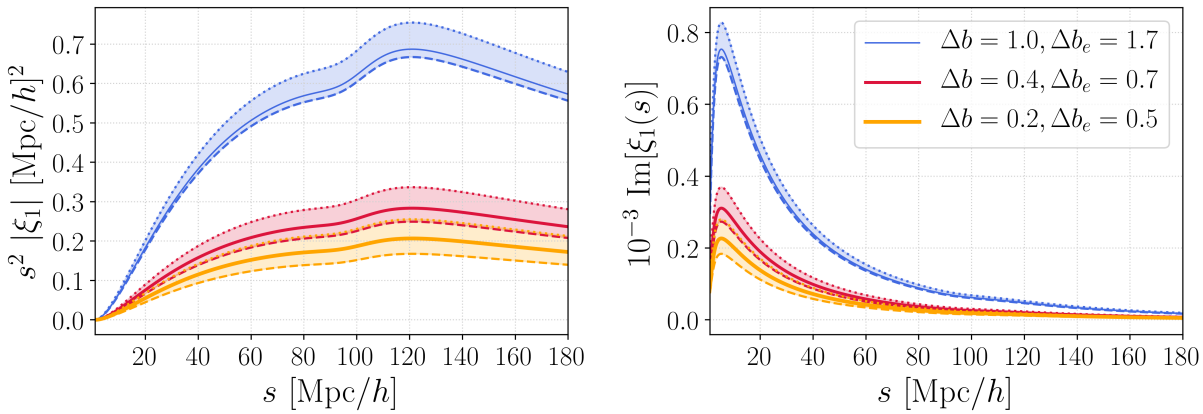


Figure 6.2: Same as Figure 6.1, but for the cross-correlation function dipole of different tracers at redshift $z = 1.9$. Differences in the linear and evolution bias are shown in the legend. Solid lines represent the case where there is no magnification bias $s_\alpha = 0$, whereas shaded regions represent the effect of different magnification biases among the tracers. Dotted lines shows the limiting case where s_α is smaller than s_β by 40%, whilst the dashed ones shows the opposite case, with s_α larger than s_β by a factor of 40%.

6.2 Simulation

In this work we make use of a large N-body simulation performed with the relativistic code *gevolution* (Adamek et al., 2016a,b). The simulation has a comoving volume of $(2.4\ \mathrm{Gpc}/h)^3$ with dark matter particles of mass $2.64 \times 10^9\ (M_\odot/h)$, and represents a typical ΛCDM cosmology: $h = 0.67556$, $\omega_b = 0.022032$, $\omega_{\mathrm{cdm}} = 0.12038$, $T_{\mathrm{CMB}} = 2.7255$ K, $A_s = 2.215 \times$

10^{-9}, $n_s = 0.9619$, $N_{\rm ur} = 3.046$, and $N_{\rm ncdm} = 0$. In order to avoid replications in the light cone, the pencil beam was carefully oriented in the periodic domain, with the line-of-sight direction in the polar axis given by the Cartesian vector $\boldsymbol{\ell} = (0.735, 0.485, 1)$. The initial conditions for the simulation were set at a redshift of $z = 127$.

Unlike the standard approach to building light cones (Merson et al., 2012; Smith et al., 2017; Breton et al., 2018), which consists of generating many simulation snapshots with a sufficient small redshift step between them to avoid time discretisation effects, the light cone output from *gevolution* records particle positions and velocities on the fly: during the simulation, the code identify particles that are within a proper comoving distance interval from a predefined observer that would cause them to be placed in the final catalogue. These particles are then shifted by a fractional time step and recorded on the null FLRW hypersurface given by the past light cone of the observer. Hence, there are no time discretisation artefacts and no need to generate an enormous amount of snapshots to build the light cone. In our case, no replications whatsoever were performed in order to cover the whole light cone volume, which has the advantage of removing any concerns about spurious correlations on large scales due to periodicity for example.

The *gevolution* code does not employ the adaptive mesh refinement (AMR) method and thus has a low accuracy at small scales. However, while AMR can improve the 1-halo term by better resolving halo substructures, it does not significantly impact the large scales dominated by the 2-halo term, which is the focus of this work. As will be pointed out in Section 6.2.2, all sub-halos are discarded in our analysis in any case.

6.2.1 The particle catalogue

The final cold dark matter catalogue is recorded in a Gadget format, which includes the position and velocities of more than 10^{11} particles with no special redshift distribution. In our analysis, we consider a catalogue restricted to 35×10^6 particles only, and we show the angular and radial (redshift) distribution of these particles in the histograms of Figure 6.3, where we select a specific redshift range ($2.1 < z < 2.5$). To put this particles in redshift space, we apply a ray tracing algorithm[2], which connects the particles (and further halos) with the real observed redshift and angle variables. In contrast to cases where Newtonian N-body simulations are ray traced (for example Borzyszkowski et al., 2017), in *gevolution* the metric perturbations and the source positions are both provided in the Poisson gauge. As opposed to simplistic ray-tracing algorithms that rely on the Born approximation, the version utilised in this work makes use of a

[2]The algorithm is described in Lepori et al. (2020). In Guandalin et al. (2020) we give a brief review. The reader is invited to look at these references.

shooting method to connect the light rays by integrating the geodesics backward in time.

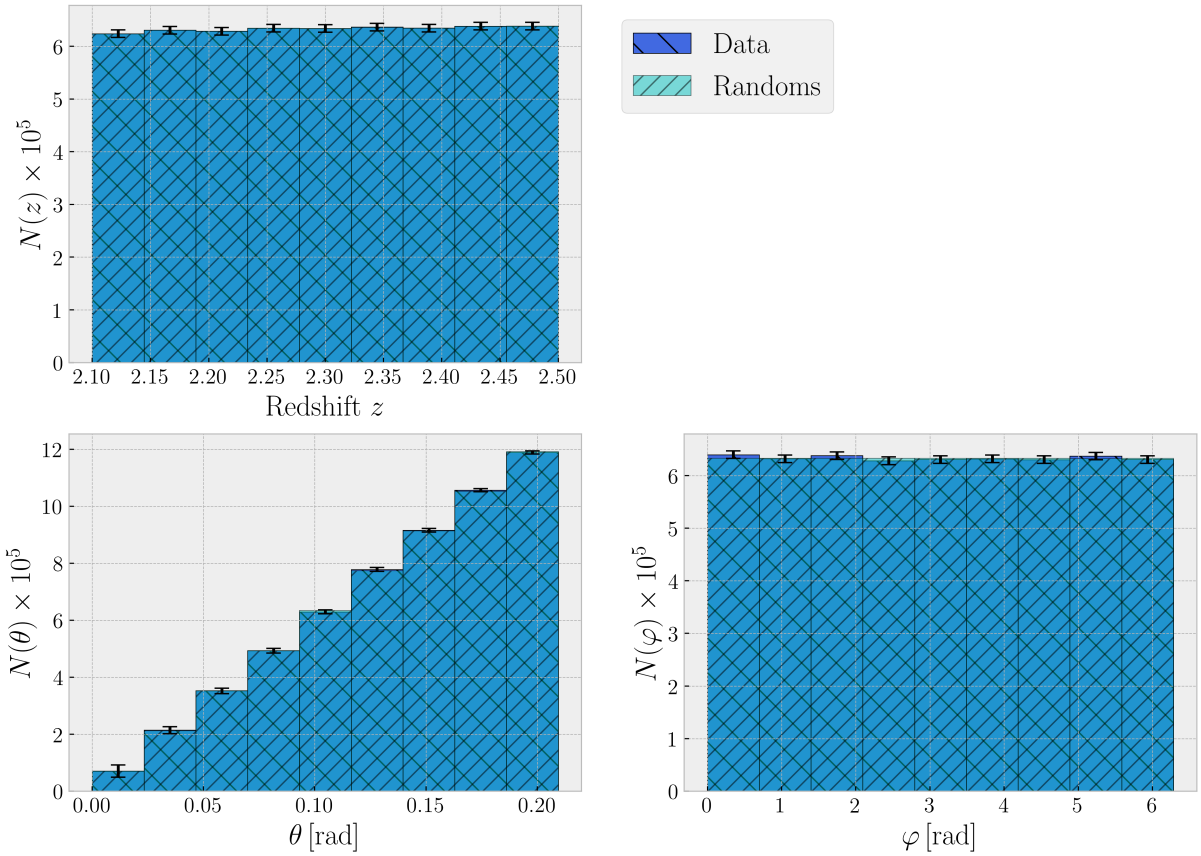


Figure 6.3: Comparison between a random catalogue following the same angular (*bottom panels*) and redshift (*top panel*) distributions as the dark matter particles (data). The random number count is already normalised by the factor $\alpha = N_h/N_r$. We can see that the redshifts of the random objects follow exactly those of the real data, whereas uniform distribution in the azimuth coordinate φ for the random catalogue differs from the data, as the random particles do not have any intrinsic clustering. Despite of that, the differences are well contained within the Poissonian error bars, which indicates that our random samples are valid for analysis.

The particles in real space will be used to test for the halo bias (see Section 6.2.2.2). As discussed in Chapter 3 and as shall be resumed in Section 6.3, to build the power spectrum estimator one needs the selection function to create a random Poissonian catalogue to define the overdensity fluctuation field $F(\boldsymbol{r})$. In this work, the selection function is sampled from an uniform catalogue containing N_r particles and then applied to the conventional FKP estimator. In Figure 6.3 we show the distribution for both the random and the data catalogues, containing respectively N_r and N_h objects. Because the particle catalogue is very dense ($N_h \sim 10^7$ particles), we took $N_r = 10N_h$. The same behaviour was observed for the halo data and random samples, with the data points agreeing with the random distribution based within the Poissonian error bars (see caption of Figure 6.3). For this reason, we do not show them here.

6.2.2 Halo catalogue

From the *real space* particles, the halo catalogue was created with the Rockstar halo finder (Behroozi et al., 2012), using a friends-of-friends (FOF) algorithm with linking length $b = 0.28$ in order to detect over 10^7 halos in the light cone.

After going through the ray tracing algorithm, which is crucial to connect the halos and the observer, the perturbed three-dimensional positions of halos were obtained and the resulting file consists of three mock surveys (see Figure 6.4) contained within the range $0.0 \lesssim z \lesssim 7.1$, with different survey areas. The survey that will be used in this work spans the range of comoving look-back distance from 275 up to 4560 Mpc/h.

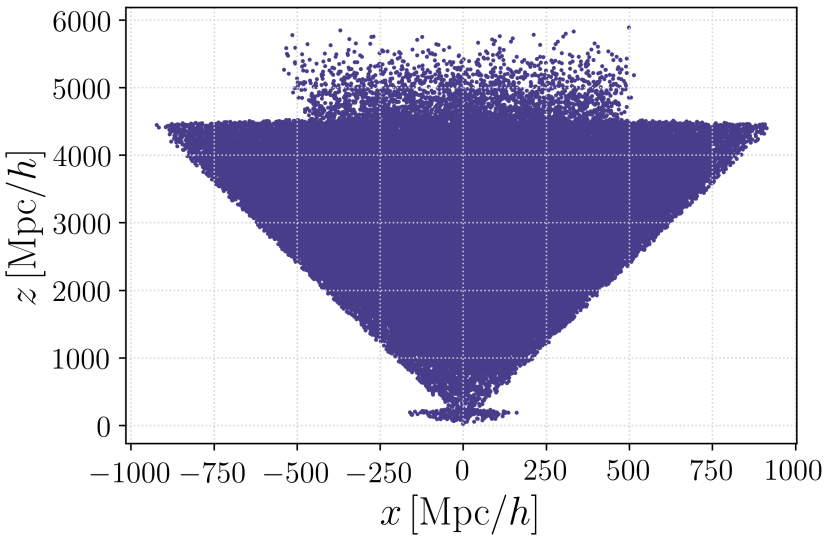


Figure 6.4: The *gevolution* light cones available for our work. Axes x and z are Cartesian coordinates in units of Mpc/h. We only consider the intermediate light cone, for which $\theta \leq 0.21$ rad, extending from $0 \lesssim z \lesssim 4000$ [Mpc/h]. In the figure, we do not shown all the halos available in the simulation.

We limit ourselves to the high-redshift region between $z_{\rm min} = 1.7$ and $z_{\rm max} = 2.9$, with redshift bins of size $\Delta z = 0.4$ which kept the variation of the growth function within the 5% limit[3]. Each redshift bin has an effective volume of ~ 0.7 (Gpc/h)^3 given the chosen cosmology and the sky fraction $f_{\rm sky} \sim 0.01$. After this redshift selection, we were left with 8.5×10^6 dark matter halos.

The high-redshift binning was chosen to deliver a reasonable volume necessary for the observation of the relativistic features at large scales, giving an effective fundamental mode of $k_{\rm F} = 2\pi/V^{1/3} \approx 7 \times 10^{-3}\,(h/{\rm Mpc})$. In a future work we will present the results of the same analysis, but in the full-sky case. The current survey area of ~ 400 deg^2 is compatible with the current survey areas available for a cross-correlation analysis (Zhao et al., 2021).

[3]This criterion was chosen to keep halos of different evolutionary stages somewhat separated.

The final halo catalogue was then separated into three halo samples per redshift bin, of different masses such that, at each redshift bin, the number of halos was the same in each sample. The main properties of these samples are detailed in Table 6.1, and we give a graphical visualisation of each redshift slice in Figure 6.5. Because more massive halos are expected at lower redshifts, the effective redshift $\bar{z}$ of each halo sample varies slightly, but only by less than 0.5%; therefore, we considered the values shown in the aforementioned table as the respective central redshift. The halo population incorporating all halos is referred to as $H_{\rm all}$ in what follows.

Table 6.1: Specifications of the halo samples, selected to match the number density for each population, yielding $\bar{n}_0 \sim \bar{n}_1 \sim \bar{n}_2$. The mean redshift $\bar{z}$ is obtained from all the halos within each redshift bin, as the effective redshift of each halo sample differs from $\bar{z}$ in a sub-percent level ($\lesssim 0.5\%$). The biases have been computed from the monopole of the power spectrum by fitting a linear polynomial to the ratio between the halo auto-spectra and the real-space linear matter power spectrum (see Section 6.2.2.2), differing from the Tinker bias by $\sim 5\%$. For the three redshift bins, the volumes are such that the fundamental mode of observation is $k_{\rm F} = 2\pi/V^{1/3} \sim 7 \times 10^{-3}\,(h/\mathrm{Mpc})$.

	# halos	Mean mass $[M_\odot/h]$	Bias (fit)	$\bar{n}(\bar{z})$ $[\mathrm{Mpc}/h]^{-3}$
		$\bar{z} = 1.89$		
All	480643	4.41×10^{12}	2.927	7.053×10^{-4}
H_0	160081	1.86×10^{12}	2.551	2.349×10^{-4}
H_1	160547	2.83×10^{12}	2.758	2.356×10^{-4}
H_2	160015	8.54×10^{12}	3.477	2.348×10^{-4}
		$\bar{z} = 2.29$		
All	326899	3.85×10^{12}	3.469	4.666×10^{-4}
H_0	109003	1.83×10^{12}	3.020	1.556×10^{-4}
H_1	108809	2.66×10^{12}	3.270	1.553×10^{-4}
H_2	109087	7.05×10^{12}	4.154	1.557×10^{-4}
		$\bar{z} = 2.69$		
All	205678	3.44×10^{12}	4.214	2.947×10^{-4}
H_0	68501	1.80×10^{12}	3.735	9.815×10^{-5}
H_1	68550	2.52×10^{12}	4.006	9.822×10^{-5}
H_2	68627	5.98×10^{12}	4.932	9.833×10^{-5}

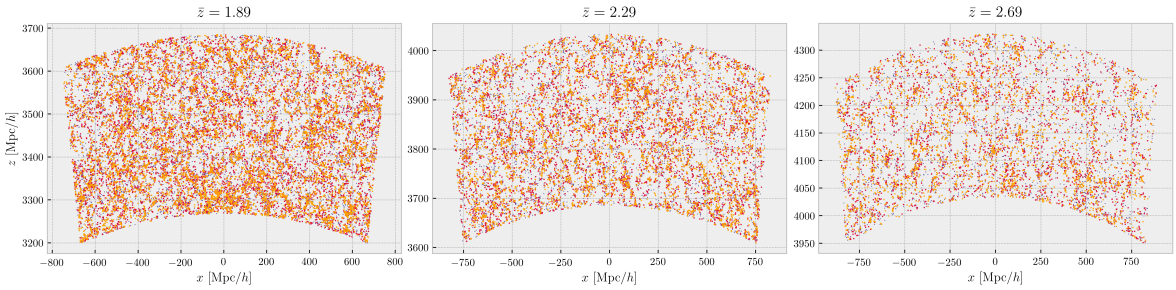


Figure 6.5: The three halo samples H_0 (*blue*), H_1 (*red*) and H_2 (*orange*), for the three redshift slices, as described in Table 6.1. The apparent constant density comes from the fact that our selection does not come from the halo mass, but from bins containing a constant number of halos.

6.2.2.1 Halo mass function

For completeness we computed the mass function of our full halo sample between $z = 0.05$ and $z = 0.465$ (effective redshift $\bar{z} = 0.34$) for comparison with the full-sky RayGal simulation[4] employed in the analysis of Breton et al. (2018). The halo mass function describes the probability of having a comoving number density of halos at redshift z in the range $[\ln M, \ln M + \mathrm{d}\ln M]$:

$$\frac{\mathrm{d}\,\bar{n}(M,z)}{\mathrm{d}\ln M} = \frac{\bar{\rho}_{m,0}}{M} f(\sigma) \frac{\mathrm{d}\ln \sigma^{-1}}{\mathrm{d}\ln M}, \tag{6.3}$$

with $\bar{\rho}_{m,0}$ the comoving background matter density today, $f(\sigma)$ the multiplicity function, and σ the overdensity variance smoothed in a sphere of radius R. The multiplicity function can be computed analytically from the spherical collapse model (Press and Schechter, 1974) or the ellipsoidal collapse (Sheth and Tormen, 1999), or from numerical fits (Tinker et al., 2008).

The RayGal simulation consists of a set of high-resolution Newtonian N-body simulations, whose halos have been ray-traced to the redshift space position, rendering them with almost all properties of our halos. The RayGal light cone was build from 300 snapshots to avoid time discretisation effects.

Our analysis was based on halo masses $M_{200b} \equiv M$ defined within the density thresholds of $\Delta = 200$, whose correspondence with the parameters fit of the Tinker mass function (Tinker et al., 2008) (solid lines in Figure 6.6) is straightforward. In Corasaniti et al. (2018), RayGal halo mass functions were computed from the snapshots and were based on the Sheth-Tormen (Sheth and Tormen, 1999) fit, with the halo identified with the spherical overdensity (SO) method, and thus the halos are more closely connected to the ellipsoidal collapse employed in the Sheth-Tormen fit (Desjacques et al., 2018).

However, for the RayGal light cone, halos were identified via a friends-of-friends (FOF) algorithm, just like in our catalogue. In Smith et al. (2017), differences with numerical fits seem at the low-mass end are also present, and they conclude that such discrepancies are associated with the comparison between different halo finder methods (SO and FOF). They computed the mass function using the SO correspondent, and just as in our case, found the same behaviour at low masses. In the catalogue employed in our analysis, $\sim 1.1 \times 10^6$ halos with $M_{\rm vir} \in [0.518, 4.862] \times 10^{12}$ $(M_\odot/h)$ were discarded for having their respective M_{200b} null. As pointed out in Smith et al. (2017), small overdensities in large FOF groups might be identified as part of the larger group, leading to a lack of such structures.

Such discrepancies are important if one wishes to paint galaxies to the halos via, e.g., a halo occupation distribution. For our current purposes, the lack of a proper function to describe

[4]The RayGal simulation is contained within $0.05 < z < 0.465$, their effective redshift is the same as the low redshift considered in this section for comparative purposes.

the light cone halo mass function impacted only our ability to predict the linear halo bias (see Section 6.2.2.2), and thus did not pose an issue for the analysis.

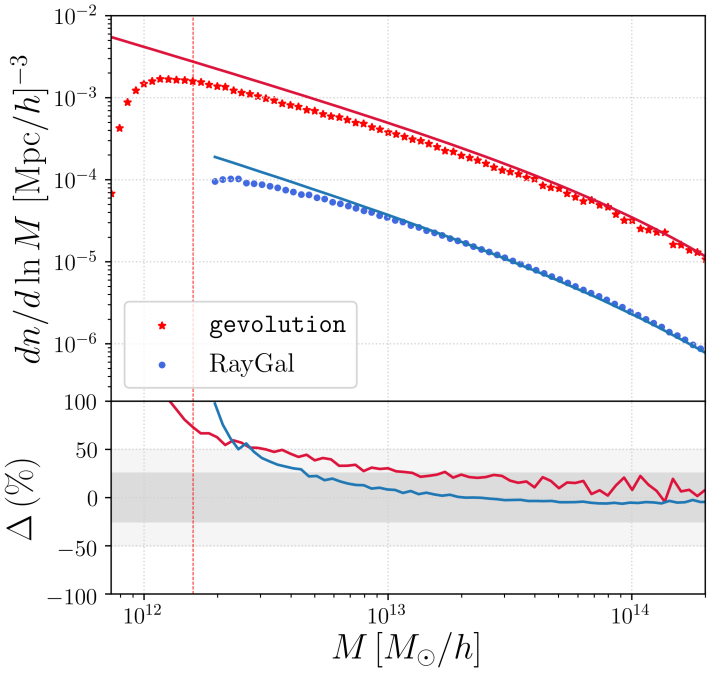


Figure 6.6: Comparison between the halo mass function of the catalogue employed in this analysis (*stars*), with the one from the full-sky RayGal simulation (*dots*) from Breton et al. (2018), with both catalogues in real space. Solid curves correspond to the Tinker mass function fit (Tinker et al., 2008), while the bottom panel shows the relative difference between the fit prediction and the mass function computed from the simulations. Vertical dashed line corresponds to the limit of *gevolution* halos with at least 600 particles. The cosmological parameters in the RayGal simulation that differ from ours are $h = 0.72$, $A_s = 2.431 \times 10^{-9}, \Omega_m = 0.257$, $T_{\mathrm{cmb}} = 2.726$. We stress that the RayGal mass function was multiplied by a 0.1 factor for a cleaner visualisation, as the values were very similar. Differences between the Tinker and Sheth-Tormen (Sheth and Tormen, 1999) mass functions were minor, so we only present the former.

6.2.2.2 Halo bias

The halo mass function describes the fraction of matter inside dark matter halos. So in order to obtain the correct halo statistics, we must account for their position in space. The halo bias, which is defined by the ratio of the halo power spectrum, $P_{hh}(k)$, to the linear dark matter power spectrum, $P_{\mathrm{lin}}(k)$ (Tinker et al., 2010),

$$b^2(k) = \frac{P_{hh}(k)}{P_{\mathrm{lin}}(k)}, \tag{6.4}$$

is best understood within the context of the Peak-Background Split (PBS), where the long-wavelength modes enhances the probability of forming halos by decreasing the threshold $\delta_c(z = 0) = 1.686$ for overdensities which are located at the peak of large-scale (background) fluctu-

ations. It can be either derived from analytical mass functions, giving the Press-Schechter and Sheth-Tormen halo biases, or from Equation (6.4) via numerical simulations. From Tinker et al. (2010), the bias is given by

$$b(\nu) = 1 - A\frac{\nu^a}{\nu^a + \delta_c^a} + B\nu^b + C\nu^c, \tag{6.5}$$

where $\nu = \delta_c/\sigma$ and A, B, C, a, b, and c are parameters fitted from simulations, depending on the matter perturbations at virialisation which is chosen to be $\Delta = 200$. This phenomenological fit proved to be unsatisfactory for our halo samples, for the reasons described in Section 6.2.2.1.

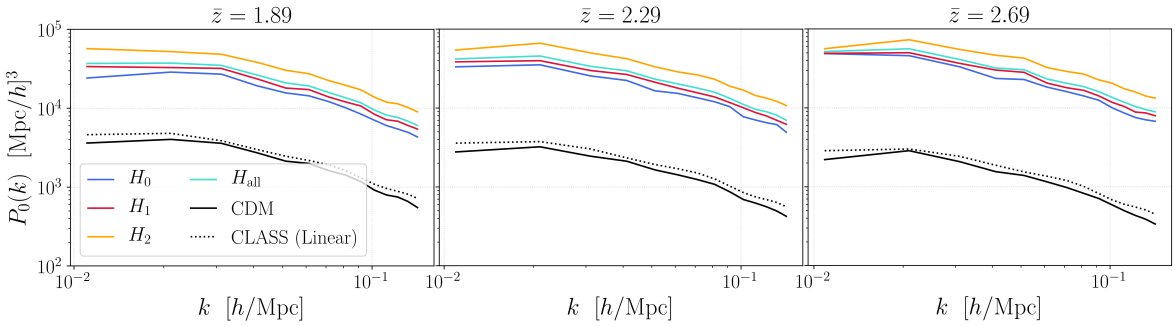


Figure 6.7: Real-space power spectrum monopole of all four halo samples described in Table 6.1, and the cold dark matter particles, for the three high redshift slices considered in our analysis.

We proceeded then with the definition of Equation (6.4) and employed a polynomial fit

$$b^2(k) = b_1^2 + b_2^2 k, \tag{6.6}$$

considering the linear term as the fit for the linear halo biases, neglecting the scale dependence emerging from non-linear effects in the power spectrum. Notice that the estimated spectra P_{hh} employed in this fit are for the halos in real space. The results are shown in Table 6.2 for each halo sample, where we use different methods to obtain the matter power. In one case (PS) we used the theoretical power spectrum obtained from the CLASS Boltzmann solver (Blas et al., 2011) as denominator in Equation (6.4). In a second case (CDM), we computed the biases using the real-space matter power spectrum (Figure 6.7) computed from the cold dark matter particle ensemble instead of CLASS. As we can see, it differs by $\sim 9\%$ from the Tinker value, which we believe comes from fluctuations of the estimator itself.

Figure 6.8 shows the application of this method to estimate the linear bias of our halos, using the linear (*upper panel*), non-linear (HALOFIT, *middle panel*), and the power spectra for the cold dark matter (CDM) estimated directly from the light cone.

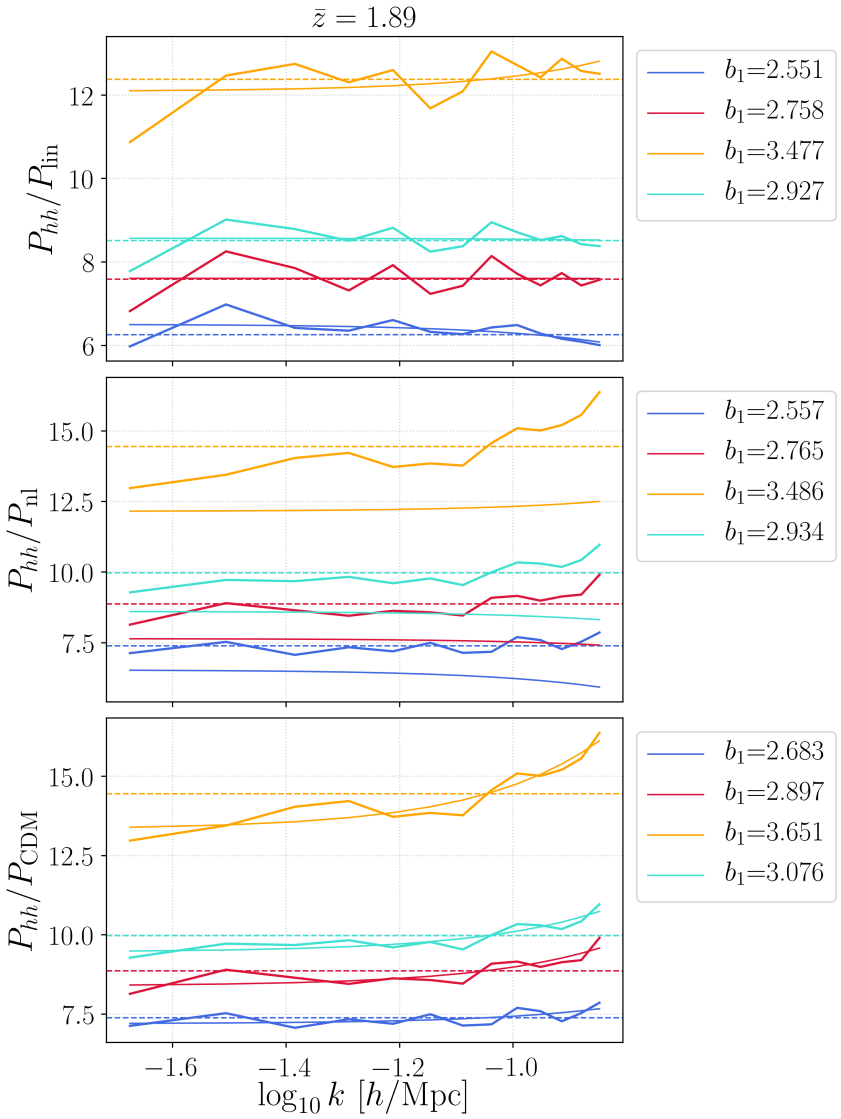


Figure 6.8: Bias estimates following Equation (6.4) and the polynomial fit of Equation (6.6) with the linear (*top*), non-linear (*middle*) and the estimated power spectrum from the cold dark matter particles (*bottom*), for the first redshift slice. Dashed lines represent the mean of the ratios and overestimate the results for larger halo masses. As we can see, the linear bias b_1 differs slightly between the linear and non-linear cases, whereas it is overestimated with respect to the CLASS fit because the particles spectra is slightly below the CLASS prediction. Results from the fit with the linear power spectrum are shown in Table 6.2, column *Bias (PS)*, for all the three redshift slices, and will be adopted for the remaining of the work. This plot follows the same colour reference of Figure 6.7.

In order to verify the consistency of the method and the possible impacts of the window function in the bias estimation, we tested for the approach of Breton et al. (2018): the linear bias was computed by fitting a constant function to the ratio

$$b = \sqrt{\frac{\xi^{\ell=0}_{hh}}{\xi_0}}, \tag{6.7}$$

where ξ_0 is the monopole of the matter autocorrelation function, computed from

$$\xi_0(x) = \frac{1}{2\pi^2} \int \mathrm{d}k\ k^2\ j_0(kx)\ P^{(r)}(k), \tag{6.8}$$

and $\xi_{hh}^{\ell=0}$ is monopole of the halo-halo autocorrelation, computed from the real space catalogues with `CUTE`[5] (Alonso, 2012). The real space linear matter power spectrum[6] was obtained from the CLASS Boltzmann solver (Blas et al., 2011) with the input parameters of the simulation.

This method is not perfect though: we observed a shift in the BAO peak scale for the redshift-space halo samples if compared to the theoretical prediction. Also, we limited ourselves to the range $28 < r_{\mathrm{fit}} < 68$ in units of Mpc/h to proceed with smaller noise. Although time consuming, this method is naturally safe from the mode coupling induced by the window function. The results from this fit is shown in Table 6.2. Using the fits from Equation (6.6) or (6.7) did not lead to significant differences to the overall results, with both methods being tantamount.

Table 6.2: Linear biases for the halo samples considered in this work, estimated with different methods. The Tinker bias is computed from the fit of Tinker et al. (2010), Equation (6.5). The biases in the next two columns were obtained via the polynomial fit of Equation (6.6), with the linear power spectrum from CLASS (PS) or with the power spectrum estimated from the real-space cold dark matter particles (CDM), with the former estimate differing from the Tinker bias by $\sim 3-5\%$. The final column shows the bias parameters estimated from the correlation function, differing from the Tinker bias by $\sim 6\%$. Larger discrepancies for the CDM bias estimates with respect to the others will be due to fluctuations coming from the estimator.

	Mass $[M_\odot/h]$	Bias (Tinker)	Bias (PS)	Bias (CDM)	Bias (CF)
		$\bar{z} = 1.89$			
All	4.4×10^{12}	2.881	2.927	3.076	2.809
H_0	1.9×10^{12}	2.273	2.551	2.683	2.437
H_1	2.8×10^{12}	2.540	2.758	2.897	2.652
H_2	8.5×10^{12}	3.539	3.477	3.651	3.349
		$\bar{z} = 2.29$			
All	3.9×10^{12}	3.473	3.469	3.749	3.196
H_0	1.8×10^{12}	2.803	3.020	3.280	2.772
H_1	2.7×10^{12}	3.114	3.270	3.546	2.993
H_2	7.1×10^{13}	4.212	4.154	4.449	3.827
		$\bar{z} = 2.69$			
All	3.4×10^{12}	4.140	4.214	4.455	3.959
H_0	1.8×10^{12}	3.414	3.735	3.950	3.613
H_1	2.5×10^{12}	3.765	4.006	4.237	3.891
H_2	5.9×10^{13}	4.955	4.932	5.211	4.468

[5] https://github.com/damonge/CUTE/

[6] Differences of using the non-linear matter power spectrum were below the percent level.

6.2.2.3 Halo evolution bias

The evolution bias of LSS tracers quantifies the intrinsic variation in the number of sources in the Universe, and thus gives information about the time evolution of tracers. It is defined in Equation (5.14), and depends on the comoving number density of sources in real space $\bar{n}$. In the case of dark matter halos, this parameter describes the mean number density of halos. The evolution bias account for the true tracer: in the case of a halo catalogue, coming from a simulation, the underlying dark matter distribution is known and the comoving number density of halos is complete (i.e. all halos that were supposed to be found are included in the catalogue), and the evolution bias is completely faithful to the intrinsic cosmological variations of the number density.

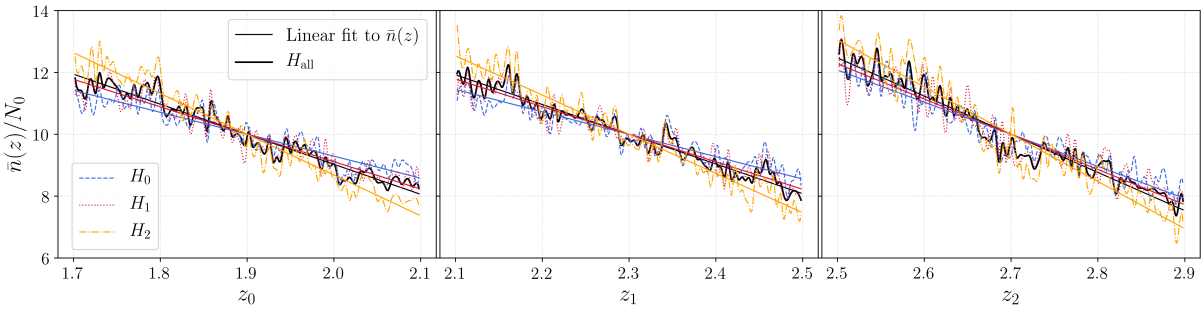


Figure 6.9: The comoving number density of halos, normalised by the total amount N_0 of objects inside each mass bin. Solid lines represent the linear fit of Equation (6.9), whereas dashed lines represent the true $\bar{n}(z)$ of the simulation, computed by dividing each redshift bin into 100 points. The smooth curves are the result of a cubic spline interpolation, for visual reasons.

In Figure 6.9 we show the comoving number density of real space halos, inside each redshift bin considered in the main analysis, normalised by the total number of halos for each mass bin defined for $\bar{n}_0 \approx \bar{n}_1 \approx \bar{n}_2$. As one can see, there is a large variation explained by the fact that more massive halos are more common at lower redshifts, which can be explicitly seem by the slope of the curves. This intrinsic variation is captured by the evolution bias b_e, which is a major parameter entering the relativistic corrections.

Following Beutler and Dio (2020), we fit a linear function to the (unnormalised) comoving number density,

$$\bar{n}(z) = a + bz, \tag{6.9}$$

which leads to the analytical expression for b_e,

$$b_e(z) = \frac{c+1}{c-z} - 1, \tag{6.10}$$

where $c \equiv a/(-b)$.

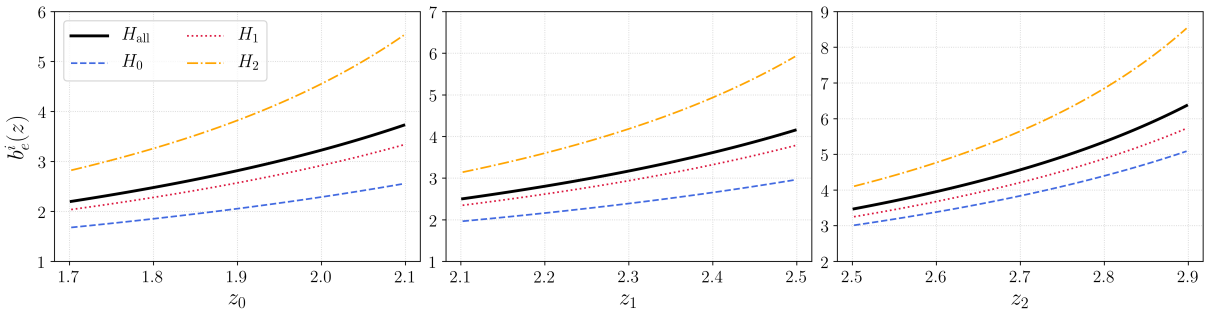


Figure 6.10: Evolution biases for the halo samples described in Table 6.1, at each redshift bin considered in the analysis. Even though the number density is approximately the same for all the halo populations, the intrinsic evolution of each halo population with the redshift gives rise to very different evolution biases. The computation followed the work of Beutler and Dio (2020).

Notice that, even though the comoving mean number density is the same for all the three samples H_0, H_1 and H_2, the different evolution with redshift between the different halo populations, defined by different halo masses, leads to distinct evolution biases, as can be seen in Figure 6.10.

Table 6.3: Parameters entering equations (6.9) and (6.10) to derive the evolution bias of the halo samples.

	a	b	c	$b_e(\bar{z})$
		$\bar{z} = 1.89$		
All	20.105	-6.863	2.929	2.781
H_0	5.521	-1.668	3.310	2.035
H_1	6.337	-2.093	3.027	2.542
H_2	8.247	-3.102	2.658	3.761
		$\bar{z} = 2.29$		
All	14.990	-4.488	3.340	3.132
H_0	4.153	-1.128	3.680	2.367
H_1	4.740	-1.385	3.422	2.905
H_2	6.098	-1.974	3.088	4.120
		$\bar{z} = 2.69$		
All	12.763	-3.636	3.510	4.500
H_0	3.728	-1.017	3.664	3.788
H_1	4.000	-1.118	3.578	4.153
H_2	5.035	-1.501	3.355	5.551

6.3 Power-spectrum multipoles estimator

To compute the power spectrum multipoles we make use of the standard approach proposed by Yamamoto et al. (2006); Bianchi et al. (2015) and Scoccimarro (2015) (for pioneering work see also Yamamoto et al., 2000), which we dub YBS estimator. It is built upon the practical FKP algorithm (Feldman et al., 1994) (see Section 3.2.2.1) to optimally estimate the

power spectrum of galaxy surveys with a varying selection function. As mentioned in Section 6.1, the selection function $\bar{N}$ encodes the spatial modulations of the mean number density of objects. For both spectroscopic and photometric surveys, the selection function accounts for all non-cosmological effects, being sensitive, for example, to the different intrinsic brightness of galaxies.

The selection function gives an estimate of the probability that a galaxy brighter than a certain threshold, at a distance $\boldsymbol{s}$, is included in the sample. Hence, it is intrinsically related to the notion of luminosity function $\Phi(L)$ (Martínez and Saar, 2001). In Wang et al. (2020) a clear example of such fact is given, with the luminosity function of eBOSS quasars (QSO) being used to fit the QSO number density and derive the evolution and magnification biases.

To resume the construction of the estimator, $N_X(\boldsymbol{x}_{ijk})$ denotes either the count-in-cells of the the random catalogue, $X = r$, or the data (halo) catalogue, $X = h$, where $\boldsymbol{x}_{ijk}$ is the position of each cell in a three-dimensional grid obtained by a mass assignment scheme, e.g. Nearest Grid Point (NGP), Cloud In Cell (CIC), or Triangular Shaped Cloud (TSC). In this analysis we consider the simplest NGP assignment.

Recall the definition of the weighted galaxy fluctuation, or the overdensity field:

$$F(\boldsymbol{x}) = \frac{w(\boldsymbol{x})}{\mathcal{N}} \left[n_h(\boldsymbol{x}) - \alpha\, n_r(\boldsymbol{x}) \right], \tag{6.11}$$

where $n_h(\boldsymbol{x}) = \sum_{i=1}^{N_h} \delta^D(\boldsymbol{x} - \boldsymbol{x}_i)$ is the number density which will be written as a grid, after mass assignment scheme is chosen. Therefore, in practice $n_h(\boldsymbol{x}) = N_h(\boldsymbol{x}_{ijk})$ is the count-in-cells grid and $n_r(\boldsymbol{x})$ is the corresponding quantity for the random catalogue, which is obtained by randomly sampling α^{-1} times more objects within the survey volume, with the same selection function as the real data (see Chapter 3).

The results presented here do not employ a weighting scheme, i.e. $w(\boldsymbol{x}) = 1$, and we follow Jeong (2010) for the implementation of the quadratic estimator. The normalisation factor will be given by

$$\mathcal{N} \approx \frac{\alpha^2}{\ell_x \ell_y \ell_z} \sum_{\boldsymbol{x}_{ijk}} N_r^2(\boldsymbol{x}_{ijk}), \tag{6.12}$$

and the shot noise, only relevant for the monopole term, will be

$$P_{\rm shot} \approx \ell_x \ell_y \ell_z \left(\frac{1+\alpha}{\alpha} \right) \frac{\sum_{\boldsymbol{x}_{ijk}} N_r(\boldsymbol{x}_{ijk})}{\sum_{\boldsymbol{x}_{ijk}} N_r^2(\boldsymbol{x}_{ijk})}, \tag{6.13}$$

where $\ell_i \equiv L_i/n_i$ is the size of each cell dimension in units of Mpc$/h$. In this work we choose $\ell_i = 10$ Mpc$/h$.

After the construction of these quantities, the power spectrum multipoles can be obtained by the YBS estimator, which we now briefly discuss. The whole idea of this method relies in

generalising the power spectrum to *local* regions in space, where statistical homogeneity may be assumed. These regions are defined by a single middle line-of-sight $\boldsymbol{d} = (\boldsymbol{s}_1 + \boldsymbol{s}_2)/2$, as shown on the left of Figure 6.11. Then, the corresponding power spectrum at this region is

$$P(\boldsymbol{k}_1, \boldsymbol{k}_2) = \int \mathrm{d}^3 s_1 \int \mathrm{d}^3 s_2 \; \xi(\boldsymbol{s}_1, \boldsymbol{s}_2)\, \mathrm{e}^{i\boldsymbol{k}_1\cdot\boldsymbol{s}_1}\, \mathrm{e}^{-i\boldsymbol{k}_2\cdot\boldsymbol{s}_2}. \tag{6.14}$$

Notice that

$$\mathrm{e}^{i\boldsymbol{k}_1\cdot\boldsymbol{s}_1}\, \mathrm{e}^{-i\boldsymbol{k}_2\cdot\boldsymbol{s}_2} = \mathrm{e}^{i\boldsymbol{d}\cdot\boldsymbol{q}}\, \mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{s}}, \tag{6.15}$$

where the Fourier transform of the local configuration is shown on the right of Figure 6.11. With this change of coordinates, one can see that the local power spectrum can be obtained by taking the Fourier transform of the first component of the local correlation function $\xi(\boldsymbol{s}_1, \boldsymbol{s}_2) = \xi(\boldsymbol{s}, \boldsymbol{d})$. Finally, the multipoles of the local power spectrum can be obtained from the Legendre expansion

$$P(\boldsymbol{k}, \boldsymbol{d}) = \sum_\ell P_\ell(k, d) \mathcal{L}_\ell(\hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{d}}). \tag{6.16}$$

Lastly, the inversion of this relation yields the power spectrum multipoles:

$$\hat{P}_\ell(k) = \Big\langle \frac{2\ell + 1}{2\mathcal{N}} \int \mathrm{d}^3 s_1 \int \mathrm{d}^3 s_2 \, F(\boldsymbol{s}_1) F(\boldsymbol{s}_2)\, \mathrm{e}^{-i\boldsymbol{k}\cdot(\boldsymbol{s}_2 - \boldsymbol{s}_1)} \, \mathcal{L}_\ell(\hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{d}}) - S_\ell(\boldsymbol{k}) \Big\rangle, \tag{6.17}$$

where the brackets correspond to an average over k-shells and S_ℓ is the shot-noise term, only relevant for the monopole.

To speed up the computation of the multipoles by means of FFTs, the YBS estimator takes the end-point LOS $\boldsymbol{s}_1$. It is worth mentioning that this LOS intrinsically generates odd multipoles that may impact the signal we are trying to measure. This happens simply because we are breaking the symmetry between the galaxy pair by fixing the LOS to the position of a specific galaxy (see Figure 1 of Beutler et al., 2019). Hence, this must be accounted for in the window function, as will be discussed in Section 6.3.1.

With the adoption of this LOS, the monopole can estimated as

$$\hat{P}_0(k) = \frac{1}{\mathcal{N}} \langle F_0(\boldsymbol{k}) F_0^*(\boldsymbol{k}) - S_0 \rangle, \tag{6.18}$$

where

$$F_0(\boldsymbol{k}) = \int \mathrm{d}^3 x \; F(\boldsymbol{k})\, \mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{x}} \tag{6.19}$$

is the Fourier transform of the overdensity field with no weight,

$$F(\boldsymbol{x}) = n_h(\boldsymbol{x}) - \alpha\, n_r(\boldsymbol{x}), \tag{6.20}$$

and the dipole is obtained by

$$\hat{P}_1(k) = \frac{3}{\mathcal{N}} \langle F_0(\boldsymbol{k}) F_1^*(\boldsymbol{k}) \rangle, \tag{6.21}$$

with

$$F_1(\boldsymbol{k}) = \sum_{i=x,y,z} \hat{k}_i \, f_{1,i}(\boldsymbol{k}), \tag{6.22}$$

and

$$f_{1,i}(\boldsymbol{k}) = \int \mathrm{d}^3 r \; \hat{r}_i F(\boldsymbol{r}) \; \mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{r}}. \tag{6.23}$$

For higher order multipoles, we refer the interested reader to Bianchi et al. (2015) and Beutler et al. (2019) for the even and odd ones, respectively.

Finally, to compute the cross-dipole, $F_0(\boldsymbol{k})$ and $F_1(\boldsymbol{k})$ are built from the first and second tracers, respectively, and the normalisation becomes (Beutler and Dio, 2020):

$$\mathcal{N} \approx \frac{\alpha_{(1)}\alpha_{(2)}}{\ell_x \ell_y \ell_z} \sum_{\boldsymbol{x}_{ijk}} N_r^{(1)}(\boldsymbol{x}_{ijk}) \, N_r^{(2)}(\boldsymbol{x}_{ijk}). \tag{6.24}$$

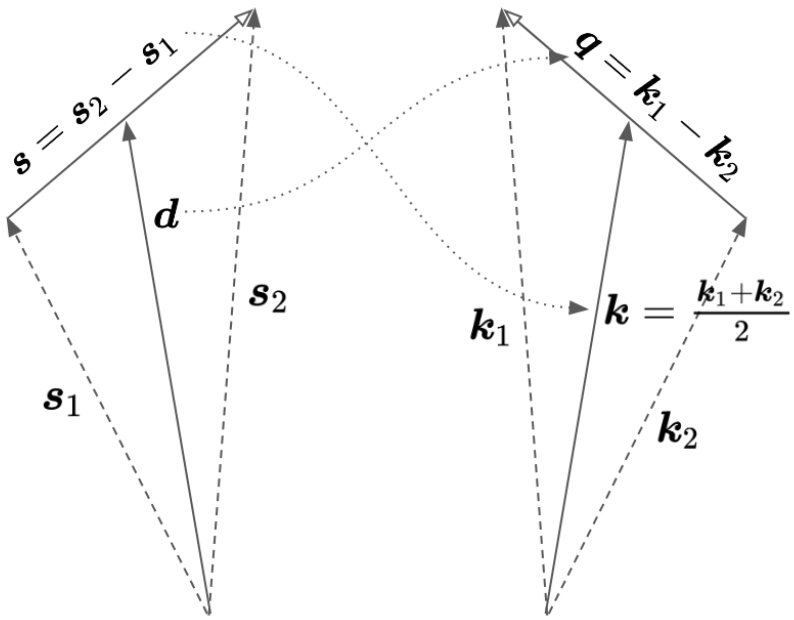


Figure 6.11: The mirrored scheme at the right-hand side corresponds to the Fourier transform of the left hand configuration. The Fourier counterpart of $\boldsymbol{d}$ and $\boldsymbol{s}$ are, respectively, $\boldsymbol{q}$ and $\boldsymbol{k}$ (Reimberg et al., 2016). On the left side, the configuration for the local estimator $P(\boldsymbol{k}, \boldsymbol{d})$ is depicted. The YBS estimator corresponds to integrating over all possible lines-of-sight $\boldsymbol{d}$ and averaging over k-bins. In this sketch, $\boldsymbol{s} = \boldsymbol{s}_2 - \boldsymbol{s}_1$, with the observer located at the lower vertex.

6.3.1 Window function

Because we are not working inside a box, with a constant selection function and a trivial geometry, we need to account for the survey window function. It is employed in the observed power spectra multipoles from the theoretical predictions for comparative purposes and imparts substantial effects on large scales and on the odd multipole moments. Therefore, its inclusion is mandatory.

We begin by recalling that the observed density field $\hat{\delta}$ is given by

$$\hat{\delta}(\boldsymbol{x}) = W(\boldsymbol{x})\,\delta(\boldsymbol{x}), \tag{6.25}$$

where $\delta(\boldsymbol{x})$ is the true underlying density field and $W(\boldsymbol{x}) = w(\boldsymbol{x})\bar{n}(\boldsymbol{x})$ accounts for the survey geometry and local weighting w scheme (Feldman et al., 1994). Therefore, the observed correlation function is given by

$$\hat{\xi}(\boldsymbol{s}_1, \boldsymbol{s}_2) = W(\boldsymbol{s}_1)\,W(\boldsymbol{s}_2)\,\xi(\boldsymbol{s}_1, \boldsymbol{s}_2). \tag{6.26}$$

Notice that we can write $\boldsymbol{s}_2 = \boldsymbol{s}_1 + \boldsymbol{s}$, where $\boldsymbol{s}$ is the pair separation, so that the correlation function may also be written as $\xi(\boldsymbol{s}_1, \boldsymbol{s})$.

In Fourier space we obtain the well-known convolution result for the overdensity field,

$$\hat{\delta}(\boldsymbol{k}) = \int \frac{\mathrm{d}^3k'}{(2\pi)^3}\; W(\boldsymbol{k} - \boldsymbol{k}')\,\delta(\boldsymbol{k}'), \tag{6.27}$$

yielding the three-dimensional observed power spectrum:

$$\begin{aligned} \hat{P}(\boldsymbol{k}) &= \int \frac{\mathrm{d}^3q}{(2\pi)^3}\; |W(\boldsymbol{k} - \boldsymbol{q})|^2 P(\boldsymbol{q}), \\ &= \int \frac{\mathrm{d}^3q}{(2\pi)^3}\; |W(\boldsymbol{q})|^2\, P(\boldsymbol{k} - \boldsymbol{q}), \end{aligned} \tag{6.28}$$

where we made use of the fact that δ and W are real quantities.

One possible way to compare theory and estimates is to deconvolve the survey window from $\hat{P}$; however, since convolution in Fourier space destroys information, the deconvolution of the window is an attempt to recover this intrinsic information loss in the signal analysis. The standard procedure (Wilson et al., 2016; Beutler et al., 2016) consists, instead, in computing the multipoles of $|W(\boldsymbol{q})|^2$ to convolve the theoretical power spectrum to obtain $\hat{P}$, where

$$|W(\boldsymbol{q})|^2 = \int \mathrm{d}^3s\; \mathrm{e}^{-i\boldsymbol{q}\cdot\boldsymbol{s}}\, W^2(\boldsymbol{s}), \tag{6.29}$$

and

$$W^2(\boldsymbol{s}) \equiv \int \mathrm{d}^3s_1\; W(\boldsymbol{s}_1)\,W(\boldsymbol{s}_1 + \boldsymbol{s}). \tag{6.30}$$

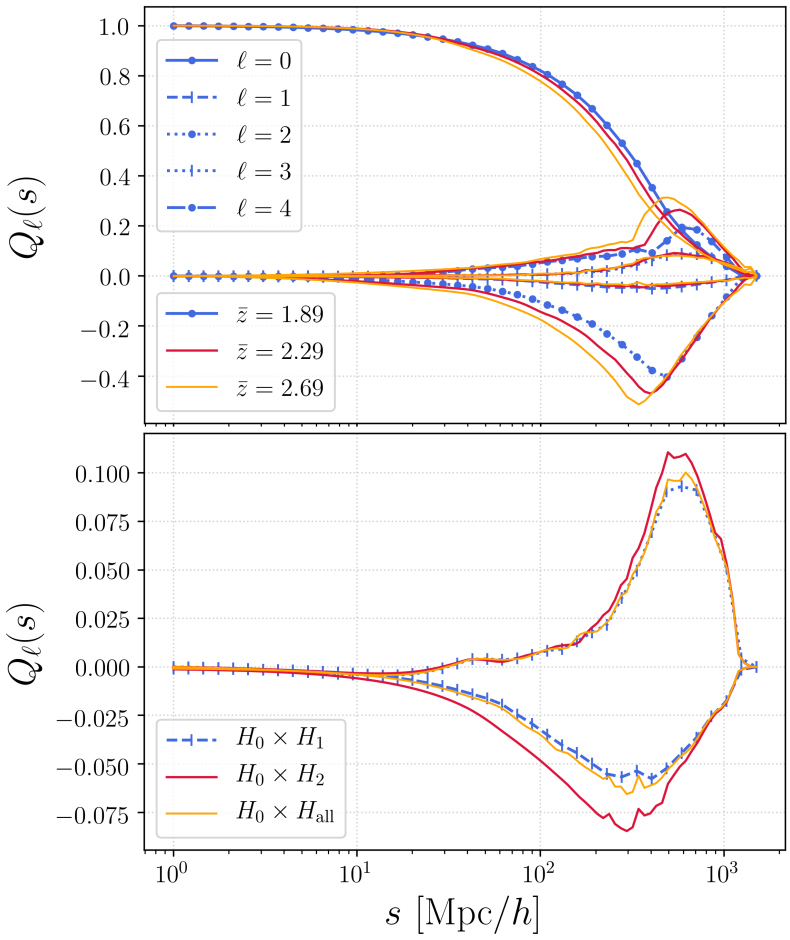


Figure 6.12: Multipoles of the window function computed semi-analytically from the sample radial and angular distribution. The top panel shows the window for the catalogues of the halo population H_0 at the redshift slices considered for analysis (different colours), with different line styles representing different multipoles: dots depict the even, while vertical bars the odd ones. Differences between the tracers as a result of different selection functions are negligible, and thus we only show the autocorrelation of the H_0; however, the cross-correlation among tracers has a larger impact for the odd multipole moments, as shown in the bottom panel for a fixed redshift ($\bar{z} = 1.89$).

We shall write $W^2(\boldsymbol{s}) \equiv Q(\boldsymbol{s})$ and $|W(\boldsymbol{k})|^2 = Q(\boldsymbol{k})$. Notice that this depends on the local LOS, which is taken to be $\boldsymbol{s}_1$ (end-point LOS) in the case of the YBS multipoles estimator. Hence, the multipoles of the "window function", with respect to a LOS $\hat{\boldsymbol{d}} = \hat{\boldsymbol{s}}_1$, are given by

$$Q_\ell(s) = \frac{2\ell+1}{4\pi} \int \mathrm{d}^3 d \int \mathrm{d}\Omega_s \; Q(\boldsymbol{s}, \boldsymbol{d})\, \mathcal{L}_\ell(\hat{\boldsymbol{s}} \cdot \hat{\boldsymbol{d}}) \tag{6.31}$$

in configuration space and

$$Q_\ell(k) = \frac{2\ell+1}{4\pi} \int \mathrm{d}^3 d \int \mathrm{d}\Omega_k \; Q(\boldsymbol{k}, \boldsymbol{d})\, \mathcal{L}_\ell(\hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{d}}). \tag{6.32}$$

in Fourier space. The integrals on $d\Omega_s$ and $d\Omega_k$ run over the angles between, respectively, $\hat{\boldsymbol{s}}$

and $\hat{\boldsymbol{k}}$ with the LOS $\hat{\boldsymbol{d}}$: $\int d\Omega = \int_0^{\pi} d\theta \sin\theta \int_0^{2\pi} d\varphi$. After the integration over all angles, $Q_\ell(s)$ can be obtained by the final integration over all possible LOS. Because our "survey" geometry is well-behaved[7], consisting of a simple angular selection delimited by the light cone opening angle (which sets the upper limit in the θ integral, $Q_\ell(s)$ can be obtained semi-analytically by considering the halos radial selection function (shown in Figure 6.9), and performing a plain angular integration, with the radial integral limited to the redshift range of the sample through the proper inclusion of Heaviside step functions.

With the semi-analytical window function, obtained from its definition in Equation (6.31), differences between halo populations as a result of different selection functions were below the 2% level for the autocorrelation case, with the same behaviour being observed at the estimated windows from the random catalogues. In contrast, differences between redshift bins are more relevant, as can be seen in Figures 6.12 and 6.13, and are fully included in our analysis.

If we plug in the expressions for $Q(\boldsymbol{k}, \boldsymbol{d})$ and $Q(\boldsymbol{s}, \boldsymbol{d})$, just as in the local power spectrum case, we see that to estimate the multipoles $Q_\ell(k)$ and $Q_\ell(s)$ we just apply the usual power spectrum and correlation function estimators. Since the explicit convolution of Equation (6.28) is computationally expensive, and so is the computation of $Q_\ell(s)$ directly from the random pair correlation, as the survey window function (random catalogue) contains 10^8 particles to completely fill the survey region, one possibility is to compute the power spectrum multipoles of the random catalogues. We tried this approach, obtaining $Q_\ell(k)$ by means of FFTs, which are then taken to configuration space.

From the straightforward product of Equation (6.26), the Legendre expansion of (6.26) results in (Wilson et al., 2016; Beutler et al., 2016, 2019; Beutler and Dio, 2020):

$$\hat{\xi}_0(s) = \xi_0(s)Q_0(s) + \frac{1}{5}\xi_2(s)Q_2 + \frac{1}{9}\xi_4(s)Q_4(s) + \dots \tag{6.33}$$

and

$$\hat{\xi}_1(s) = \xi_0(s)Q_1(s) + \xi_2(s)\left[\frac{2}{5}Q_1(s) + \frac{9}{35}Q_3(s)\right] + \frac{4}{21}\xi_4(s)Q_3(s) + \dots, \tag{6.34}$$

and from an inverse Hankel (1D Fourier) transform of these equations we finally obtain the convolved power spectrum multipoles of Equation (6.28). For the monopole we have

$$\hat{P}_0(k) = 4\pi \int s^2 \, \mathrm{d}s \; j_\ell(ks)\,\hat{\xi}_0(s). \tag{6.35}$$

[7]For surveys whose angular selection is too complicated, precluding an analytical derivation of the window function, one can compute it from FFTs of a random catalogue, as we discuss further, or from a random-random pair count (Wilson et al., 2016; Beutler et al., 2016); this latter option, however, can be too time consuming for surveys that require a large number of random objects and Breton and de la Torre (2021) proposed a similar approach to "analytically" compute the pair counts with high precision and without much computational effort.

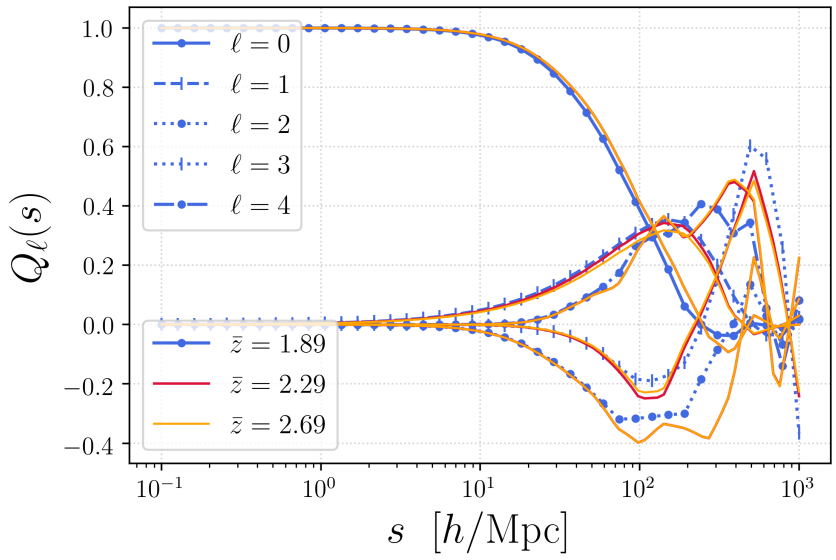


Figure 6.13: Multipoles of the window function computed from the inverse Hankel transform of YBS estimator for the random catalogues of the halo population H_0. Different line styles represent different multipoles, dots depict the even and vertical bars the odd ones (which have been multiplied by 10 for visual reasons). Different colours show the window function at different redshift slices. Differences between the tracers as a result of different selection functions are negligible, and thus we only show the autocorrelation of the H_0; however, the intrinsic fluctuations of the FFT estimator makes the window function very noisy at large scales, causing a variation between redshifts (and also tracers) larger than what is expected from Equation (6.31).

We compared the second (FFT) approach with the semi-analytical window and found disagreements for large scales due to possible instabilities in the Hankel transform, and the limited k range and fluctuations from the FFT estimator that made the window function very noisy for large s. In addition to these differences at larger scales, we have found an inconsistency with the octupole that we are still investigating. Also, for a given number of bins (e.g. 29 Fourier bins), as discussed in Section 6.4, this FFT approach worked just fine; however, for a smaller number of bins (e.g. 14 bins) it yielded nonphysical results. Still, the convolution procedure described above gave a consistent monopole for the semi-analytical window function. For this reason, we opted to analyse the impact of the window by using the semi-analytical result.

In Figure 6.14 we show the relative difference between what is obtained from Equation (6.35) and the linear power spectrum monopole. The latter is computed from the Legendre expansion, with the Newtonian monopole coefficient c_0 shown in Equation (5.21) and the linear real-space power spectrum extracted from CLASS. The former is computed in three different ways. First, we only include the first term in Equation (6.33), which corresponds to the case where there is no coupling between higher order multipoles with $\ell = 0$. Then we account for the leakage of the quadrupole (second term in Equation 6.33), and of the hexadecapole (full Equation 6.33). As we can see, the inclusion of the quadrupole becomes relevant for scales

$k \lesssim 0.025\ h/\mathrm{Mpc}$, while the hexadecapole contribution is negligible. For our range of scales, the differences between the convolved and the redshift-space theory (unconvolved) is of the order of 5%. Thus, we do not see a large impact in face of our error bars (see Figure 6.16). We stress that this behaviour is consistent with the BOSS DR12 anisotropic analysis (Beutler et al., 2016).

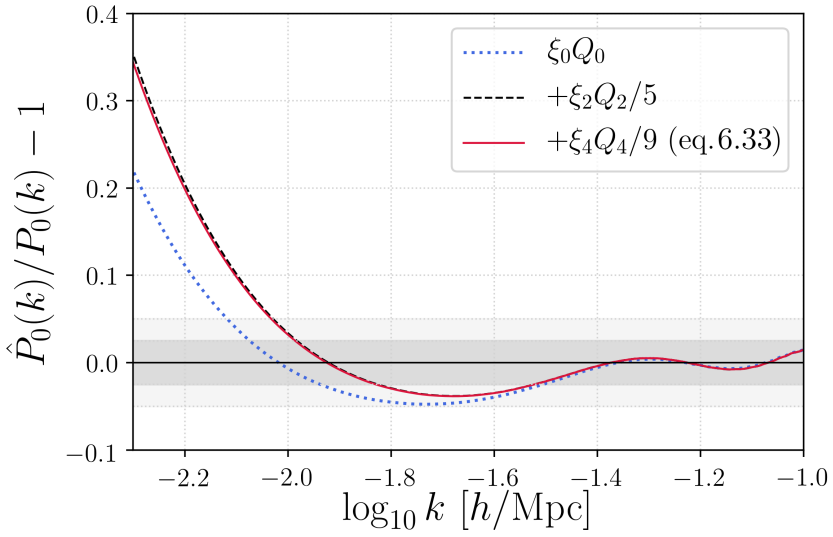


Figure 6.14: Relative difference between the convolved power spectrum $\hat{P}_0$ and the linear power spectrum monopole obtained from the Newtonian redshift-space coefficient (as described in equations 5.17, 5.20 and 5.21). The convolved theory was obtained via Equation (6.35), for the cases in which there is no leakage of the quadrupole and hexadecapole to the monopole (blue dashed curve), when we only account for the quadrupole contribution to the monopole (black dashed curve), that is, considering the first two terms in the right-hand side of Equation (6.33), and when the full expression is given (red solid curve). Shaded regions corresponds to differences of 0.05 (light grey) and 0.025 (dark grey). The inclusion of the quadrupole becomes relevant for scales $k \lesssim 10^{-1.6} = 0.025\ h/\mathrm{Mpc}$, while the hexadecapole contribution is negligible.

Finally, following Beutler et al. (2019) and Beutler and Dio (2020), the convolved dipole is given by:

$$\hat{P}_1(k) = -3i \int \mathrm{d}s\ s^2\, j_1(ks)\, \hat{\xi}_1(s) - iQ_1(k) \int \mathrm{d}s\ s^2\, \hat{\xi}_0(s). \tag{6.36}$$

This represents the leakage of even multipoles to the dipole, and must be accounted for if one wishes to analyse the pure signal of the cross-dipole $P_1^{\alpha\beta}(k)$.

6.4 Results

For the objects under analysis (dark matter halos), the concept of luminosity function can replaced by the halo mass function $\mathrm{d}\bar{n}/\mathrm{d}\ln M$, which gives the probability of having a mean number of halos, within some comoving volume, with mass in the range $[\ln M_i, \ln M_i + dM]$.

Thereby, the selection function coincides with the comoving mean number density of halos within a certain mass bin $[M_i, M_{i+1}]$, in complete analogy to the definition of selection function from $\Phi(L)$:

$$\bar{n}(z, \Delta M) = \int_{M_i}^{M_{i+1}} \mathrm{d}\ln M \, \frac{\mathrm{d}\bar{n}}{\mathrm{d}\ln M}. \tag{6.37}$$

Still for the specific case of dark matter halos, the magnification bias is identically zero, $s = 0$, and the only term accounting for the mass function variations is the evolution bias, which explores its dependency with time: this is related to the fact that halos can merge to form more massive structures and, thereby, their number counts are not conserved. The evolution biases $b_e^i(z)$ of each halo sample $i = \{H_0, H_1, H_2, H_{\mathrm{all}}\}$ considered in this work, within each redshift bin, is shown in Figure 6.10. The procedure to compute $b_e^i(z)$ is described in Section 6.2.2.3 and is based on the work of Beutler and Dio (2020).

We estimated our error bars from the standard deviation of 100 log-normal mocks generated with the same characteristics of the original data: box dimensions, evolution and linear biases, selection function and survey mask. For this last step, we first generated the log-normal mocks for the whole box encompassing the different redshift bins of the light cone, and then applied the proper mask to select the specific angular region. However, because these error bars only quantify the variance of the estimator, the particular survey features do not matter much, and thus it should be possible to compute the variance of mocks inside the box, with the only caveat of following the mean number density of tracers to properly incorporate the shot noise. The impact of the window function in the standard deviation of the log-normal samples generated minor changes at very large scales, and we do not explore this further.

We first attempted to work with 29 bandpowers (Fourier bins) linearly spaced between $k_{\mathrm{min}} \approx 0.006$ and $k_{\mathrm{max}} \approx 0.157$ with $\Delta k \approx 0.005$. Apart from the amplitude of the monopole, not much change occurs between different redshifts, hence we only show the first z-bin here.

The binning described above resulted in oscillating error bars due to a poor sampling of modes for the average of the dipole, so we opted for a different binning: we considered 14 bandpowers linearly spaced between $k_{\mathrm{min}} \approx 0.006\ h/\mathrm{Mpc}$ and $k_{\mathrm{max}} \approx 0.157\ h/\mathrm{Mpc}$, with $\Delta k \approx 0.01\ h/\mathrm{Mpc} \approx 1.67\, k_{\mathrm{F}}$. However, in this case the window function convolution yielded nonphysical results, and we are still trying to understand what is the impact of the binning for the procedure described in Section 6.3.1.

We proceeded working with these 14 band powers and neglected the impact of the window function for the monopole. The reason for that was due to its impact being of the order of 5% or less at the scales we considered here, once we adopted the theoretical window function (see Section 6.3.1). Because its impact is larger for $k \lesssim 0.01\ h/\mathrm{Mpc}$, where the measurement is completely dominated by cosmic variance, we find it more instructive to compare the halo monopole with the one estimated from the real-space power spectrum of cold dark mat-

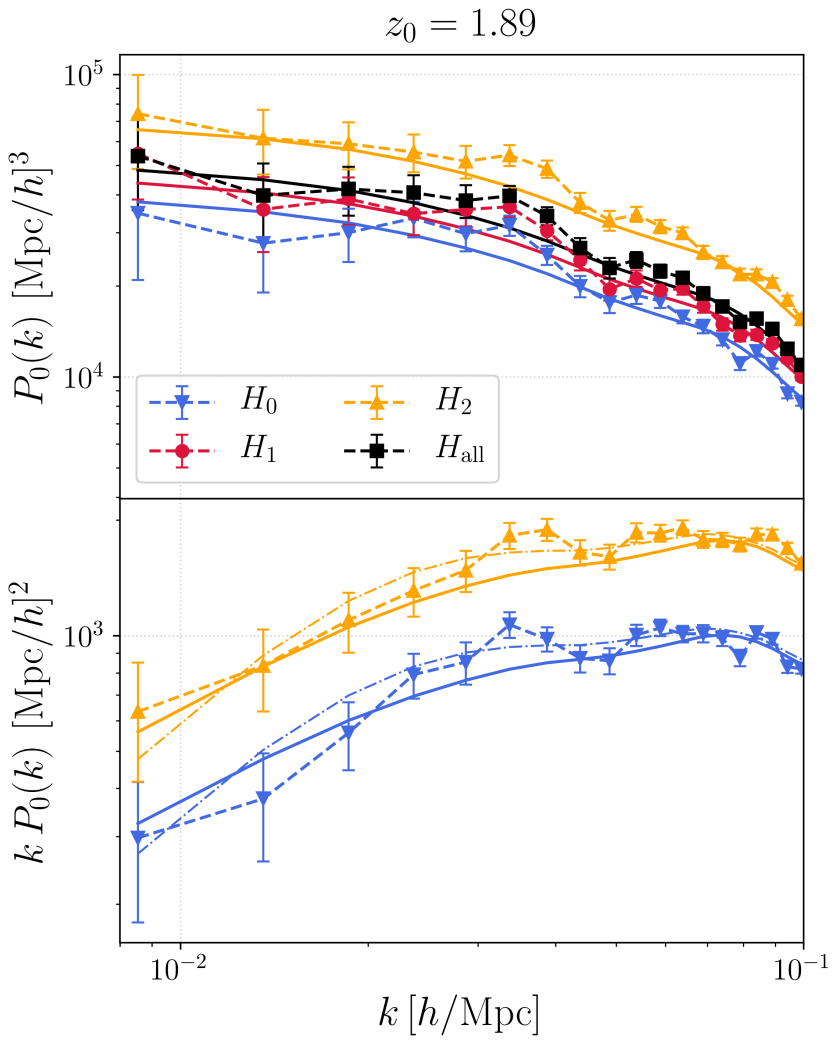


Figure 6.15: Monopole estimated from the four halo samples in the first redshift bin ($\bar{z} \approx 1.89$) using 29 bandpowers. Error bars are computed from the standard deviation of 100 log-normal mocks generated with the same box dimensions, evolution and linear biases, selection function and survey mask. Solid lines represent the convolution of the theory with the window function. The bottom panel shows $kP_0(k)$ for the less and more massive halo samples (lower and upper curve, respectively) for better visibility of the errors and the larger scales; it also shows the theoretical monopoles without considering the window function (dash-dotted lines).

ter (CDM) particles multiplied by the proper coefficient of the Legendre expansion, Equation (5.21). This is shown in Figure 6.16, for the four halo samples in the first redshift bin ($\bar{z} \approx 1.89$). This shows us that the theoretical connection between the CDM particles in real space is consistent with the halo measurements obtained from the redshift-space catalogue. This also has the advantage of naturally incorporating any window function effect one might be concerned with. For completeness, we also compare the measurements with the same monopole coefficient, but using the linear real-space power spectrum extracted from the CLASS Boltzmann solver (Blas et al., 2011). The bias parameters used for CLASS are presented in Table 6.1 and

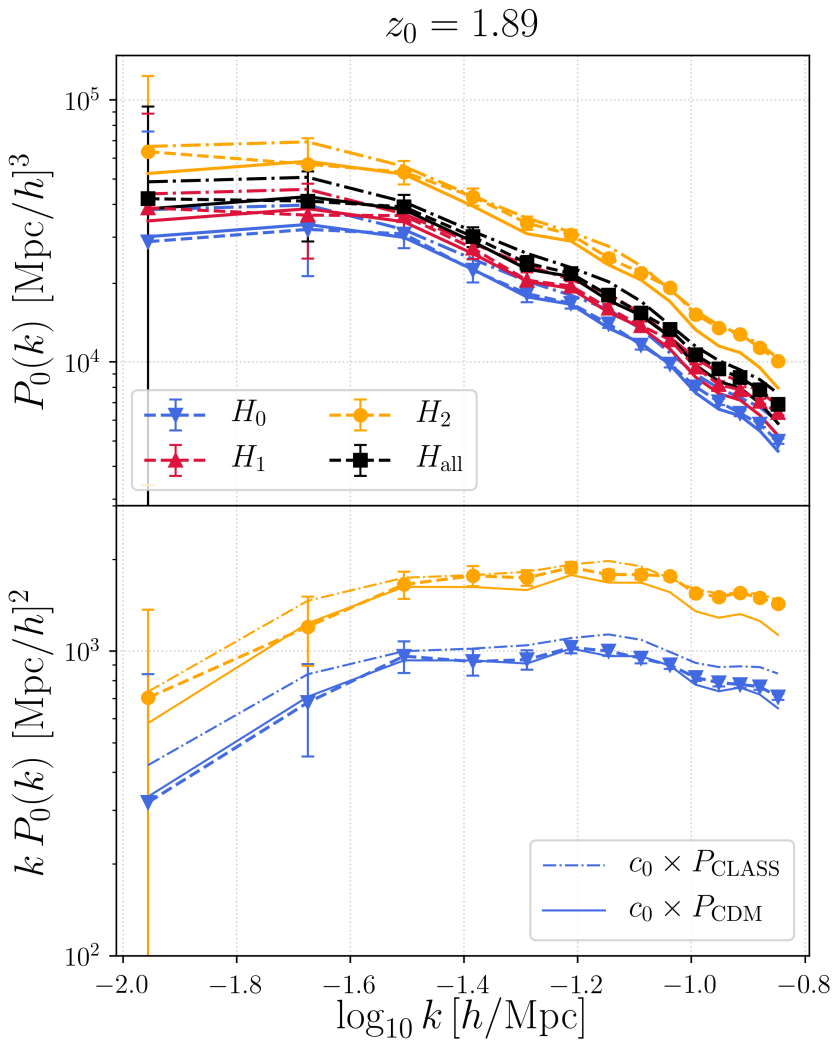


Figure 6.16: Monopole estimated from the four halo samples in the first redshift bin: $\bar{z} \approx 1.89$. Error bars are computed from the standard deviation of 100 log-normal mocks generated with the same box dimensions, evolution and linear biases, selection function and survey mask. Solid lines represent the redshift-space halo power spectrum computed from the real space CDM estimated spectrum. The bottom panel shows $kP_0(k)$ for the less and more massive halo samples (lower and upper curve, respectively) for better visibility of the errors and the larger scales; it also shows the theoretical monopoles without considering the window function (dash-dotted lines).

it is used throughout this work. To compute the monopole coefficient with the real-space CDM power spectrum, we consider the bias parameters shown in Table 6.2, fourth column (CDM).

To what concerns the variance of our measurements, the window function had a negligible impact. Nonetheless, to avoid any sort of complications, we explore the asymmetry of the relativistic signal as suggested in Beutler and Dio (2020): by computing

$$\Delta P_1 = P_1^{\alpha\beta} - P_1^{\beta\alpha} \tag{6.38}$$

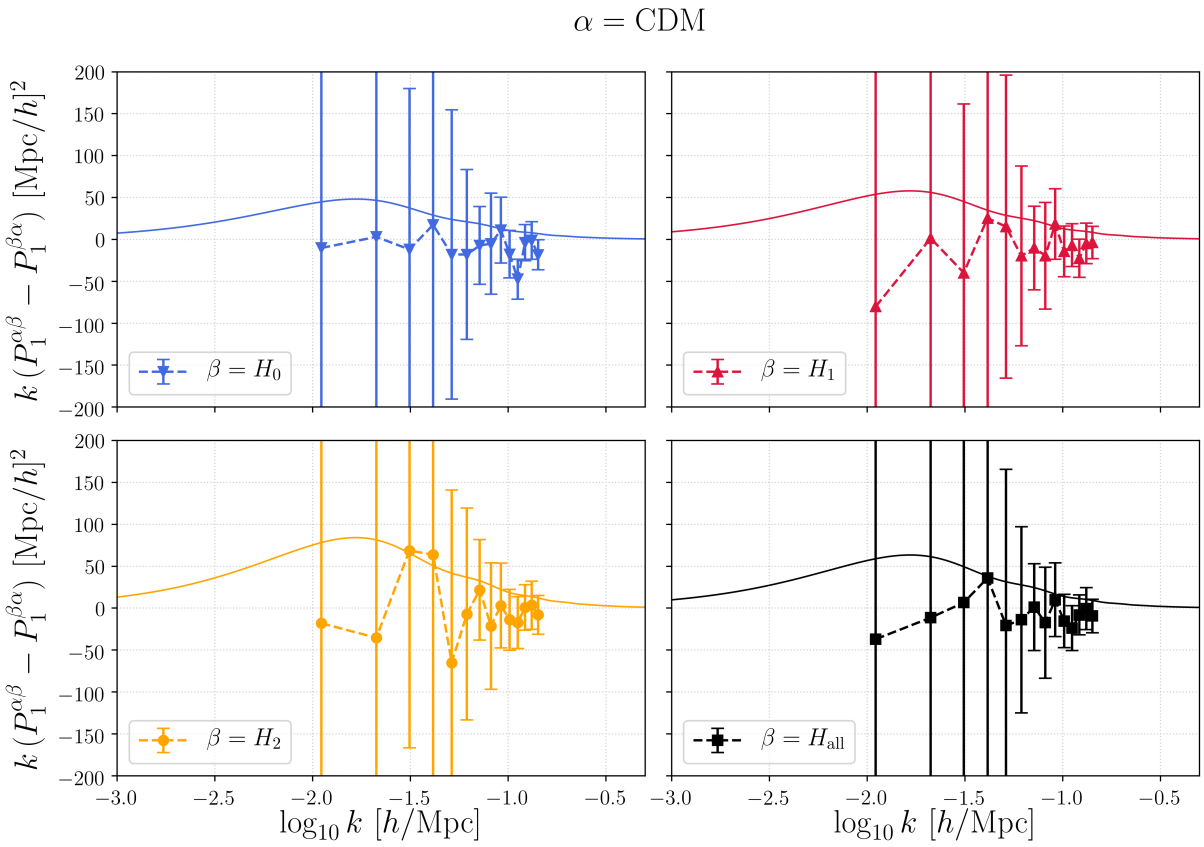


Figure 6.17: Dipole estimated from the four halo samples in the first redshift bin: $\bar{z} \approx 1.89$. Error bars are computed from the standard deviation of 100 log-normal mocks generated with the same box dimensions, evolution and linear biases, selection function and survey mask. Solid lines represent the theory, as $P_1^{\alpha\beta} - P_1^{\beta\alpha}$ is free from the window function contribution. We show all possible combinations of halos with cold dark matter.

it is possible to isolate the relativistic contribution and get rid of the impact of the window function, which is symmetric. As pointed out in Section 6.1 and as can be seen in Chapter 5, the Doppler term is antisymmetric, $\langle \delta_\alpha(\boldsymbol{k}) \delta^*_\beta(\boldsymbol{k}) \rangle = -\langle \delta_\beta(\boldsymbol{k}) \delta^*_\alpha(\boldsymbol{k}) \rangle$, and thus $\Delta P_1 \sim 2\langle \delta_\alpha(\boldsymbol{k}) \delta^*_\beta(\boldsymbol{k}) \rangle$. For this reason, it was not necessary the inclusion of the convolution effect described in Equation (6.36).

Consistently, the theoretical prediction for the last cross-correlation $H_2 \times H_{\rm all}$ is positive, as $\Delta b = b_2 - b_{\rm all} > 0$. However, as shown in Figure 6.18, for the cross-dipole among halos, and in Figure 6.17, for the cross-correlation between halos and CDM particles (the latter in redshift space as well), we conclude that no detection can be claimed with this pencil-beam light cone. Very similar results were obtained for the other two redshift samples $z_1 \approx 2.29$ and $z_2 \approx 2.69$.

Still, we point out the possibility of exploring optimal weighting schemes to enhance the signal in the light of the work carried out by Castorina et al. (2019). Lastly, the advantages of employing different tracers, coupled with the low densities of the catalogues, as well as the need for robust statistics, suggests the use of optimal weights for a more efficient combination

of tracers (Abramo et al., 2015; Montero-Dorta et al., 2020).

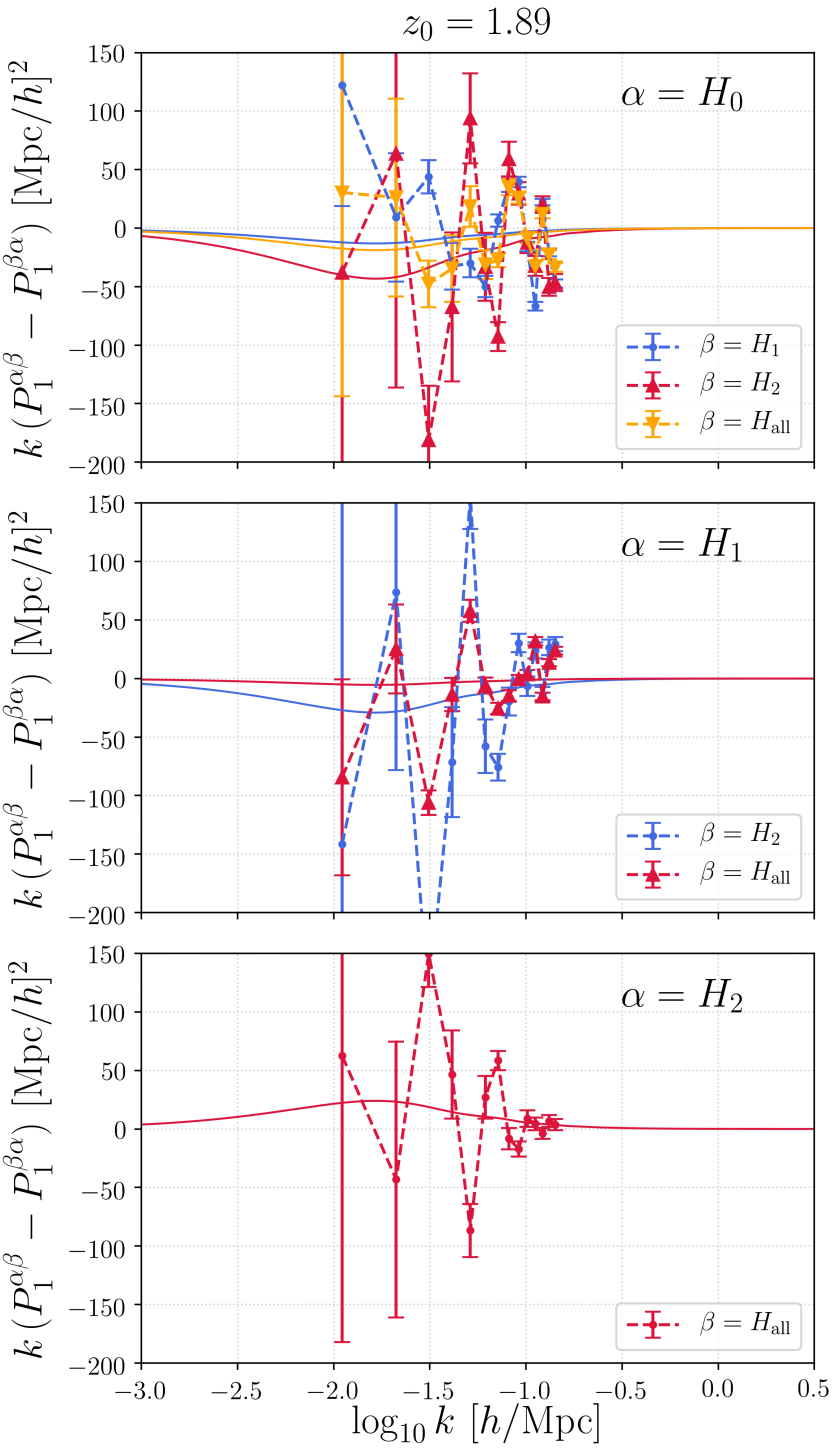


Figure 6.18: Dipole estimated from the four halo samples in the first redshift bin ($\bar{z} \approx 1.89$) with 14 bandpowers. Error bars are computed from the standard deviation of 100 log-normal mocks generated with the same box dimensions, evolution and linear biases, selection function and survey mask. Solid lines represent the theory, as $P_1^{\alpha\beta} - P_1^{\beta\alpha}$ is free from the window function contribution. We show all possible combinations of halos.

Chapter 7

Clustering redshifts with 21-cm bispectra

In this chapter we present a direct application of the multi-tracer strategy in the context of cross-correlating different tracers to reconstruct the redshift distribution of galaxies in a photometric survey. To achieve the reconstruction, we use maps of neutral hydrogen (HI) obtained from the intensity mapping technique, whose redshifts are known, as the second sample to be correlated with galaxies. Both the cross-correlation (McQuinn and White, 2013) and HI intensity mapping (Alonso et al., 2017) had already been used for the purposes of photometric redshift reconstruction; the novelty of our method is the use of the cross-bispectrum to reach scales that are in principle lost to foreground removal techniques from the mode coupling of short and long radial modes naturally incorporated in the third-order statistics. This work was devised by Dr. David Alonso, and will be submitted to the MNRAS (Guandalin et al., 2021). Both theoretical and numerical pieces have been cross-checked by the authors.

7.1 Motivation

As already mentioned in Chapter 2, the dawn of precision cosmology will be anchored by the upcoming surveys that promise very high-precision (spectroscopic surveys) redshifts, or very dense galaxy sample (photometric surveys). Joining the benefits of spectroscopic precision in a catalogue as dense as those provided by photometry is costly; nonetheless, as discussed in Section 2.4, there is another route when it comes to availing the existence of spectroscopic and photometric surveys: the use of cross-correlations between different probes (Bacon et al., 2015).

While spectroscopic surveys offer high-precision redshifts, their downside is the amount of time needed to get the spectrum of each surveyed object in order to obtain the redshifts. This results in a sparse sample that leads to a weaker signal. To circumvent this issue, one may carry out a photometric survey, where the spectra of objects are replaced by pseudo-spectra capable of estimating redshifts via strong spectral features. The byproduct of these surveys is a dense catalogue, containing millions (or even billions) of objects with photometric redshifts (photo-

zs), providing a larger signal-to-noise. However, the redshift distribution is unknown and its uncertainty will impact the summary statistics and, eventually, the inference of cosmological parameters (Alonso et al., 2017; Lee et al., 2021).

A route to diminish photo-z uncertainties and their potential impact on cosmological analyses is to cross-correlate the photometric sample with a spectroscopic one, via what is called clustering redshifts (Newman, 2008; McQuinn and White, 2013; Alonso et al., 2017; Cunnington et al., 2019b). Still, as photometric surveys start to cover a larger volume, matching the objects in both samples begins to be a difficult task due to the aforementioned limitations of spectroscopic surveys, and also due to the fact that spectroscopic and photometric surveys usually have a disjoint coverage sky area.

For the many purposes of large-scale analysis, though, resolving individual galaxies to obtain a detailed map at small scales, as is done in optical, spectroscopic or neutral hydrogen (HI) galaxy surveys[1], is not necessary. In this way, an HI intensity mapping survey, built from the integrated 21-cm emission line of many (unresolved) individual sources at a given angular scale, provides a fast way to map large volumes, at high redshifts, with spectroscopic precision. While this technique overcomes the weakness of the neutral hydrogen signal to detect one individual galaxy, delivering a larger signal-to-noise, it is prone to foreground contamination, which surpasses the strength of the cosmological signal by several orders of magnitude (Cunnington et al., 2019b). Therefore, it is of central importance in intensity mapping surveys to control and properly model the removal of such foregrounds. The effects of foregrounds in 21-cm IM surveys have been discussed in Section 2.3, to which we refer the reader for definitions.

On the other hand, although with HI intensity mapping the overlapping region with wide photometric surveys is increased, the foreground removal renders a limited range of scales accessible by the survey, which may not be optimal when it comes to employing the clustering redshifts approach to improve the photo-z calibration. For this reason, we present in this chapter a method to reconstruct the redshift distributions of photometric surveys by combining them with intensity mapping experiments at a three-point function level. As it will be discussed in Section 7.2, the clustering redshifts method performed with the power spectrum requires the overlap in k-space between galaxies and the HI map. However, as described in Chapter 2, the photometric survey act as a low-pass k-filter, retaining large radial scales only, whereas the region removed by the 21-cm foregrounds is at low-$k_{\parallel}$. Therefore, clustering redshifts with intensity mapping may be benefited by the coupling of large and small scales, information naturally accessible from higher-order statistics such as the bispectrum: we will combine a

[1] Although HI galaxy surveys have not been discussed in Section 2.2, the strategy is to detect the 21-cm line emitted by galaxies. These surveys will not provide the integrated emission of many galaxies, as done in the intensity mapping approach, but rather to resolve individual sources, as done in spectroscopic surveys.

galaxy sample, which retains the low-$k_\parallel$ information, with two hydrogen maps that, by the triangular inequality imposed by the bispectrum, will be restricted to high-$k_\parallel$ modes. Based on this idea, we forecast the ability of the cross-bispectra between galaxies and HI intensity maps to recover the redshift distribution of photometric surveys. The surveys' details are described in Sections 2.2.2 and 2.3, while the specifications utilised in this chapter are addressed in Section 7.3. However, for self-consistency of this chapter, we restate some concepts and definitions. Finally, the results are presented in Section 7.4.

7.2 Clustering redshift method

When the radial information (e.g. redshifts) is not completely available, as it is the case of photometric surveys (see Section 2.2.2), it is customary to project the cosmological field along the line-of-sight, and perform what is called a *tomographic* analysis. Therefore, let $\Delta_g(\boldsymbol{r}_\perp)$ be the projected galaxy overdensity on a given redshift bin related to the three-dimensional overdensity $\delta_g(r_\parallel, \boldsymbol{r}_\perp)$, in the flat-sky approximation, via

$$\Delta_g(\boldsymbol{r}_\perp) = \int \mathrm{d}r_\parallel \; \phi(r_\parallel)\, \delta_g(r_\parallel, \boldsymbol{r}_\perp), \tag{7.1}$$

where the radial selection function $\phi(r_\parallel)$ is nothing but the survey redshift distribution. In the presence of photo-z errors, $\phi(r_\parallel)$ can be interpreted as the probability to find a galaxy whose true (spectroscopic) redshift is z, in a given photometric redshift bin, as shown in Equation (2.5),

$$\phi_i(z) = n_g(z) \int_{z_{\rm ph}^i}^{z_{\rm ph}^{i+1}} \mathrm{d}z_{\rm ph} P(z_{\rm ph}|z), \tag{7.2}$$

where $n_g(z)$ is the number density of galaxies per unit redshift, $n(z) = \mathrm{d}N(z)/\mathrm{d}z$. Notice that, when $\sigma_z \to 0$, $P(z^{\rm ph}|z) \to \delta^D(z^{\rm ph} - z)$, and $\phi \to n_g(z)$. As it has been done in Section 2.2 (see discussion around Equation 2.1), we can write $n_g(r_\parallel) = H(z)\, c^{-1} n_g(z)$, and $\phi_i(r_\parallel) = H(z)\, c^{-1} \phi_i(z)$.

For analytical purposes, we will suppose that $P(z^{\rm ph}|z) \sim \mathcal{N}(\mu = z, \sigma = \sigma_z)$ is a Gaussian distribution centred around the spectroscopic redshift z (assuming that there are no sources of photo-z bias), with variance σ_z^2 (Equation 2.4):

$$P(z_{\rm ph}|z) = \frac{1}{\sigma_z \sqrt{2\pi}} \exp\left[-\frac{1}{2}\frac{(z_{\rm ph} - z)}{\sigma_z^2}\right]. \tag{7.3}$$

Therefore, the Fourier transform of $\phi_i(r_\parallel)$ is given by

$$\phi_i(k_\parallel) = \exp\left(-\frac{k_\parallel^2 \sigma_\parallel^2}{2}\right) \exp\left(-i k_\parallel r_\parallel^*\right), \tag{7.4}$$

where $\sigma_\parallel = c\,\sigma_z\,H^{-1}(z)$ is the photo-z error, along the radial direction, and $r^*_\parallel$ is the comoving distance corresponding to the centre of the photo-z bin. Notice that the presence of this photo-z error is analogous to the effects of redshift-space distortions (RSDs) and, therefore, erase small scale information in photometric redshifts, compromising the ability to use RSDs to constrain cosmological models.

Described for the first time in Newman (2008), and further explored in McQuinn and White (2013), the clustering redshifts approach aims at determining the form of ϕ from the cross-correlations of Δ_g with maps of a reference tracer δ_h with known redshift distributions. This is usually carried with spectroscopic samples. An intuitive way to think about this method is by considering the redshift distribution as the one depicted in Figure 2.2, in which the $\phi(r_\parallel)$ is essentially a histogram. We can thus parametrise ϕ as a linear combination of basis functions $w_a(r_\parallel)$:

$$\phi(r_\parallel) = \sum_a \phi_a\, w_a(r_\parallel), \tag{7.5}$$

where each bin of $\phi(r_\parallel)$ is weighted by ϕ_a, which is proportional to the number of galaxies in that bin. If there is no overlap between the spectroscopic and photometric samples, then ϕ_a vanishes. As desired, this cross-correlation is therefore able to determine the photometric-redshift distribution.

Here we explore this technique with $\delta_h(r_\parallel, \boldsymbol{r}_\perp)$ corresponding to the three-dimensional distribution of HI determined from intensity mapping measurements of the 21-cm line.

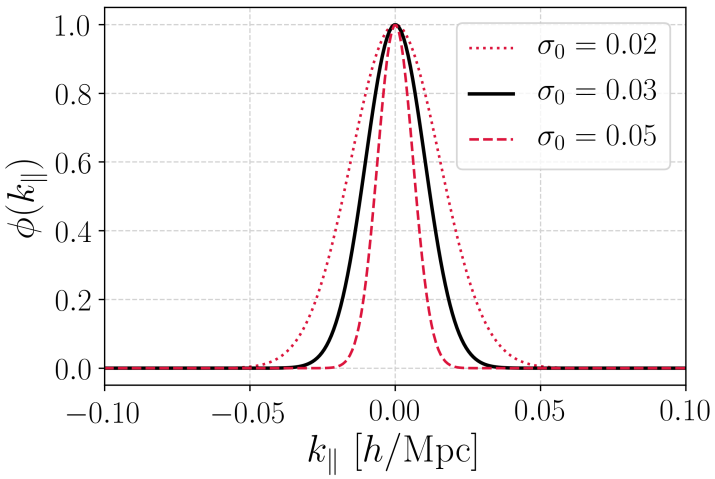


Figure 7.1: Fourier transform of the Gaussian photometric redshift distribution, Equation 7.4. The standard deviation of the Gaussian distribution, $\sigma_z = \sigma_0(1+z)$, is parametrised by σ_0, which is varied in the plot. The larger the photo-z scatter, the more information is contained in the large radial modes (small $k_\parallel$).

As can be seen in Figure 7.1, the distribution of photometric redshifts with a large scatter (i.e. large σ_z) is dominated by long wavelengths, i.e. small Fourier modes. Also, as already

mentioned in Section 2.3 (see discussion below Section 2.3.1), IM technique corresponds to a high-pass (Fourier) filter due to foreground contamination, which limits the capacity of cross-correlating IM with galaxy surveys to reconstruct distributions with a large scatter. In this work we take advantage of the non-linear evolution induced by gravitational collapse that propagates information from the long-wavelengths to the small scales. This coupling of scales allows us to recover the apparent loss of radial information from the small scales, and has been successfully employed to reconstruct the large scales of HI IM simulations Modi et al. (2019). For our purposes, we will use the bispectrum to access the modes lost to foregrounds in IM experiments and reconstruct the aforementioned photo-z distributions. Again, for simplicity we will rely on the flat-sky approximation.

With the expansion of Equation (7.5), the projected overdensity Δ_g can be written as a sum over "slices" δ_g^a:

$$\Delta_g(\boldsymbol{r}_\perp) = \sum_a \phi_a\, \delta_g^a(\boldsymbol{r}_\perp), \quad \delta_g^a(\boldsymbol{r}_\perp) \equiv \int \mathrm{d}r_\parallel\; w_a(r_\parallel)\, \delta_g(r_\parallel, \boldsymbol{r}_\perp). \tag{7.6}$$

Our aim is to quantify the accuracy with which the coefficients ϕ_a can be determined through cross-correlations. We follow the conventions of Section 1.2, carrying out our calculations in the transverse (spatial dimension $d = 2$) Fourier space.

Fourier decomposition We shall consider the Fourier basis for the expansion of Equation (7.5), in which w_a will be characterised by the radial wavenumber $k_\parallel$ as

$$w_{k_\parallel}(r_\parallel) = \exp(-i\, k_\parallel\, r_\parallel). \tag{7.7}$$

In the Fourier basis, the slices are the 3D Fourier coefficients of the galaxy overdensity $\delta_g^{k_\parallel}(\boldsymbol{k}_\perp) \equiv \delta_g(k_\parallel, \boldsymbol{k}_\perp)$, and the selection function coefficients $\phi_{k_\parallel}$ can be related to their radial Fourier transform via

$$\phi_{k_\parallel} = \frac{\Delta k_\parallel}{2\pi}\, \phi(k_\parallel) = \frac{1}{L}\, \phi(k_\parallel), \tag{7.8}$$

where $\Delta k_\parallel \equiv 2\pi/L$ is the Fourier-grid spacing.

For the redshift calibration, we will consider the 2- and 3-point correlators between the projected fields Δ_g, which we will label with a slice index T, and the HI-slices δ_h^a:

$$\begin{aligned} \langle \delta_x^a(\boldsymbol{k}_\perp)\, \delta_y^b(\boldsymbol{p}_\perp) \rangle &\equiv (2\pi)^2\, \delta^D(\boldsymbol{k}_\perp + \boldsymbol{p}_\perp)\, \mathcal{P}_{xy}^{ab}(k_\perp), \\ \langle \delta_x^a(\boldsymbol{k}_\perp)\, \delta_y^b(\boldsymbol{p}_\perp)\, \delta_z^c(\boldsymbol{q}_\perp) \rangle &\equiv (2\pi)^2\, \delta^D(\boldsymbol{k}_\perp + \boldsymbol{p}_\perp + \boldsymbol{q}_\perp)\, \mathcal{B}_{xyz}^{abc}(k_\perp, p_\perp, q_\perp), \end{aligned} \tag{7.9}$$

where $\mathcal{P}$ and $\mathcal{B}$ are the 2D power spectra and bispectra, respectively, and $x, y, z = \{g, h\}$ will label the type of tracer under consideration. The 3D counterparts of $\mathcal{P}$ and $\mathcal{B}$ are defined in

Equations (3.4) and (3.10), which we reproduce below:

$$\begin{aligned}\langle\delta_x(\boldsymbol{k})\delta_y(\boldsymbol{p})\rangle &\equiv (2\pi)^3\delta^D(\boldsymbol{k}+\boldsymbol{p})P_{xy}(k),\\ \langle\delta_x(\boldsymbol{k})\delta_y(\boldsymbol{p})\delta_z(\boldsymbol{q})\rangle &\equiv (2\pi)^3\delta^D(\boldsymbol{k}+\boldsymbol{p}+\boldsymbol{q})B_{xyz}(k,p,q).\end{aligned} \tag{7.10}$$

Above, $k = |\boldsymbol{k}|$ is the modulus of the 3D Fourier vector.

Using the Fourier radial decomposition, we can relate the different possible projected correlators, Equation (7.9), with P and B:

$$\begin{aligned}\mathcal{P}_{xy}^{k_\parallel p_\parallel}(k_\perp) &= L\ \delta^K_{k_\parallel+p_\parallel}\ P_{xy}(k),\\ \mathcal{P}_{gx}^{T\,p_\parallel}(k_\perp) &= \sum_{k_\parallel} L\ \delta^K_{k_\parallel+p_\parallel}\phi_{k_\parallel}P_{gx}(k),\\ \mathcal{P}_{gg}^{TT}(k_\perp) &= \sum_{k_\parallel p_\parallel} L\ \delta^K_{k_\parallel+p_\parallel}\phi_{k_\parallel}\phi_{p_\parallel}P_{gg}(k),\end{aligned} \tag{7.11}$$

and

$$\begin{aligned}\mathcal{B}_{xyz}^{k_\parallel p_\parallel q_\parallel}(k_\perp,p_\perp,q_\perp) &= L\ \delta^K_{k_\parallel+p_\parallel+q_\parallel}B_{xyz}(k,p,q),\\ \mathcal{B}_{gxy}^{T\,p_\parallel q_\parallel}(k_\perp,p_\perp,q_\perp) &= \sum_{k_\parallel} L\ \delta^K_{k_\parallel+p_\parallel+q_\parallel}\phi_{k_\parallel}B_{gxy}(k,p,q),\\ \mathcal{B}_{ggx}^{TT\,q_\parallel}(k_\perp,p_\perp,q_\perp) &= \sum_{k_\parallel p_\parallel} L\ \delta^K_{k_\parallel+p_\parallel+q_\parallel}\phi_{k_\parallel}\phi_{p_\parallel}B_{ggx}(k,p,q),\\ \mathcal{B}_{ggg}^{TTT}(k_\perp,p_\perp,q_\perp) &= \sum_{k_\parallel p_\parallel q_\parallel} L\ \delta^K_{k_\parallel+p_\parallel+q_\parallel}\phi_{k_\parallel}\phi_{p_\parallel}\phi_{q_\parallel}B_{ggg}(k,p,q).\end{aligned} \tag{7.12}$$

7.2.1 Fisher matrix

The Fisher matrix for a data vector made up of a set of power spectra $\mathcal{P}^{XY}(k_\perp)$ is given by:

$$F^P_{\alpha\beta} = \sum_{XX'}\frac{A}{4\pi}\int_0^\infty \mathrm{d}k_\perp\, k_\perp\, \partial_\alpha\mathcal{P}^{XY}(k_\perp)\,\partial_\beta\mathcal{P}^{X'Y'}(k_\perp)\,\mathcal{I}^{XX'}(k_\perp)\,\mathcal{I}^{YY'}(k_\perp), \tag{7.13}$$

where $\mathcal{I}^{XY}$ is the XY element of the inverse matrix of

$$\mathcal{P}^{XY} = \left(\begin{array}{c|c} \mathcal{P}_{gg}^{TT} & \mathcal{P}_{hg}^{p_\parallel T} \\ \hline \mathcal{P}_{gh}^{T\,p_\parallel} & \mathcal{P}_{hh}^{k_\parallel p_\parallel} \end{array}\right) \equiv \left(\begin{array}{c|c} \phi & v^\mathsf{T} \\ \hline v & M \end{array}\right), \tag{7.14}$$

that is,

$$\mathcal{I}^{XY} = \left(\begin{array}{c|c} \frac{1}{\phi - v^{\mathsf{T}} M^{-1} v} & -\frac{v^{\mathsf{T}} M^{-1}}{\phi - v^{\mathsf{T}} M^{-1} v} \\ \hline -\frac{M^{-1} v}{\phi - v^{\mathsf{T}} M^{-1} v} & M^{-1} + \frac{M^{-1} v v^{\mathsf{T}} M^{-1}}{\phi - v^{\mathsf{T}} M^{-1} v} \end{array}\right). \tag{7.15}$$

The ij-entry was obtained by employing the Woodbury matrix identity (Woodbury, 1950) to $\left(M - \phi^{-1} v\, v^{\mathsf{T}}\right)^{-1}$, and we made an analogy to the scalar-vector-tensor decomposition (Section B.2.1), calling the element $\mathcal{P}^{TT}_{gg} \equiv \phi$, the vector slices $\mathcal{P}^{T p_\parallel}_{gh} \equiv v$, and the diagonal matrix $\mathcal{P}^{k_\parallel p_\parallel}_{hh} \equiv M$. Finally, in Equation (7.13) $\boldsymbol{X} = \{X, Y\}$ runs over T and all possible HI slices, and A is the survey area.

Writing the matrix $\mathcal{I}^{XY}$ explicitly in terms of the power-spectrum elements, one obtains:

$$\mathcal{I}^{TT}(k_\perp) = \left\{ \sum_{k_\parallel} L\, |\phi_{k_\parallel}|^2 P_{gg}(k) \left[1 - \frac{(P_{gh}(k))^2}{P_{gg}(k) P_{hh}(k)} \right] \right\}^{-1}, \tag{7.16}$$

$$\mathcal{I}^{T k_\parallel}(k_\perp) = -\mathcal{I}^{TT}(k_\perp)\, \phi_{k_\parallel} \frac{P_{gh}(k)}{P_{hh}(k)}, \tag{7.17}$$

$$\mathcal{I}^{k_\parallel q_\parallel}(k_\perp) = \frac{\delta^K_{k_\parallel + q_\parallel}}{L\, P_{hh}(k)} + \mathcal{I}^{TT}(k_\perp)\, \phi_{k_\parallel} \phi_{q_\parallel} \frac{P_{gh}(k_\parallel, k_\perp)}{P_{hh}(k_\parallel, k_\perp)} \frac{P_{gh}(q_\parallel, k_\perp)}{P_{hh}(q_\parallel, k_\perp)} \simeq \frac{\delta^K_{k_\parallel + q_\parallel}}{L\, P_{hh}(k)}. \tag{7.18}$$

For the bispectra the Fisher matrix reads (Yadav et al., 2007; Chen et al., 2021):

$$F^B_{\alpha\beta} = \sum_{XX'} \frac{A}{4\pi} \frac{1}{6} \int_0^\infty \mathrm{d}k_\perp\, \mathrm{d}p_\perp\, \mathrm{d}q_\perp\, \frac{k_\perp\, p_\perp\, q_\perp}{\pi^2 A_T}\, \partial_\alpha \mathcal{B}^{XYZ}(k_\perp, p_\perp, q_\perp)\, \partial_\beta \mathcal{B}^{X'Y'Z'}(k_\perp, p_\perp, q_\perp) \times \mathcal{I}^{XX'}(k_\perp)\, \mathcal{I}^{YY'}(p_\perp)\, \mathcal{I}^{ZZ'}(q_\perp), \tag{7.19}$$

where $A_T = \frac{1}{2}\sqrt{2k_\perp^2 p_\perp^2 + 2p_\perp^2 q_\perp^2 + 2q_\perp^2 k_\perp^2 - k_\perp^4 - p_\perp^4 - q_\perp^4}$ is the area of a triangle with sides $(k_\perp, p_\perp, q_\perp)$, and now $\boldsymbol{X} = \{X, Y, Z\}$.

In our case, the only parameters we will consider are the selection function coefficients $\phi_{k_\parallel}$, and we choose to denote $\partial_{k_\parallel} \equiv \partial / \partial \phi_{k_\parallel}$. Thus, from Equations (7.11) and (7.12), the only non-vanishing derivatives are:

$$\begin{aligned}
\partial_{p_\parallel} \mathcal{P}^{T q_\parallel}_{gh}(k_\perp) &= L\, \delta^K_{p_\parallel + q_\parallel} P_{gh}(p_\parallel, k_\perp), \\
\partial_{p_\parallel} \mathcal{P}^{TT}_{gg}(k_\perp) &= 2L\, \phi_{-p_\parallel} P_{gg}(p_\parallel, k_\perp), \\
\partial_{l_\parallel} \mathcal{B}^{T p_\parallel q_\parallel}_{ghh}(k_\perp, p_\perp, q_\perp) &= L\, \delta^K_{l_\parallel + p_\parallel + q_\parallel} B_{ghh}(l_\parallel, p_\parallel, q_\parallel; \Delta_\perp), \\
\partial_{l_\parallel} \mathcal{B}^{TT q_\parallel}_{ggh}(k_\perp, p_\perp, q_\perp) &= L\, \phi_{-l_\parallel - q_\parallel} \left[B_{ggh}(l_\parallel, -l_\parallel - q_\parallel, q_\parallel; \Delta_\perp) + B_{ggh}(-l_\parallel - q_\parallel, l_\parallel, q_\parallel; \Delta_\perp) \right], \\
\partial_{l_\parallel} \mathcal{B}^{TTT}_{ggh}(k_\perp, p_\perp, q_\perp) &= \sum_{q_\parallel} L\, \phi_{q_\parallel} \phi_{-l_\parallel - q_\parallel} \left[B_{ggg}(l_\parallel, -l_\parallel - q_\parallel, q_\parallel; \Delta_\perp) + B_{ggg}(q_\parallel, l_\parallel, -l_\parallel - q_\parallel; \Delta_\perp) + \right.
\end{aligned}$$

$$B_{ggg}(-l_\parallel - q_\parallel, q_\parallel, l_\parallel; \Delta_\perp)\Big],$$

since $\partial_{p_\parallel}\mathcal{P}_{hh}^{k_\parallel q_\parallel}(k_\perp) = 0$ and $\partial_{l_\parallel}\mathcal{B}_{hhh}^{k_\parallel p_\parallel q_\parallel}(k_\perp, p_\perp, q_\perp) = 0$ as the autocorrelations of HI maps do not depend on $\phi_{k_\parallel}$. Above, we defined the short-hand notation $\Delta_\perp \equiv (k_\perp, p_\perp, q_\perp)$.

Power spectrum Fisher matrix The power spectrum Fisher matrix, Equation (7.13), can be decomposed into the following four non-zero contributions:

$$F^P_{k_\parallel q_\parallel} = F^{2T2T}_{k_\parallel q_\parallel} + 4F^{2T1T}_{(k_\parallel q_\parallel)} + 2F^{1T1T+}_{k_\parallel q_\parallel} + 2F^{1T1T\times}_{k_\parallel q_\parallel}, \tag{7.20}$$

where $4F^{2T1T}_{(k_\parallel q_\parallel)} = 2F^{2T1T}_{k_\parallel q_\parallel} + 2F^{2T1T}_{q_\parallel k_\parallel}$, and

$$F^{2T2T}_{k_\parallel q_\parallel} = \frac{A}{4\pi}\int \mathrm{d}k_\perp\, k_\perp\, \partial_{k_\parallel}\mathcal{P}^{TT}_{gg}(k_\perp)\partial_{q_\parallel}\mathcal{P}^{TT}_{gg}(k_\perp)\left[\mathcal{I}^{TT}(k_\perp)\right]^2 \tag{7.21}$$

$$= \frac{A}{4\pi}4L^2\phi_{-k_\parallel}\phi_{-q_\parallel}\int \mathrm{d}k_\perp\, k_\perp\, \left[P_{gg}(k_\parallel, k_\perp)\mathcal{I}^{TT}(k_\perp)\right]\left[P_{gg}(q_\parallel, k_\perp)\mathcal{I}^{TT}(k_\perp)\right] \tag{7.22}$$

$$= \frac{A}{\pi}\phi(-k_\parallel)\phi(-q_\parallel)\int \mathrm{d}k_\perp\, k_\perp\, \mathcal{R}_{gg}(k_\parallel, k_\perp)\,\mathcal{R}_{gg}(q_\parallel, k_\perp) \tag{7.23}$$

$$F^{2T1T}_{k_\parallel q_\parallel} = \frac{A}{4\pi}\sum_{p_\parallel}\int \mathrm{d}k_\perp\, k_\perp\, \partial_{k_\parallel}\mathcal{P}^{TT}_{gg}(k_\perp)\partial_{q_\parallel}\mathcal{P}^{Tp_\parallel}_{gh}(k_\perp)\mathcal{I}^{TT}(k_\perp)\mathcal{I}^{Tp_\parallel}(k_\perp) \tag{7.24}$$

$$= -\frac{A}{4\pi}2L^2\phi_{-k_\parallel}\phi_{-q_\parallel}\int \mathrm{d}k_\perp\, k_\perp\, \mathcal{R}_{gg}(k_\parallel, k_\perp)\mathcal{R}^2_{gh}(q_\parallel, k_\perp) \tag{7.25}$$

$$F^{1T1T+}_{k_\parallel q_\parallel} = \frac{A}{4\pi}\sum_{p_\parallel l_\parallel}\int \mathrm{d}k_\perp\, k_\perp\, \partial_{k_\parallel}\mathcal{P}^{Tp_\parallel}_{gh}(k_\perp)\partial_{q_\parallel}\mathcal{P}^{Tl_\parallel}_{gh}(k_\perp)\mathcal{I}^{TT}(k_\perp)\mathcal{I}^{p_\parallel l_\parallel}(k_\perp) \tag{7.26}$$

$$= \frac{AL}{4\pi}\delta^K_{k_\parallel+q_\parallel}\int \mathrm{d}k_\perp\, k_\perp \mathcal{R}^2_{gh}(k_\parallel, k_\perp) + \frac{A}{4\pi}L^2\phi_{-k_\parallel}\phi_{-q_\parallel}\int \mathrm{d}k_\perp\, k_\perp \mathcal{R}^2_{gh}(k_\parallel, k_\perp)\,\mathcal{R}^2_{gh}(q_\parallel, k_\perp), \tag{7.27}$$

$$F^{1T1T\times}_{k_\parallel q_\parallel} = \frac{A}{4\pi}\sum_{p_\parallel l_\parallel}\int \mathrm{d}k_\perp\, k_\perp\, \partial_{k_\parallel}\mathcal{P}^{Tp_\parallel}_{gh}(k_\perp)\partial_{q_\parallel}\mathcal{P}^{Tl_\parallel}_{gh}(k_\perp)\mathcal{I}^{Tp_\parallel}(k_\perp)\mathcal{I}^{Tl_\parallel}(k_\perp) \tag{7.28}$$

$$= \frac{A}{4\pi}L^2\phi_{-k_\parallel}\phi_{-q_\parallel}\int \mathrm{d}k_\perp\, k_\perp\, \mathcal{R}^2_{gh}(k_\parallel, k_\perp)\mathcal{R}^2_{gh}(q_\parallel, k_\perp) \tag{7.29}$$

where

$$\mathcal{R}_{gg}(k_\parallel, k_\perp) \equiv P_{gg}(k_\parallel, k_\perp)\mathcal{I}^{TT}(k_\perp), \quad \mathcal{R}^2_{gh}(k_\parallel, k_\perp) \equiv \frac{P^2_{gh}(k_\parallel, k_\perp)}{P_{hh}(k_\parallel, k_\perp)}\mathcal{I}^{TT}(k_\perp). \tag{7.30}$$

Joining all the terms together:

$$F^P_{k_\parallel q_\parallel} = \frac{AL^2}{2\pi}\left(\frac{\delta^K_{k_\parallel+q_\parallel}}{L}\int \mathrm{d}k_\perp\, k_\perp \mathcal{R}^2_{gh}(k_\parallel, k_\perp) + \right.$$
$$\left. 2\,\phi_{-k_\parallel}\phi_{-q_\parallel}\int \mathrm{d}k_\perp\, k_\perp \left[\mathcal{R}_{gg}(k_\parallel, k_\perp) + \mathcal{R}^2_{gh}(k_\parallel, k_\perp)\right]\left[\mathcal{R}_{gg}(q_\parallel, k_\perp) + \mathcal{R}^2_{gh}(q_\parallel, k_\perp)\right]\right) \tag{7.31}$$

$$\simeq \delta^K_{k_\parallel+q_\parallel}\frac{AL}{2\pi}\int \mathrm{d}k_\perp\, k_\perp\, \mathcal{I}^{TT}(k_\perp)\frac{P^2_{gh}(k_\parallel,k_\perp)}{P_{hh}(k_\parallel,k_\perp)}, \tag{7.32}$$

where the second equality holds to lowest order in P_{gh}/P_{hh} and neglecting any information coming from $\mathcal{P}^{TT}_{gg}(k_\perp)$.

As one can see in Equation (7.32), the clustering redshifts approach with the power spectrum requires an overlap in k-space through $P_{gh}(k_\parallel,k_\perp)$. As already discussed, when combining photometric galaxies with 21-cm intensity mapping the overlapping region is highly limited. Therefore, in what follows, we combine the maps at a three-point function (bispectrum) level.

The bispectrum Fisher matrix The bispectrum Fisher matrix can be decomposed into the following non-zero contributions:

$$F^B_{k_\parallel q_\parallel} = F^{3T3T}_{k_\parallel q_\parallel} + 6F^{3T2T}_{(k_\parallel,q_\parallel)} + 6F^{3T1T}_{(k_\parallel,q_\parallel)} + 3F^{2T2T+}_{k_\parallel q_\parallel} + 6F^{2T2T\times}_{k_\parallel q_\parallel} + 12F^{2T1T+}_{(k_\parallel,q_\parallel)} + 6F^{2T1T\times}_{(k_\parallel,q_\parallel)} + 3F^{1T1T+}_{k_\parallel q_\parallel} + 6F^{1T1T\times}_{k_\parallel q_\parallel}, \tag{7.33}$$

where $\lambda F_{(k_\parallel,q_\parallel)} = (\lambda/2)\,F_{k_\parallel,q_\parallel} + (\lambda/2)\,F_{q_\parallel,k_\parallel}$. We leave a detailed explanation to Appendix (D.3), reproducing only the final expression for the Fisher matrix, at lowest order in P_{gh}/P_{hh}, and neglecting any information from B^{TTT} or $B^{TTk_\parallel}$. With these approximations, which are properly justified in Section 7.4.1, the only non-zero component contributing to the Fisher matrix $F^B_{k_\parallel q_\parallel}$ is $F^{1T1T+}_{k_\parallel q_\parallel}$:

$$F^B_{k_\parallel p_\parallel} \simeq \delta^K_{k_\parallel+p_\parallel}\frac{AL}{4\pi}\frac{1}{6}\int \mathrm{d}k_\perp \mathrm{d}p_\perp \mathrm{d}q_\perp\, \frac{k_\perp p_\perp q_\perp}{\pi^2 A_T}\mathcal{I}^{TT}(k_\perp)\int\frac{\mathrm{d}p_\parallel}{2\pi}\frac{B^2_{ghh}(k_\parallel,p_\parallel,-k_\parallel-p_\parallel,\Delta_\perp)}{P_{hh}(p_\parallel,q_\perp)P_{hh}(-k_\parallel-p_\parallel,p_\perp)}. \tag{7.34}$$

Notice now that, as opposed to Equation (7.32), the $k_\parallel$ modes from the photometric sample can be limited to large radial scales, which are those not damped by the photo-z errors, while the available 21-cm modes, $p_\parallel$ and $q_\parallel$, are restricted to small scales through the triangular condition: $q_\parallel = -k_\parallel - p_\parallel$. That is, the apparent loss of large-scale information from the foreground removal process is regained from the coupling induced by gravitational collapse, which we are able access through the photometric galaxy sample.

7.3 Forecast specifications

As explained in Chapter 3, when dealing with the bispectrum, one must consider the second-order bias b_2 induced by the non-linear clustering. At large scales, we can consider the local (in matter density) bias expansion for the overdensity field of some tracer x (Fry and Gaztañaga, 1993),

$$\delta_x(\boldsymbol{x}) = b^{(1)}_x\delta(\boldsymbol{x}) + \frac{b^{(2)}_x}{2}\delta^2(\boldsymbol{x}) + \cdots, \tag{7.35}$$

and the evolved matter density contrast, up to second order in perturbation theory, $\delta(\boldsymbol{k}) = \delta^{(1)}(\boldsymbol{k}) + \delta^{(2)}(\boldsymbol{k}) + \cdots$, to write the cross-bispectrum of any biased tracer as

$$B_{xyz}(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3) = b_x^{(1)} b_y^{(1)} b_z^{(1)} B(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3) + \\ b_x^{(1)} b_y^{(1)} b_z^{(2)} P(k_1)\, P(k_2) + b_x^{(1)} b_y^{(2)} b_z^{(1)} P(k_1)\, P(k_3) + b_x^{(2)} b_y^{(1)} b_z^{(1)} P(k_2)\, P(k_3), \tag{7.36}$$

where $P(k)$ is the linear matter power spectrum, and $B(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3)$ is the matter bispectrum, driven by the non-linear gravitational dynamics in the dark matter field. In the remaining of this work, we will adopt the fitting formula proposed in Gil-Marín et al. (2012), which is performed within ΛCDM N-body simulations for cold dark matter:

$$B(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3) = 2\, F_2^{\rm eff}(\boldsymbol{k}_1, \boldsymbol{k}_2)\, P_{\rm nl}(\boldsymbol{k}_1)\, P_{\rm nl}(\boldsymbol{k}_2) + \text{ perms}, \tag{7.37}$$

with

$$F_2^{\rm eff}(\boldsymbol{k}_i, \boldsymbol{k}_j) = \frac{5}{7}\, \tilde{a}(k_i)\, \tilde{a}(k_j) + \frac{1}{2} \cos(\theta_{ij}) \left(\frac{k_i}{k_j} + \frac{k_j}{k_i} \right) \tilde{b}(k_i)\, \tilde{b}(k_j) + \frac{2}{7} \cos^2(\theta_{ij})\, \tilde{c}(k_i)\, \tilde{c}(k_j) \tag{7.38}$$

recovering the tree-level result in the limit of $\tilde{a}, \tilde{b}, \tilde{c} \to 1$, with the replacement of the non-linear matter power spectrum $P_{\rm nl}(\boldsymbol{k})$ by the linear one (see Equation 3.15). In the above equation,

$$\theta_{ij} = \frac{k_k^2 - k_i^2 - k_j^2}{2\, k_i\, k_j} \tag{7.39}$$

is the angle between the vectors $\boldsymbol{k}_i$, and $\boldsymbol{k}_j$, as they form a triangle with the third wavemode $\boldsymbol{k}_k$.

The functions $\tilde{a}, \tilde{b}$ and $\tilde{c}$ are given by

$$\begin{aligned} \tilde{a}(k) &= \frac{1 + \sigma_8^{a_6}(z)\, [0.7\, Q_3(n)]^{1/2}\, (q\, a_1)^{n+a_2}}{1 + (q\, a_1)^{n+a_2}}, \\ \tilde{b}(k) &= \frac{1 + 0.2\, a_3\, (n+3)(q\, a_7)^{n+3+a_8}}{1 + (q\, a_7)^{n+3.5+a_8}}, \\ \tilde{c}(k) &= \frac{1 + 4.5\, a_4\, (q\, a_5)^{n+3+a_9} \left[1.5 + (n+3)^4\right]^{-1}}{1 + (q\, a_5)^{n+3.5+a_9}}, \end{aligned} \tag{7.40}$$

where

$$\begin{aligned} a1 &= 0.484, \quad & a4 &= 0.392, \quad & a7 &= 0.128, \\ a2 &= 3.740, \quad & a5 &= 1.013, \quad & a8 &= -0.722, \\ a3 &= -0.849, \quad & a6 &= -0.575, \quad & a9 &= -0.926, \end{aligned}$$

are the best-fitting parameters of Equation (7.40) for the simulations[2] considered in Gil-Marín et al. (2012), $\sigma_8(z)$ is the amplitude of the linear density fluctuations[3], as a function of redshift, on a scale of 8 Mpc/h, $q \equiv k/k_{\rm nl}$, with $k_{\rm nl}$ such that

$$\frac{k_{\rm nl}^3 P_{\rm lin}(k_{\rm nl})}{2\pi^2} = 1, \tag{7.41}$$

the function $Q_3(n)$ is given by

$$Q_3(n) = \frac{4 - 2^n}{1 + 2^{n+1}}, \tag{7.42}$$

and

$$n \equiv \frac{{\rm d}\log P_{\rm lin}(k)}{{\rm d}\log k} \tag{7.43}$$

is the slope of the linear power spectrum $P_{\rm lin}(k)$ after the smoothing described in Figure 7.2.

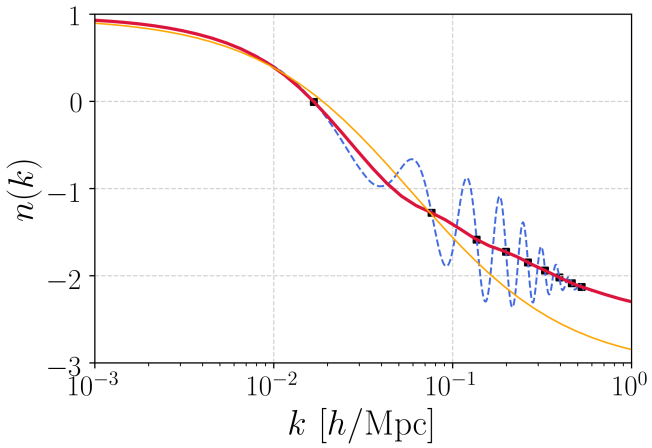


Figure 7.2: Slope of the linear power spectrum computed as in Equation (7.43) (*blue-dashed*), by interpolating the middle-of-the-amplitude splines (black dots) of the wiggles (*thick-red*), and fitting a smooth curve that approximates the damping of the red curve (*orange*).

The smoothing process was designed to improve the BAO wiggles in the bispectrum fit. As shown in Figure 7.3, using the slope without smoothing yields the blue-dashed curves in both plots, for a fixed redshift $z = 0.1$: although we get $\sim$ 20% differences with respect to the spline smoothing slope (red curve of Figure 7.2) for the $B(k_1, 2k_1, \theta)$ configuration (left panel of Figure 7.3), the equilateral configuration (on the right) gets nonphysical wiggles, and a large discrepancy that goes beyond 50% for some range of wavemodes. Using a simple smooth function (thin orange curve) that has an approximate behaviour of what is shown in red, in Figure 7.2, gives a $\approx$ 10% difference in the bispectrum amplitude for the $B(k_1, 2k_1, \theta)$

[2]The fit is obtained in the k-range $0.03 \leq k \leq 0.4$ [h/Mpc] and for redshifts in the range $0 \leq z \leq 1.5$.
[3]This can be defined, for example, as a top-hat region of radius $R = 8\,{\rm Mpc}/h$ in configuration space.

configuration, but it still has a large mismatch for the amplitude around $0.1 \lesssim k \lesssim 1$ h/Mpc for the equilateral configuration.

For the "wiggled" slope of Equation (7.43), the differences become smaller as we go to larger redshifts, while the generic smooth curve remains with a large mismatch in amplitude for the equilateral configuration (Figure 7.4). Notice that overestimating the bispectrum amplitude misleads the results, as it yields better constraints than what one would actually obtain. Therefore, we consider the smoothed slope (red curve in Figure 7.2), which results in the red bispectra of Figures 7.3 and 7.4.

Finally, non-linearities in the bias will be modelled via the phenomenological fit proposed by Lazeyras et al. (2016), which relates $b^{(2)} \equiv b_2$ with the linear bias $b^{(1)} \equiv b_1$ for dark matter halos in a flat ΛCDM universe as

$$b_2(b_1) = 0.412 - 2.143\, b_1 + 0.929\, b_1^2 + 0.008\, b_1^3. \tag{7.44}$$

Because we are interested in assessing the impact of a second-order bias in our constraints, we shall assume Equation (7.44) as a rough approximation for both of our tracers (galaxies and neutral hydrogen). Nonetheless, the implementation of any other relation between b_1 and b_2 is straightforward in our formalism.

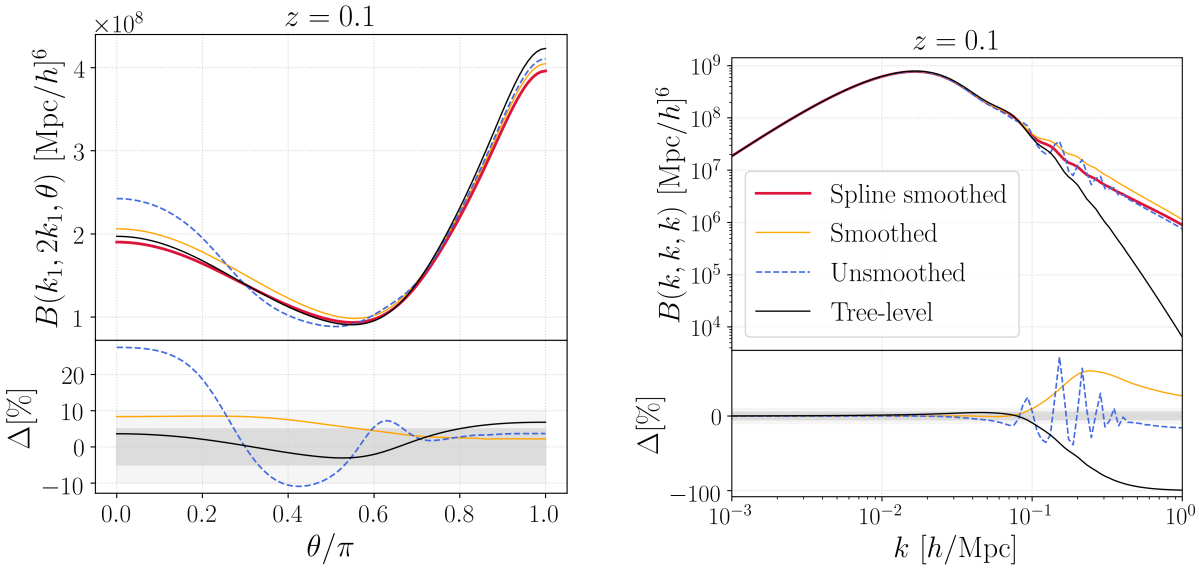


Figure 7.3: Real-space matter bispectrum, at redshift $z = 0.1$, for two configurations (*left*: $k_2 = 2k_1$, with the third wavemode k_3 parametrised by the angle θ between k_1 and k_2; *right*: equilateral configuration, $k_1 = k_2 = k_3 = k$) computed with the fit proposed in Gil-Marín et al. (2012) considering the "wiggled" power-spectrum slope (*blue-dashed*), Equation (7.2), the spline-smoothed slope (*red*), red curve of Figure 7.2, and the simple fitting function (*thin-orange*), also shown in orange in Figure 7.2. We show the tree-level bispectra (*thin-black*) for completeness. The bottom panels show the percentage difference between the blue-dashed and thin-orange bispectra with respect to the red one. Light and darker-grey shaded regions represent the 10% and 5% regions, respectively.

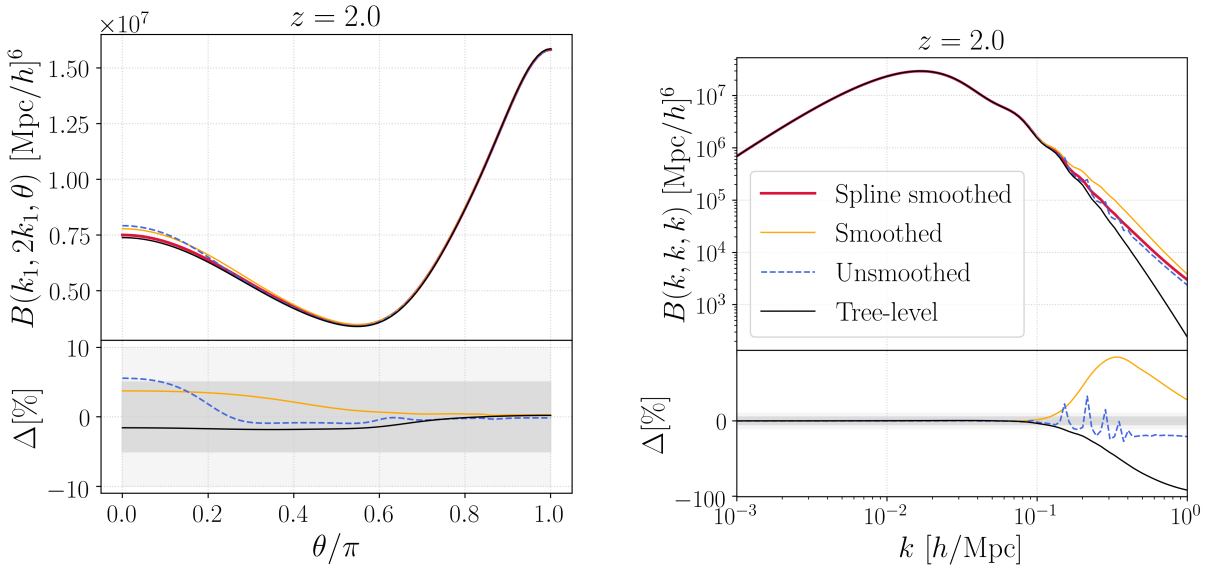


Figure 7.4: Same as Figure 7.3, but at redshift $z = 2$. As expected, the differences become smaller between the different bispectra (*left*), and the scales in which the tree-level result starts to break down are pushed to higher k-values (*right*).

7.3.1 Photometric galaxy survey

We focus on the LSST-like survey described in Chapter 2; more specifically, the redshift distribution $n_g(z)$ and the bias $b_g(z) = 0.95/D(z)$ are described in Section 2.2, with the parameters given in Table 2.2.

Power spectrum We will consider a simplistic model[4] for the 3D auto-spectrum of galaxies, namely

$$P_{gg}(k, z) = b_g^2(z) P(|\boldsymbol{k}|, z) + \bar{n}_g^{-1}(z), \tag{7.45}$$

where $P(|\boldsymbol{k}|, z)$ is the matter power spectrum at some redshift z. For our survey specifications, we show Equation (7.45) in Figure 7.5 centred at two different redshifts, for both the linear and non-linear (`halofit`) matter spectra.

Bispectrum The bispectrum of galaxies will be modelled as

$$B_{ggg}(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3) = b_{g,1}^3 B(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3) + b_{g,1}^2 b_{g,2} \left[P(k_1)P(k_2) + \text{perms}\right] + SN, \tag{7.46}$$

[4]We leave a thorougher analysis, including redshift-space distortions and higher-order biases, for example, to future work.

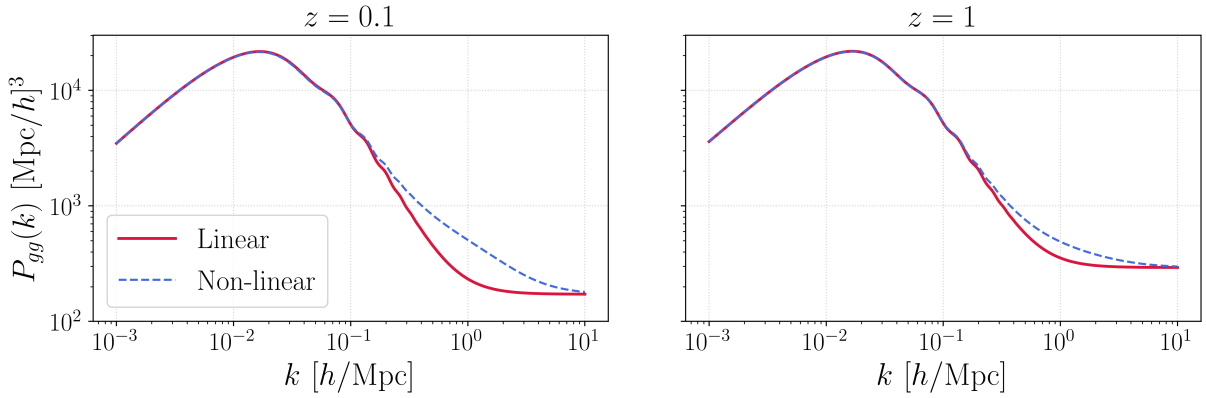


Figure 7.5: Linear (*red-solid*) and non-linear (*blue-dashed*) galaxy auto-spectrum at two different redshifts, $z = 0.1$ (*left*) and $z = 1$ (*right*). Shot-noise contribution is included.

where $B(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3)$ is the Gil-Marín et al. (2012) fit, Equation (7.37), $b_{g,1}$ and $b_{g,2}$ are the linear and second-order biases, Equations (2.2) and (7.44), respectively, and SN is the shot-noise contribution that we will omit in what follows, as the galaxies will contribute only to one "leg" in the cross-bispectrum, i.e. $B_{xyz} \to B_{ghh}$. Hence, bispectrum shot noise will not be relevant for our purposes.

Neglecting the shot-noise contribution, we show Equation (7.46) in Figure 7.6 for two redshifts, and a fixed configuration, including and neglecting the second-order bias contribution for completeness.

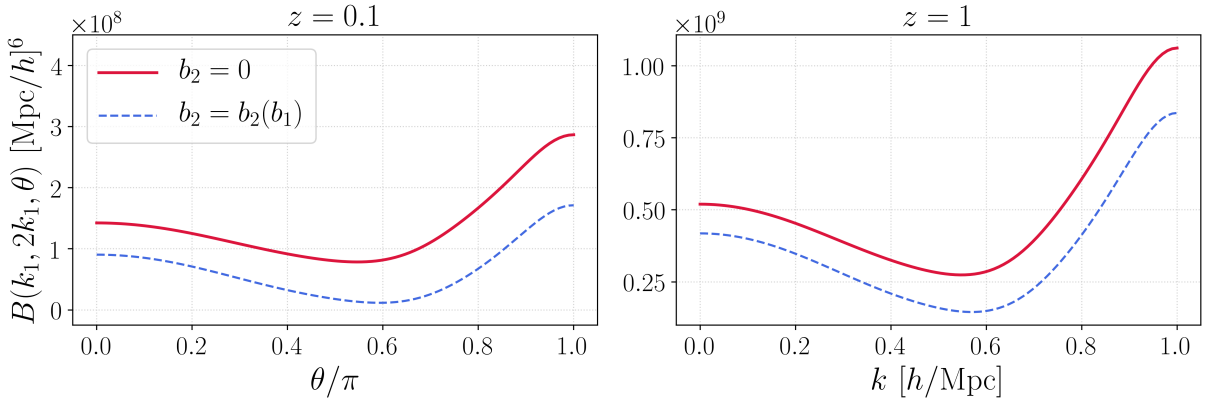


Figure 7.6: Auto-bispectrum of galaxies for $k_2 = 2k_1$, at redshifts $z = 0.1$ (*left*), and $z = 1$ (*right*), neglecting the shot-noise contribution. When the second-order bias contribution is accounted for (*blue-dashed*), the amplitude (and hence the signal) of the bispectrum gets decreased for this particular setting.

7.3.2 HI intensity mapping

As discussed in Section 2.3, intensity mapping of the HI signal maps the fluctuations δT_h of the brightness temperature T_h around its homogeneous component $\bar{T}_h$,

$$\delta T_h(\boldsymbol{x}) \equiv \bar{T}_h(z)\delta_h(\boldsymbol{x}) \equiv \bar{T}_h(z)\left[b_{h,1}(z)\delta(\boldsymbol{x}) + \frac{b_{h,2}(z)}{2}\delta^2(\boldsymbol{x})\right]. \tag{7.47}$$

Once we account for the systematic effects of beam smoothing and foreground damping, the observed temperature fluctuation is given by Equation (2.16), which we replicate below:

$$\delta T_h^{\rm obs}(\boldsymbol{k}) = \mathcal{S}_b(k_\perp)\,\mathcal{S}_{\rm FG}(k_\parallel)\,\delta T_h(\boldsymbol{k}). \tag{7.48}$$

We will consider two cases based on the functions proposed in Cunnington et al. (2020, 2021), Equation (2.14): **(1)** the Heaviside (HS) model, in which there is only a top-hat function eliminating every mode below some cut-off scale $k_\parallel^{\rm min}$, and **(2)** the smooth (SM) foregrounds model given by the exponential damping of low-$k_\parallel$ modes. They will be written as:

$$\mathcal{S}_{\rm FG}^{\rm HS}(k_\parallel) = \Theta_{\rm FG}(k_\parallel) \equiv \begin{cases} 0, & k_\parallel < k_\parallel^{\rm min} \\ 1, & k_\parallel > k_\parallel^{\rm min} \end{cases}, \quad \text{and} \quad \mathcal{S}_{\rm FG}^{\rm SM}(k_\parallel) = \left(1 - \exp\left[-\frac{k_\parallel}{k_\parallel^{\rm FG}}\right]\right). \tag{7.49}$$

Above, $k_\parallel^{\rm FG}$ is the damping scale below which wavemodes are suppressed when the foregrounds are removed (for the case of a perfect cleaning procedure, $k_\parallel^{\rm FG} = 0$).

Power spectrum The 21-cm power spectrum will be modelled as

$$P_{hh}(z, k_\parallel, \boldsymbol{k}_\perp) = b_h^2(z)\,\bar{T}_h(z)^2\,P(z,k)\,\mathcal{S}_b^2(|\boldsymbol{k}_\perp|)\,\mathcal{S}_{\rm FG}^2(k_\parallel) + P_{hh}^{\rm noise}, \tag{7.50}$$

where

$$P_{hh}^{\rm noise} = \sigma_{\rm pix}^2 V_{\rm pix}, \tag{7.51}$$

is the instrumental noise in the HI intensity mapping and

$$V_{\rm pix} = \Omega_{\rm pix}\int_{z_{\rm min}}^{z_{\rm max}} {\rm d}z \frac{c\,r^2(z)}{H(z)} \tag{7.52}$$

is the volume of a pixel.

Recall that

$$\sigma_{\rm pix}^2 = T_{\rm sys}^2 \frac{1}{\Delta\nu\,t_{\rm tot}} \frac{\Omega_{\rm tot}}{\Omega_{\rm pix}} \frac{1}{N_{\rm dishes} N_{\rm beams}}. \tag{7.53}$$

In our particular setting, the receiver noise will be given by (Bacon et al., 2020):

$$T_{\rm rx} = 15\,{\rm K} + 30\,{\rm K}\left(\frac{\nu}{\rm GHz} - 0.75\right)^2. \tag{7.54}$$

This is assumed for the SKA Band 1, which covers the redshift range $0.35 < z < 3$. We shall also consider $N_{\rm dishes} = 133$ with 15 meters in diameter, and $N_{\rm beams} = 1$. The Band 1 survey

will have an integration time of $t_{\rm tot} = 10000$ hours, covering $\Omega_{\rm tot} = 20000$ deg^2, with frequency resolution $\Delta\nu = 15.2$ kHz.

Bispectrum The 21cm auto-bispectrum will be given by

$$\langle \delta T_h(\boldsymbol{k}_1)\, \delta T_h(\boldsymbol{k}_2)\, \delta T_h(\boldsymbol{k}_3)\rangle \equiv (2\pi)^3 B_{hhh}(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3)\delta^D(\boldsymbol{k}_1 + \boldsymbol{k}_2 + \boldsymbol{k}_3), \tag{7.55}$$

where

$$B_{hhh}(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3) = \bar{T}_h^3(z)\left\{b_{h,1}^3 B(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3) + b_{h,1}^2 b_{h,2}\left[P(k_1)P(k_2) + \text{perms.}\right]\right\}. \tag{7.56}$$

From the definition of $\delta T_h^{\rm obs}$, Equation (2.16), beam smoothing and foreground removal will impact the 21-cm bispectrum as

$$B_{hhh}^{\rm obs}(\boldsymbol{k}, \boldsymbol{p}, \boldsymbol{q}) = B_{hhh}(\boldsymbol{k}, \boldsymbol{p}, \boldsymbol{q})\, \mathcal{S}_b(k_\perp, p_\perp, q_\perp)\, \mathcal{S}_{\rm FG}(k_\parallel, p_\parallel, q_\parallel), \tag{7.57}$$

where $\mathcal{S}_b(k_\perp, p_\perp, q_\perp) \equiv \mathcal{S}_b(k_\perp)\mathcal{S}_b(p_\perp)\mathcal{S}_b(q_\perp)$ and $\mathcal{S}_{\rm FG}(k_\parallel, p_\parallel, q_\parallel) \equiv \mathcal{S}_{\rm FG}(k_\parallel)\mathcal{S}_{\rm FG}(p_\parallel)\mathcal{S}_{\rm FG}(q_\parallel)$ are just a product of the functions defined in Equations (2.13) and (2.14).

As pointed out by Cunnington et al. (2021, Section 3.4 and references therein), statistical fluctuations of the thermal (Gaussian) noise that might contribute to the bispectrum noise is irrelevant for realistic intensity mapping calibrations. Hence, we do not include them in our analysis.

7.3.3 Cross-correlations

Finally, the cross-spectrum between galaxies and neutral hydrogen IM is

$$P_{gh}(k_\parallel, k_\perp, z) = \bar{T}_h(z)\, b_h(z)\, b_g(z)\, P(k, z)\, \mathcal{S}_b(k_\perp)\, \mathcal{S}_{\rm FG}(k_\parallel). \tag{7.58}$$

We shall assume the noises between both surveys to be uncorrelated.

Analogously, for the bispectrum we have:

$$B_{ghh}(\boldsymbol{k}, \boldsymbol{p}, \boldsymbol{q}) = \bar{T}_h^2(z)\mathcal{S}_b(p_\perp, q_\perp)\mathcal{S}_{\rm FG}(p_\parallel, q_\parallel) \times \left[b_{g,1}b_{h,1}^2 B(\boldsymbol{k}, \boldsymbol{p}, \boldsymbol{q}) + b_{g,1}b_{h,1}b_{h,2}P(k)P(p) + b_{g,1}b_{h,2}b_{h,1}P(k)P(q) + b_{g,2}b_{h,1}^2 P(p)P(q)\right]. \tag{7.59}$$

7.4 Application to a Gaussian distribution

To get a quantitative intuition on how the power spectra and bispectra help to constrain the redshift distribution parameters, we consider the Gaussian case, Equation (7.4). The Fourier coefficients of the distribution are given by $\phi_{k_\parallel} = L^{-1} \mathrm{e}^{-\frac{k_\parallel^2 \sigma_\parallel^2}{2}} \mathrm{e}^{-ik_\parallel r_\parallel^*}$. We are able, thus, to propagate the uncertainties over $\phi_{k_\parallel}$, estimated through the Fisher matrices of Section 7.2.1, to the free parameters of the Gaussian distribution $\phi(r_\parallel)$, namely $\theta_1 = r_\parallel^*$ and $\theta_2 = \sigma_\parallel$, through a change of variables according to

$$F_{\theta_1\theta_2} = \sum_{k_\parallel q_\parallel} \partial_{\theta_1}\phi_{k_\parallel} \partial_{\theta_2}\phi_{q_\parallel} F_{k_\parallel q_\parallel}. \tag{7.60}$$

Therefore,

$$\frac{\partial \phi_{k_\parallel}}{\partial r_\parallel^*} = -\frac{ik_\parallel}{L}\,\phi(k_\parallel) \Rightarrow F^{P,B}_{r_\parallel^* r_\parallel^*} = \sum_{k_\parallel q_\parallel} -\frac{k_\parallel\,\phi(k_\parallel)}{L}\,\frac{q_\parallel\,\phi(q_\parallel)}{L}\,F^{P,B}_{k_\parallel q_\parallel}, \tag{7.61}$$

$$\frac{\partial \phi_{k_\parallel}}{\partial \sigma_\parallel} = -\frac{k_\parallel^2 \sigma_\parallel}{L}\,\phi(k_\parallel) \Rightarrow F^{P,B}_{\sigma_\parallel \sigma_\parallel} = \sum_{k_\parallel q_\parallel} \frac{k_\parallel^2\,\phi(k_\parallel)}{L}\,\frac{q_\parallel^2\,\phi(q_\parallel)}{L}\,\sigma_\parallel^2\,F^{P,B}_{k_\parallel q_\parallel}. \tag{7.62}$$

7.4.1 Power-spectrum constraints

We begin by looking at the constraining power of using two-point statistics, i.e. applying Equations (7.31) and (7.32) to Equation (7.60). We obtain the following result:

$$\sigma_P^{-1}(\theta) = \frac{A}{2\pi}\int \mathrm{d}k_\perp\, k_\perp \mathcal{I}^{TT}(k_\perp) \int \frac{\mathrm{d}k_\parallel}{2\pi}\,\mathcal{K}_\theta^2(k_\parallel,\sigma_\parallel)\,\mathrm{e}^{-k_\parallel^2\sigma_\parallel^2}\,\frac{P_{gh}^2(k_\parallel,k_\perp)}{P_{hh}(k_\parallel,k_\perp)} +$$
$$-\frac{A}{\pi}\int \mathrm{d}k_\perp k_\perp \left[\mathcal{I}^{TT}(k_\perp)\right]^2 \left\{\int \frac{\mathrm{d}k_\parallel}{2\pi}\,\frac{\mathrm{d}q_\parallel}{2\pi}\,\mathcal{K}_\theta(k_\parallel,\sigma_\parallel)\,\mathcal{K}_\theta(q_\parallel,\sigma_\parallel)\,\mathrm{e}^{-\sigma_\parallel^2(k_\parallel^2+q_\parallel^2)}\,\mathcal{R}(k_\parallel,k_\perp)\,\mathcal{R}(q_\parallel,k_\perp),\right. \tag{7.63}$$

where

$$\mathcal{R}(k_\parallel,k_\perp) \equiv P_{gg}(k_\parallel,k_\perp) - \frac{P_{gh}^2(k_\parallel,k_\perp)}{P_{hh}(k_\parallel,k_\perp)}, \tag{7.64}$$

$\mathcal{K}_{\theta=r_\parallel^*} = k_\parallel$ and $\mathcal{K}_{\theta=\sigma_\parallel^*} = \sigma_\parallel k_\parallel^2$, and $\mathrm{Cov}(r_\parallel^*,\sigma_\parallel) = 0$. Neglecting galaxy auto-correlations, we end up with:

$$\sigma_P^{-1}(\theta) \simeq \frac{A}{2\pi}\int \mathrm{d}k_\perp k_\perp \mathcal{I}^{TT}(k_\perp) \int \frac{\mathrm{d}k_\parallel}{2\pi}\,\mathcal{K}_\theta^2(k_\parallel,\sigma_\parallel)\,\mathrm{e}^{-k_\parallel^2\sigma_\parallel^2}\,\frac{P_{gh}^2(k_\parallel,k_\perp)}{P_{hh}(k_\parallel,k_\perp)}. \tag{7.65}$$

By confronting the full Equation (7.63) with its approximation, Equation (7.65), we obtained the constraints shown in Figure 7.7 for the case of $\theta = \sigma_\parallel$, while the differences between both equations were absent for $\theta = r_\parallel^*$. The effect of including information from $\mathcal{P}_{gg}^{TT}$ is to highly constrain the variance of the redshift distribution, as it can be seen in the right panel of

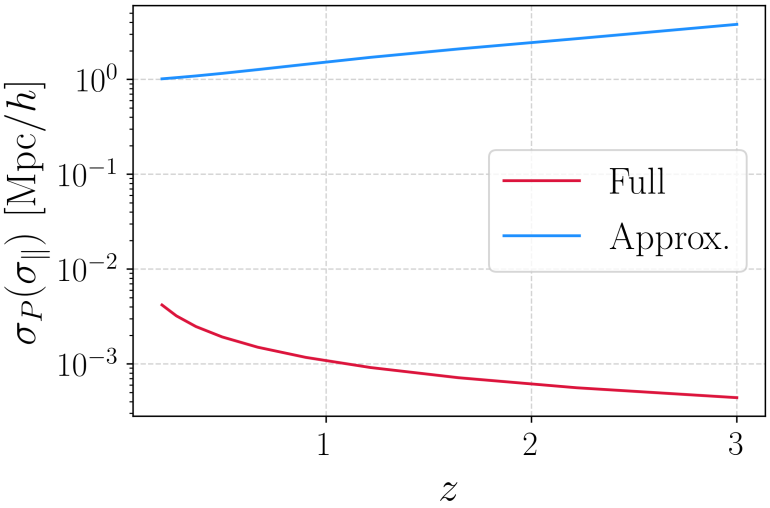


Figure 7.7: The full constraints (*thick red*) on the parameter $\sigma_\parallel$ obtained from Equation (7.63), and considering the approximation (*thin blue*) in which we neglect the autocorrelations of the projected galaxy field, Equation (7.65). We assume the smooth foregrounds (only exponential damping) with $k_\parallel^{\rm FG} = 0.02$ [h/Mpc], and photo-z width $\sigma_0 = 0.03$.

Figure 7.7. However, this information piece is highly uncertain due to the presence of unknown galaxy biases, the latter being degenerated with $\phi(z)$. Therefore, if the only unknown quantity was $\sigma_\parallel$, including this additional information is promising. However, due to the presence of bias uncertainties, $\mathcal{P}_{gg}^{TT}$ will also be a highly uncertain quantity. For that reason, we neglect the $\mathcal{P}_{gg}^{TT}$ contribution from the Fisher matrix, which results in the blue curve of Figure 7.7. We show the results for the cross-spectra in Figure 7.8 for different photo-z scattering amplitudes, σ_0, and different foreground cut-off scales, $k_\parallel^{\rm FG}$.

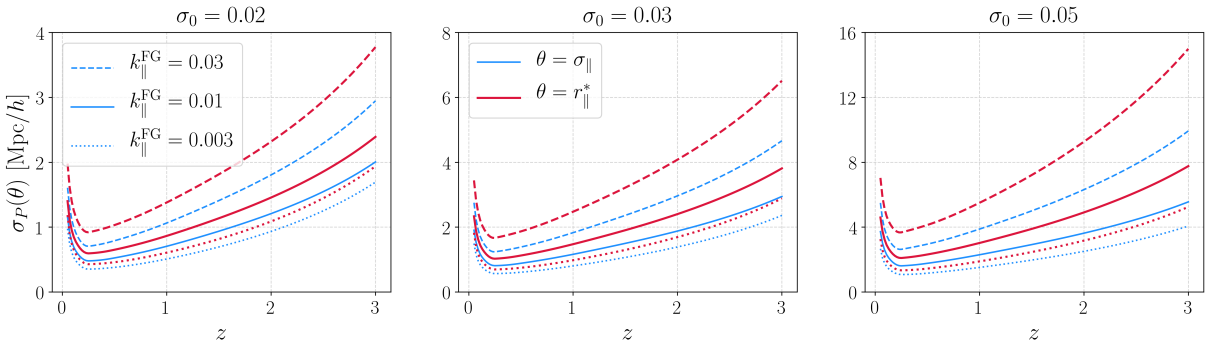


Figure 7.8: Power-spectrum constraints as a function of redshift z on both parameters $\sigma_\parallel$ (*thin blue*) and $r_\parallel^*$ (*thick red*) of the Gaussian photo-z bin, Equation (7.4). The curves assume the smooth foreground model, $\mathcal{S}_{\rm FG}^{\rm SM}$. Dashed, solid and dotted curves have cut-off scale $k_\parallel^{\rm FG} = 0.03, 0.01$, and 0.003, respectively. Different panels consider different photo-z amplitudes, from the LSST goal (*left*), LRG expectation (*middle*), and general galaxy population (*right*). We call the attention to the y-axes scale, which are doubled from the left to the right. The integrals are carried up to $k_{\rm max} = 0.3$ [h/Mpc].

7.4.2 Bispectrum constraints

The highly uncertain galaxy bias, discussed in Section 7.4.1, motivates the simplification of the bispectrum case. For that, we used the approximated Fisher matrix of Equation (7.34) to derive the uncertainties in the parameters, from Equation (7.60). They are:

$$\sigma_B^{-1}(\theta) = \frac{1}{6}\frac{A}{4\pi}\int \mathrm{d}k_\perp \mathrm{d}p_\perp \mathrm{d}q_\perp \frac{k_\perp p_\perp q_\perp}{\pi^2 A_T}\, \mathcal{I}^{TT}(k_\perp) \times \int \frac{\mathrm{d}p_\parallel}{2\pi}\frac{\mathrm{d}k_\parallel}{2\pi}\, \mathcal{K}_\theta^2(k_\parallel, \sigma_\parallel)\, \mathrm{e}^{-k_\parallel^2\sigma_\parallel^2}\, \frac{B_{ghh}^2(k_\parallel, p_\parallel, -k_\parallel - p_\parallel; k_\perp, p_\perp, q_\perp)}{P_{hh}(p_\parallel, q_\perp) P_{hh}(-k_\parallel - p_\parallel, p_\perp)}, \tag{7.66}$$

with $\mathcal{K}_{\theta=r_\parallel^*} = k_\parallel$ and $\mathcal{K}_{\theta=\sigma_\parallel^*} = \sigma_\parallel\, k_\parallel^2$ and $\mathrm{Cov}(r_\parallel^*, \sigma_\parallel) = 0$ as before.

In what follows, we show our preliminary results on the bispectrum constraints, assuming smooth foregrounds $\mathcal{S}_{\rm FG}^{\rm SM}$ and the Heaviside model $\mathcal{S}_{\rm FG}^{\rm HS}$, with $k_{\rm min} = k_\parallel^{\rm FG}$. We take the fiducial values for the photo-z scattering amplitude, foreground cut-off scale and maximum wavenumber in the integrals to be, respectively, $\sigma_0 = 0.03$, $k_\parallel^{\rm FG} = 0.02$ $[h/\mathrm{Mpc}]$ and $k_{\rm max} = 0.3$ $[h/\mathrm{Mpc}]$.

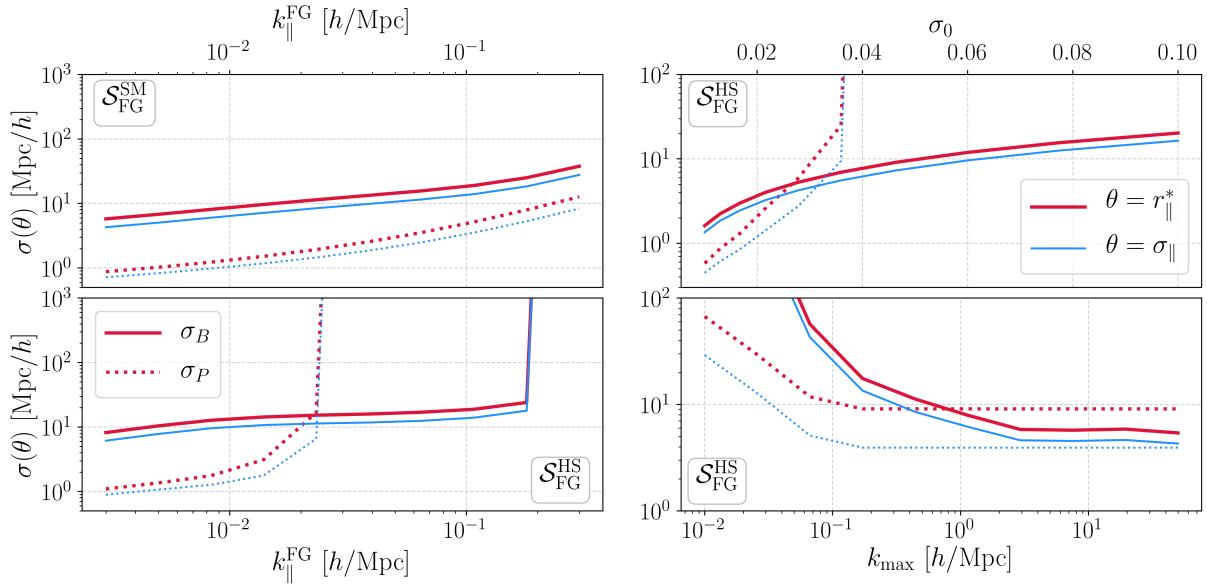


Figure 7.9: Power-spectrum (*dotted*) and bispectrum (*solid*) constraints, on both Gaussian photo-z parameters, $\theta = r_\parallel^*$ (*thick red*) and $\theta = \sigma_\parallel$ (*thin blue*), as a function of foreground damping/cut-off scale $k_\parallel^{\rm FG}$ (*left, top and bottom*), photo-z scatter σ_0 (*right, top*), and maximum wavenumber $k_{\rm max}$ (*right, bottom*). In the left plot on the top panel, foregrounds are treated via the smooth exponential damping, $\mathcal{S}_{\rm FG}^{\rm SM}(k_\parallel)$ in Equation (7.49), while the bottom panel shows the same constraints for the Heaviside model, $\mathcal{S}_{\rm FG}^{\rm HS}(k_\parallel)$. The plots on the right panels consider only the $\mathcal{S}_{\rm FG}^{\rm HS}(k_\parallel)$ modelling.

In Figure 7.9, solid curves show the bispectrum constraints, neglecting the second-order bias corrections, in comparison with what can be attainable with the power spectrum (*dashed*), for a single redshift bin centred at $z = 0.8$. The corresponding linear bias and number density for the photometric galaxies are fixed at the values $b \approx 1.4$ and $\bar{n} \approx 0.004$ $[h/\mathrm{Mpc}]^3$, respectively.

For completeness, we show the constraints on both parameters: bin centre $r_{\parallel}^{*}$ (thick-red) and photo-z dispersion (thin-blue).

As we can see, for the case in which some information is still present below the foreground cut-off scale $k_{\parallel}^{\rm FG}$ (modelled by $\mathcal{S}_{\rm FG}^{\rm SM}$), then the power-spectrum constraints are less dependent on $k_{\parallel}^{\rm FG}$ (Figure 7.9, dotted curves shown in the top-left panel) if compared with the bispectrum (solid curves). However, if we are unable to recover any information coming from modes $k_{\parallel} < k_{\parallel}^{\rm FG}$, then the bispectrum is the best statistics to reconstruct the photometric redshifts: while the power spectrum constraints diverge at scales $k_{\parallel}^{\rm FG} \gtrsim 0.02$ $[h/\mathrm{Mpc}]$ (bottom-left panel), the bispectrum is capable of dealing with foreground scales $k_{\parallel}^{\rm FG} \lesssim 0.2$ $[h/\mathrm{Mpc}]$.

Another interesting feature is that the bispectrum is less sensitive to the photo-z scattering amplitude σ_0 than the power spectrum (Figure 7.9, top-right panel). Finally, although the bispectrum is highly dependent on the scales $k_{\rm max}$ included in the analysis (bottom-right), modelling the theoretical signal up to $k_{\rm max} \sim 1$ should be enough.

In Figure 7.10, we present the forecast results with the fiducial values already mentioned and adopting the Heaviside model for foregrounds (i.e. losing all information below the cut-off scale $k_{\parallel}^{\rm FG}$). The left panel neglects second-order bias b_2 contributions to the bispectrum. Nonetheless, to show how these constraints depend on b_2, in the right panel we assumed b_2 as a function of the linear bias, via Equation (7.44). While the bispectrum outperforms the power-spectrum constraints up to $z \approx 1$ and $z \approx 0.75$, respectively for $r_{\parallel}^{*}$ and $\sigma_{\parallel}$, the decrease in its signal caused by b_2 makes the power spectrum outperforms the bispectrum sooner. Therefore, the second-order bias modelling have a major impact on the clustering redshifts approach via the 3-point statistics.

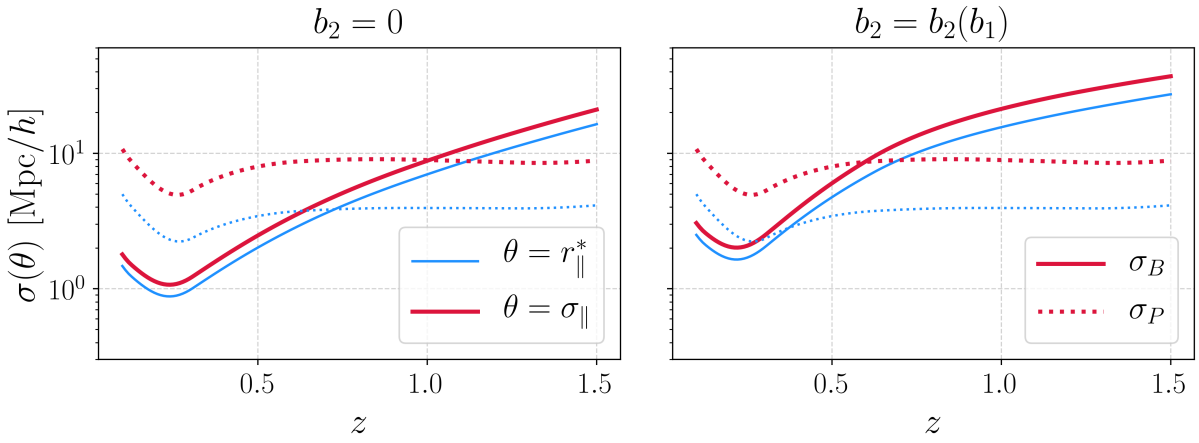


Figure 7.10: Forecast for the errors attainable on both Gaussian photo-z parameters, $\theta = r_{\parallel}^{*}$ (*thick red*) and $\theta = \sigma_{\parallel}$ (*thin blue*), when utilising the power spectrum (*dotted*) and bispectrum (*solid*) correlators. The Heaviside foreground model $\mathcal{S}_{\rm FG}^{\rm HS}$ is assumed, with the cut-off scale fixed at $k_{\parallel}^{\rm FG} = 0.02$ $[h/\mathrm{Mpc}]$. The left panel neglects the second-order bias contributions, while on the right we show the impact of including b_2, following Equation (7.44). Photo-z scatter is given by $\sigma_z = 0.03(1+z)$ and the maximum scales included in the integrals is $k_{\rm max} = 0.3$ $[h/\mathrm{Mpc}]$.

Whilst these results hold for a SKA-like survey operating in the single-dish mode, considering the loss of radial modes due to different foreground removal modelling and the beam smoothing caused by the radio telescope, we are redoing the analysis for an interferometer array with the *Hydrogen Intensity and Real-time Analysis eXperiment* (HIRAX) specifications (Newburgh et al., 2016). This type of experiment has an additional contribution to the foregrounds coming from the so-called "wedge effect" (Seo and Hirata, 2016), which is responsible for the removal of radial modes as a function of the angular scales $k_\perp$ due to the chromatic response of the interferometer set-up.

Chapter 8

Conclusions and Final Remarks

In this thesis we concerned ourselves with the clustering statistics of large-scale structure tracers. We saw how they can be used to probe primordial non-Gaussianities (PNG, Chapter 4) and general relativistic effects (RE, Chapters 5 and 6). It is worth remarking that, besides the observables considered in this work, namely 2- and 3-point summary statistics, and the galaxy/halo overdensity fields, effects on the halos and voids abundance, on the reionisation epoch, and the use of Minkowski functionals provide alternative approaches to the study of both topics.

Regarding primordial-non Gaussianites, we are currently implementing the local type, responsible for the appearance of the k^{-2} scaling in the bias of dark matter tracers, in the `ExSHalos` numerical code Voivodic et al. (2019). We plan to generate snapshots at a certain redshift to obtain power spectrum and bispectrum constraints on $f_{\rm NL}^{\rm loc}$ via the multi-tracer technique, employing halos of different masses for that matter, and hence different biases. We also will include a halo-occupation distribution to simulate galaxies and introduce real-life effects (e.g. photometric redshift errors and relativistic effects) to obtain realistic constraints.

We plan to include relativistic effects with the `LIGER` numerical code (Borzyszkowski et al., 2017), which is able to convert the Newtonian simulations generated by `ExSHalos` into a set of light cones with all the first order relativistic effects included. For that, we will focus on one redshift bin, centred at the redshift of the initial analysis, sampled with a specific bin Δz suitable for the future surveys explored in Chapter 2. The binding between PNG and RE, the two most relevant effects on the largest scales, and a realistic simulation setting, is necessary for a robust analysis of both effects. Since `ExSHalos` allows us to generate fast halo and galaxy mocks, we will also be able to explore how these different features of the LSS impact covariance matrices.

A more complete and self-consistent route to implement PNG in the presence of relativistic effects is to use the *gevolution* code, employed in the analysis of Chapter 6. Recently, Adamek et al. (2021) developed a framework to generate second-order initial conditions in the

relativistic context, which is necessary correctly model the interplay between both effects on large scales: if we were merely interested in the Newtonian evolution of a density field in the presence of primordial non-Gaussian initial conditions, applying the relevant PNG corrections to the initial potentials, e.g. Φ, is enough. However, due to the non-linear evolution of gravitational modes, PNG should be evolved beyond the linear theory[1].

Regarding relativistic effects *per se*, we have performed a power-spectrum multipole analysis on the data extracted from a light cone generated with the fully relativistic *gevolution* N-body code. We focused on the dipole signal in the cross-correlation between different dark matter halo sub-populations, which is a purely relativistic (non-Newtonian) effect. The simulation employed a novel ray-tracing method to connect the halos with the observer, being capable of incorporating all relevant general relativistic effects on the cosmologically relevant distance scales. We showed in detail how the survey window function and quantities such as the evolution bias can be estimated on the past light cone, allowing a rigorous comparison with gauge-invariant theoretical calculations at linear order.

Our analysis was restricted to a particular high-redshift bin in the range $1.7 \leq z \leq 2.9$, covering a limited sky fraction of $f_{\rm sky} = 0.01$. This is a similar sky area to the overlap region between luminous red and emission line galaxies in the multi-tracer analysis of the final eBOSS data (see Table 2 of Zhao et al., 2021), although in the latter the data are from lower redshifts ($z \sim 1$). While we were able to robustly test our analysis methods using these simulated data, no conclusive detection of the dipole signature was possible due to the limited volume of the redshift bin, a challenge that is of paramount importance for current surveys as well.

To improve the analysis and perhaps detect the signal within these simulations, we plan to use optimal weighting schemes to enhance the signal (as done in Castorina et al., 2019, for example). In particular, using the multi-tracer estimators of Abramo et al. (2015), explained in Sections 3.2.3 and D.2, could help to enhance the signal's detectability, which shall be explored in future work.

We also did not incorporate wide-angle effects in our modelling, as they were not relevant for the solid angle and redshift range of our analysis. A thorough account of these effects should also be explored in the context of wider survey areas however, particularly in the case of future surveys like Euclid, LSST and SKA, which are expected to cover an appreciable fraction of the sky. We will include these effects and repeat the analysis on a full-sky *gevolution* simulation that we already have and include integrated effects (e.g. lensing) in the analytical model: they were fully included in the data considered in this work, but were neglected in the present analysis; however, they impact large angular scales. Despite the Doppler term being the largest contribution to the relativistic effects for our particular setup, non-local terms should also be

[1] Private communication with Julian Adamek, 2020.

modelled and properly included for analyses that go to larger scales.

Furthermore, our analysis was performed by considering that the halos' pairs, used to compute the multipoles estimator, were at a fixed redshift ($\bar{z} = 1.89, 2.29$, and 2.69). However, this is not true, as different pairs have different effective redshifts. In Obuljen and Percival (2021) it was found an additional scale-dependent clustering anisotropy coming from these differences in the effective redshift assumed for the pairs. While they have found 4% and 40% differences in the quadrupole and hexadecapole, respectively (with negligible effects in the monopole), no estimate was made for the dipole. Because RE, as well as PNG, rely on large scales, constructing a narrow redshift bin can degrade the scales available for analysis. Therefore, it is interesting to study this effect and its impact on these measurements.

In this work, we have limited our analysis to a single high-redshift bin with a relatively narrow survey area, and have pursued only a limited set of observables, i.e. the multipoles of the relativistic power spectrum. In future work, we will relax these limitations by moving to larger survey volumes more representative of the next generation of large-scale structure surveys, while also including wide-angle and integrated effects, and extending our analysis to two-point correlation functions and multipoles of the relativistic bispectrum.

In what comes to the application of the multi-tracer method to the reconstruction of photometric redshifts (Chapter 7), we cross-correlated photometric galaxies with intensity maps of neutral hydrogen. Because the latter suffers from a loss of long-radial modes, as discussed in Section 2.3, the constraints of redshift distribution parameters via this clustering redshifts approach can be severely degraded when the photometric redshifts are poorly determined. Therefore, we explored the coupling between long- and short-wavelength modes induced by the non-linear evolution of the gravitational collapse to trying to recover the modes lost to foregrounds. For that, we chose the bispectrum as the summary statistics for this exercise.

At the moment of these conclusions, we have found that the bispectrum is an interesting statistics for the clustering redshifts approach as it is less sensitive to the foreground cut-off scale $k_{\parallel}^{\mathrm{FG}}$ and the photometric redshifts scattering amplitude σ_0, which enters the photo-z error like $\sigma_z = \sigma_0 (1 + z)$. We also saw how the maximum scale k_{max} up to which we can model the theory is highly relevant for the clustering-z with three-point statistics and that the constraints can be improved by including smaller scales, as long as the impact of non-linearities can be quantified. Nevertheless, we found that, given the expected sensitivity of SKA, the constraining power saturates at $k_{\mathrm{max}} \sim 1$ $[h/\mathrm{Mpc}]$. It is important to use scales slightly inside the non-linear regime ($k_{\mathrm{max}} \sim 0.3$ $[h/\mathrm{Mpc}]$), as the constraints can be severely degraded if only modes $k \sim 0.15$ $[h/\mathrm{Mpc}]$ are used. Hence, it is worth exploring the impact of higher-order biases and the non-linear effects, which plague these scales, on the clustering-redshifts method. We point out, however, that a conservative stand that includes only linear scales ($k_{\mathrm{max}} \sim 0.15$ $[h/\mathrm{Mpc}]$) is not strictly necessary, and in fact traditional applications of the clustering redshifts approach

using 2-point functions have made use of non-linear scales, e.g. down to separations of $\sim$ 0.5 Mpc (Chiang et al., 2019), showing robust constraints on the tomographically reconstructed quantities. Therefore, $k_{\rm max} \sim 0.3$ $[h/{\rm Mpc}]$ is reasonable for this calibration procedure. Finally, we showed how the second-order bias b_2 is relevant for this method, as it controls the amplitude of the signal, setting the redshifts up to which the bispectrum outperforms the power-spectrum constraints. Therefore, having a proper and reliable way to model b_2 for both galaxies and neutral hydrogen is of great importance.

We stress that this is a work in progress: so far we only showed the forecast for a SKA-like survey operating in the single-dish mode. We are extending the analysis for interferometer arrays, like HIRAX, where an additional loss of radial modes is present due to the so-called "wedge effect". As future prospects, we aim at using real/simulated data, as opposed to the simplistic forecast described in Chapter 7. Because the estimators used in this work (Section D.3) are relatively simple, they could be computed in a feasible amount of time. Therefore, we plan to apply our method to a more realistic scenario, including a more complete description of the bispectrum and possibly breaking the flat-sky assumption, using simulations to study the impact of non-linearities and $k_{\rm max}$, and targeting subsequent applications to real galaxy surveys data.

We also point that this method can be applied to other LSS probes, such as Ly-alpha forest and quasars – with the latter suffering from systematics at large-radial scales similarly to 21-cm IM –, and that it is a promising tool to tackle the detection of primordial non-Gaussianities via 21-cm surveys: while this science case is benefited by the vast volumes reachable with the intensity-mapping technique, it is exactly at the large-radial scales lost by foreground removal that PNG leave distinct imprints at the power spectrum. Therefore, we leave these prospects to future work.

Finally, we are examining the possibility to extend the estimator presented in Section 3.2.4 to multiple tracers of the LSS, generalising the idea of Abramo et al. (2015) to the bispectrum. We are analysing the possibility to include the simplified Fisher matrix presented in Section D.3. The goal is to estimate the bispectrum of a single tracer with the smallest error allowed by the information that can be extracted from the multi-tracer technique. In principle, this estimator may help us to improve the detectability of the relativistic bispectrum dipole, which in spite of not requiring cross-correlations to be present in the bispectrum signal, can benefit from the extra information encoded in various tracers of the large-scale structure.

To close these final remarks: cosmology is in the realm in which we can test different theories. Although it was not the topic of this thesis, there are high-energy phenomena and quantum gravity effects that cannot be probed in the laboratory. However, because the energy scales involved in the early universe are orders of magnitude higher than what can be achieved by particle accelerators, the primordial universe may help us to narrow down some possibilities

of quantum gravity and test high-energy physics. All the information that is present in the clustering of galaxies, from the largest to the smallest scales, is a unique tool to shed light on fundamental physics, from the macro to the microdynamics of our theories. With all the new experiments soon to be operating, the future could not seem brighter.

Bibliography

Bibliography

Abazajian, K., Arnold, K., Austermann, J., Benson, B., Bischoff, C., Bock, J., Bond, J., Borrill, J., Buder, I., Burke, D. and et al., Inflation physics from the cosmic microwave background and large scale structure. *Astroparticle Physics*, vol. 63, pp. 55–65, 2015.

Abramo, L. R. and Leonard, K. E., Why multitracer surveys beat cosmic variance. *Monthly Notices of the Royal Astronomical Society*, vol. 432, no. 1, pp. 318–326, 2013.

Abramo, L. R., Secco, L. F. and Loureiro, A., Fourier analysis of multitracer cosmological surveys. *Monthly Notices of the Royal Astronomical Society*, vol. 455, no. 4, pp. 3871–3889, 2015.

Acquaviva, V., Bartolo, N., Matarrese, S. and Riotto, A., Gauge-invariant second-order perturbations and non-Gaussianity from inflation. *Nuclear Physics B*, vol. 667, no. 1-2, pp. 119–148, 2003.

Adamek, J., Calles, J., Montandon, T., Noreña, J. and Stahl, C., Relativistic second-order initial conditions for simulations of large-scale structure, 2021.

Adamek, J., Daverio, D., Durrer, R. and Kunz, M., General relativity and cosmic structure formation. *Nature Phys.*, vol. 12, no. 4, pp. 346–349, 2016a.

Adamek, J., Daverio, D., Durrer, R. and Kunz, M., gevolution: a cosmological N-body code based on General Relativity. *Journal of Cosmology and Astroparticle Physics*, vol. 2016, no. 07, p. 053, 2016b.

Adamek, J. and Fidler, C., The large-scale general-relativistic correction for Newtonian mocks. *Journal of Cosmology and Astroparticle Physics*, vol. 2019, no. 09, p. 026, 2019.

Aghanim, N., Akrami, Y., Ashdown, M., Aumont, J., Baccigalupi, C., Ballardini, M., Banday, A. J., Barreiro, R. B., Bartolo, N. and et al., Planck 2018 results. VI. Cosmological parameters. *Astronomy & Astrophysics*, vol. 641, p. A6, 2020.

Akrami, Y., Arroja, F., Ashdown, M., Aumont, J., Baccigalupi, C., Ballardini, M., Banday, A., Barreiro, R., Bartolo, N., Basak, S. et al., Planck 2018 results. IX. Constraints on primordial non-Gaussianity. *Astronomy & Astrophysics*, vol. 641, p. A9, 2020.

Alonso, D., CUTE solutions for two-point correlation functions from large cosmological datasets. *arXiv preprint: 1210.1833*, 2012.

Alonso, D., Bellini, E., Hale, C., Jarvis, M. J. and Schwarz, D. J., Cross-correlating radio continuum surveys and CMB lensing: constraining redshift distributions, galaxy bias, and cosmology. *Monthly Notices of the Royal Astronomical Society*, vol. 502, no. 1, p. 876–887, 2021.

Alonso, D. and Ferreira, P. G., Constraining ultralarge-scale cosmology with multiple tracers in optical and radio surveys. *Physical Review D*, vol. 92, p. 063525, 2015.

Alonso, D., Ferreira, P. G., Jarvis, M. J. and Moodley, K., Calibrating photometric redshifts with intensity mapping observations. *Physical Review D*, vol. 96, no. 4, p. 043515, 2017.

Alonso, D., Ferreira, P. G. and Santos, M. G., Fast simulations for intensity mapping experiments. *Monthly Notices of the Royal Astronomical Society*, vol. 444, no. 4, p. 3183–3197, 2014.

Alpher, R. A. and Herman, R., Evolution of the Universe. *Nature*, vol. 162, no. 4124, pp. 774–775, 1948.

Alpher, R. A. and Herman, R. C., Remarks on the evolution of the expanding universe. *Physical Review*, vol. 75, no. 7, p. 1089, 1949.

Alvarez, M., Baldauf, T., Bond, J. R., Dalal, N., de Putter, R., Doré, O., Green, D., Hirata, C., Huang, Z., Huterer, D. et al., Testing inflation with large scale structure: connecting hopes with reality. *arXiv preprint: 1412.4671*, 2014.

Amendola, L. and Tsujikawa, S., *Dark energy: theory and observations*. Cambridge University Press, 2010.

Appleby, S., Park, C., Pranav, P., Hong, S. E., Hwang, H. S., Kim, J. and Buchert, T., Minkowski functionals of SDSS-III BOSS: hints of possible anisotropy in the density field?, 2021.

Bacon, D., Bridle, S., Abdalla, F. B., Brown, M., Bull, P., Camera, S., Fender, R., Grainge, K., Ivezic, Z., Jarvis, M. et al., Synergy between the Large Synoptic Survey Telescope and the Square Kilometre Array. *arXiv preprint: 1501.03977*, 2015.

Bacon, D. J., Battye, R. A., Bull, P., Camera, S., Ferreira, P. G., Harrison, I., Parkinson, D., Pourtsidou, A., Santos, M. G. and et al., Cosmology with Phase 1 of the Square Kilometre Array Red Book 2018: Technical specifications and performance forecasts. *Publications of the Astronomical Society of Australia*, vol. 37, 2020.

Bagdonas, A., Zanetic, J. and Gurgel, I., Who discovered the expansion of the universe? Priority disputes as way to teach cosmology using history and philosophy of science. *Revista Brasileira de Ensino de Física*, vol. 39, no. 2, 2017.

Baker, T. and Bull, P., Observation signatures of modified gravity on ultra-large scales. *The Astrophysical Journal*, vol. 811, no. 2, p. 116, 2015.

Ballardini, M., Matthewson, W. L. and Maartens, R., Constraining primordial non-Gaussianity using two galaxy surveys and CMB lensing. *Monthly Notices of the Royal Astronomical Society*, vol. 489, no. 2, pp. 1950–1956, 2019.

Bardeen, J. M., Gauge-invariant cosmological perturbations. *Physical Review D*, vol. 22, no. 8, p. 1882, 1980.

Barreira, A., On the impact of galaxy bias uncertainties on primordial non-Gaussianity constraints. *Journal of Cosmology and Astroparticle Physics*, vol. 2020, no. 12, p. 031–031, 2020.

Barreira, A., Cabass, G., Schmidt, F., Pillepich, A. and Nelson, D., Galaxy bias and primordial non-Gaussianity: insights from galaxy formation simulations with IllustrisTNG. *Journal of Cosmology and Astroparticle Physics*, vol. 2020, no. 12, p. 013–013, 2020.

Bartolo, N., Komatsu, E., Matarrese, S. and Riotto, A., Non-Gaussianity from inflation: theory and observations. *Physics Reports*, vol. 402, no. 3-4, pp. 103–266, 2004.

Bates, D. J., Tojeiro, R., Newman, J. A., Gonzalez-Perez, V., Comparat, J., Schneider, D. P., Lima, M. and Streblyanska, A., Mass functions, luminosity functions, and completeness measurements from clustering redshifts. *Monthly Notices of the Royal Astronomical Society*, vol. 486, no. 3, pp. 3059–3077, 2019.

Baumann, D., Cosmology: Part III Mathematical Tripos. *University lecture notes*, 2014.

Baumann, D. and McAllister, L., *Inflation and string theory*. Cambridge University Press, 2015.

Behroozi, P. S., Wechsler, R. H. and Wu, H.-Y., The ROCKSTAR phase-space temporal halo finder and the velocity offsets of clsuter cores. *The Astrophysical Journal*, vol. 762, no. 2, p. 109, 2012.

Beisbart, C., Can we justifiably assume the cosmological principle in order to break model underdetermination in Cosmology? *Journal for general philosophy of science*, vol. 40, no. 2, pp. 175–205, 2009.

Bernardeau, F., Colombi, S., Gaztañaga, E. and Scoccimarro, R., Large-scale structure of the Universe and cosmological perturbation theory. *Physics reports*, vol. 367, no. 1-3, pp. 1–248, 2002.

Bertone, G. and Hooper, D., History of dark matter. *Reviews of Modern Physics*, vol. 90, no. 4, 2018.

Beutler, F., Castorina, E. and Zhang, P., Interpreting measurements of the anisotropic galaxy power spectrum. *Journal of Cosmology and Astroparticle Physics*, vol. 2019, no. 03, p. 040, 2019.

Beutler, F. and Dio, E. D., Modeling relativistic contributions to the halo power spectrum dipole. *Journal of Cosmology and Astroparticle Physics*, vol. 2020, no. 07, p. 048, 2020.

Beutler, F., Seo, H.-J., Saito, S., Chuang, C.-H., Cuesta, A. J., Eisenstein, D. J., Gil-Marín, H., Grieb, J. N., Hand, N., Kitaura, F.-S., Modi, C., Nichol, R. C., Olmstead, M. D., Percival, W. J., Prada, F., Sánchez, A. G., Rodriguez-Torres, S., Ross, A. J., Ross, N. P., Schneider, D. P., Tinker, J., Tojeiro, R. and Vargas-Magaña, M., The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: anisotropic galaxy clustering in Fourier space. *Monthly Notices of the Royal Astronomical Society*, vol. 466, no. 2, pp. 2242–2260, 2016.

Bianchi, D., Gil-Marín, H., Ruggeri, R. and Percival, W. J., Measuring line-of-sight-dependent Fourier-space clustering using FFTs. *Monthly Notices of the Royal Astronomical Society: Letters*, vol. 453, no. 1, pp. L11–L15, 2015.

Blas, D., Lesgourgues, J. and Tram, T., The Cosmic Linear Anisotropy Solving System (CLASS). Part II: Approximation schemes. *Journal of Cosmology and Astroparticle Physics*, vol. 2011, no. 07, p. 034, 2011.

Bonvin, C., Isolating relativistic effects in large-scale structure. *Classical and Quantum Gravity*, vol. 31, no. 23, p. 234002, 2014.

Bonvin, C. and Durrer, R., What galaxy surveys really measure. *Physical Review D*, vol. 84, no. 6, 2011.

Bonvin, C., Durrer, R. and Gasparini, M. A., Fluctuations of the luminosity distance. *Physical Review D*, vol. 73, no. 2, p. 023523, 2006.

Bonvin, C. and Fleury, P., Testing the equivalence principle on cosmological scales. *Journal of Cosmology and Astroparticle Physics*, vol. 05, p. 061, 2018.

Bonvin, C., Hui, L. and Gaztañaga, E., Asymmetric galaxy correlation functions. *Physical Review D*, vol. 89, p. 083535, 2014.

Borzyszkowski, M., Bertacca, D. and Porciani, C., `liger`: mock relativistic light cones from Newtonian simulations. *Monthly Notices of the Royal Astronomical Society*, vol. 471, no. 4, pp. 3899–3914, 2017.

Brandenberger, R. H., String Gas Cosmology, 2008.

Breton, M.-A., Rasera, Y., Taruya, A., Lacombe, O. and Saga, S., Imprints of relativistic effects on the asymmetry of the halo cross-correlation function: from linear to non-linear scales. *Monthly Notices of the Royal Astronomical Society*, vol. 483, no. 2, pp. 2671–2696, 2018.

Breton, M.-A. and de la Torre, S., Fast analytical calculation of the random pair counts for realistic survey geometry. *Astronomy & Astrophysics*, vol. 646, p. A40, 2021.

Bull, P., Akrami, Y., Adamek, J., Baker, T., Bellini, E., Beltrán Jiménez, J., Bentivegna, E., Camera, S., Clesse, S., Davis, J. H. and et al., Beyond ΛCDM: Problems, solutions, and the road ahead. *Physics of the Dark Universe*, vol. 12, p. 56–99, 2016.

Bull, P., Camera, S., Raccanelli, A., Blake, C., Ferreira, P., Santos, M. and Schwarz, D. J., Measuring baryon acoustic oscillations with future SKA surveys. *Proceeding of science*, vol. AASKA14, p. 024, 2015a.

Bull, P., Ferreira, P. G., Patel, P. and Santos, M. G., Late-time cosmology with 21-cm intensity mapping experiments. *The Astrophysical Journal*, vol. 803, no. 1, p. 21, 2015b.

Butterfield, J., On under-determination in cosmology. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, vol. 46, pp. 57–69, 2014.

Camarena, D., Marra, V., Sakr, Z. and Clarkson, C., The Copernican principle in light of the latest cosmological data, 2021.

Camera, S., Santos, M. G. and Maartens, R., Probing primordial non-Gaussianity with SKA galaxy redshift surveys: a fully relativistic analysis. *Monthly Notices of the Royal Astronomical Society*, vol. 448, no. 2, pp. 1035–1043, 2015.

Castorina, E., Hand, N., Seljak, U., Beutler, F., Chuang, C.-H., Zhao, C., Gil-Marín, H., Percival, W. J., Ross, A. J., Choi, P. D. and et al., Redshift-weighted constraints on primordial non-Gaussianity from the clustering of the eBOSS DR14 quasars in Fourier space. *Journal of Cosmology and Astroparticle Physics*, vol. 2019, no. 09, p. 010, 2019.

Catelan, P., Lucchin, F., Matarrese, S. and Moscardini, L., Eulerian perturbation theory in non-flat universes: second-order approximation. *Monthly Notices of the Royal Astronomical Society*, vol. 276, no. 1, pp. 39–56, 1995.

Challinor, A. and Lewis, A., Linear power spectrum of observed source number counts. *Physical Review D*, vol. 84, no. 4, 2011.

Chavez, H. O. C., *Estruturas em larga escala e o Dark Energy Survey*. Master's thesis, Instituto de Física, Universidade de São Paulo, São Paulo, 2014.

Chen, S.-F., Lee, H. and Dvorkin, C., Precise and accurate cosmology with CMB×LSS power spectra and bispectra. *Journal of Cosmology and Astroparticle Physics*, vol. 2021, no. 05, p. 030, 2021.

Chiang, Y.-K., Ménard, B. and Schiminovich, D., Broadband intensity tomography: spectral tagging of the cosmic UV background. *The Astrophysical Journal*, vol. 877, no. 2, 150, 2019.

Chisari, N. E. and Zaldarriaga, M., Connection between Newtonian simulations and general relativity. *Physical Review D*, vol. 83, no. 12, 2011.

Clarkson, C., Clifton, T. and February, S., Perturbation theory in Lemaître-Tolman-Bondi cosmology. *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 06, p. 025–025, 2009.

Clowe, D., Bradač, M., Gonzalez, A. H., Markevitch, M., Randall, S. W., Jones, C. and Zaritsky, D., A direct empirical proof of the existence of dark matter. *The Astrophysical Journal*, vol. 648, no. 2, p. L109–L113, 2006.

Coates, L., Adamek, J., Bull, P., Guandalin, C. and Clarkson, C., Observing relativistic features in large-scale structure surveys – II. Doppler magnification in an ensemble of relativistic simulations. *Monthly Notices of the Royal Astronomical Society*, vol. 504, no. 3, p. 3534–3543, 2021.

Cohen-Tannoudji, C., Diu, B., Laloe, F. and Dui, B., *Quantum Mechanics, Volume 2: Angular Momentum, Spin, and Approximation Methods*. John Wiley & Sons, 2006.

Condon, J. J. and Ransom, S. M., *Essential radio astronomy*. Princeton University Press, 2016.

Cooray, A. and Sheth, R., Halo models of large scale structure. *Physics Reports*, vol. 372, no. 1, p. 1–129, 2002.

Corasaniti, P. S., Ettori, S., Rasera, Y., Sereno, M., Amodeo, S., Breton, M.-A., Ghirardini, V. and Eckert, D., Probing cosmology with dark matter halo sparsity using X-ray cluster mass measurements. *The Astrophysical Journal*, vol. 862, no. 1, p. 40, 2018.

Cunnington, S., Camera, S. and Pourtsidou, A., The degeneracy between primordial non-Gaussianity and foregrounds in 21cm intensity mapping experiments. *Monthly Notices of the Royal Astronomical Society*, 2020.

Cunnington, S., Harrison, I., Pourtsidou, A. and Bacon, D., HI intensity mapping for clustering-based redshift estimation. *Monthly Notices of the Royal Astronomical Society*, vol. 482, no. 3, pp. 3341–3355, 2019a.

Cunnington, S., Watkinson, C. and Pourtsidou, A., The HI intensity mapping bispectrum including observational effects, 2021.

Cunnington, S., Wolz, L., Pourtsidou, A. and Bacon, D., Impact of foregrounds on HI intensity mapping cross-correlations with optical surveys. *Monthly Notices of the Royal Astronomical Society*, vol. 488, no. 4, pp. 5452–5472, 2019b.

Dalal, N., Doré, O., Huterer, D. and Shirokov, A., Imprints of primordial non-Gaussianities on large-scale structure: Scale-dependent bias and abundance of virialized objects. *Physical Review D*, vol. 77, no. 12, p. 123514, 2008.

Desjacques, V., Jeong, D. and Schmidt, F., Large-scale galaxy bias. *Physics Reports*, vol. 733, pp. 1–193, 2018.

Dizgah, A. M., Biagetti, M., Sefusatti, E., Desjacques, V. and Noreña, J., Primordial non-Gaussianity from biased tracers: likelihood analysis of real-space power spectrum and bispectrum. *Journal of Cosmology and Astroparticle Physics*, vol. 2021, no. 05, p. 015, 2021.

Donath, Y. and Senatore, L., Biased tracers in redshift space in the EFTofLSS with exact time dependence. *Journal of Cosmology and Astroparticle Physics*, vol. 2020, no. 10, p. 039, 2020.

Durrer, R., *The cosmic microwave background*. Cambridge University Press, 2020.

Ehlers, J., Geren, P. and Sachs, R. K., Isotropic solutions of the Einstein-Liouville equations. *Journal of Mathematical Physics*, vol. 9, no. 9, pp. 1344–1349, 1968.

Einstein, A., *Relativity: The Special and General Theory*. Henry Holt and Company, New York, 1920.

Ellis, G. F., Does the multiverse really exist? *Scientific American*, vol. 305, no. 2, pp. 38–43, 2011.

Ellis, G. F., Maartens, R. and MacCallum, M. A., *Relativistic cosmology*. Cambridge University Press, 2012.

Ellis, G. F. and Stoeger, W., The 'fitting problem' in cosmology. *Classical and Quantum Gravity*, vol. 4, no. 6, p. 1697, 1987.

Ellis, G. F. R., Issues in the Philosophy of Cosmology. In: J. Butterfield and J. Earman (Editors), *Philosophy of Physics*, vol. Part B of *Handbook of the Philosophy of Science*, Elsevier, 2007, pp. 1183–1285.

Faraoni, V. and Capozziello, S., *Beyond Einstein Gravity*. Springer, Dordrecht, 2011.

Feldman, H. A., Kaiser, N. and Peacock, J. A., Power-spectrum analysis of three-dimensional redshift surveys. *The Astrophysical Journal*, vol. 426, p. 23, 1994.

Fidler, C., Rampf, C., Tram, T., Crittenden, R., Koyama, K. and Wands, D., General relativistic corrections toN-body simulations and the Zel'dovich approximation. *Physical Review D*, vol. 92, no. 12, 2015.

Fidler, C., Sujata, N. and Rampf, C., A relativistic interpretation of bias in Newtonian simulations. *Journal of Cosmology and Astroparticle Physics*, vol. 2019, no. 02, p. 049–049, 2019.

Fidler, C., Tram, T., Rampf, C., Crittenden, R., Koyama, K. and Wands, D., Relativistic initial conditions for N-body simulations. *Journal of Cosmology and Astroparticle Physics*, vol. 2017, no. 06, p. 043, 2017.

Fixsen, D., The temperature of the cosmic microwave background. *The Astrophysical Journal*, vol. 707, no. 2, p. 916, 2009.

Frieden, B. R., *Physics from Fisher information: a unification*. Cambridge University Press, 1998.

Friedman, M., *Foundations of Space-Time Theories: Relativistic Physics and Philosophy of Science*. Princeton University Press, 1983.

Fry, J. N. and Gaztañaga, E., Biasing and hierarchical statistics in large-scale structure. *The Astrophysical Journal*, vol. 413, no. astro-ph/9302009, pp. 447–452, 1993.

Gil-Marín, H., Norena, J., Verde, L., Percival, W. J., Wagner, C., Manera, M. and Schneider, D. P., The power spectrum and bispectrum of SDSS DR11 BOSS galaxies–I. Bias and gravity. *Monthly Notices of the Royal Astronomical Society*, vol. 451, no. 1, pp. 539–580, 2015.

Gil-Marín, H., Verde, L., Norena, J., Cuesta, A. J., Samushia, L., Percival, W. J., Wagner, C., Manera, M. and Schneider, D. P., The power spectrum and bispectrum of SDSS DR11 BOSS galaxies–II. Cosmological interpretation. *Monthly Notices of the Royal Astronomical Society*, vol. 452, no. 2, pp. 1914–1921, 2015.

Gil-Marín, H., Wagner, C., Fragkoudi, F., Jimenez, R. and Verde, L., An improved fitting formula for the dark matter bispectrum. *Journal of Cosmology and Astroparticle Physics*, vol. 2012, no. 02, p. 047–047, 2012.

Giovannini, M., *A primer on the physics of the cosmic microwave background*. World Scientific, 2008.

Griffiths, D. J., Hyperfine splitting in the ground state of hydrogen. *American Journal of physics*, vol. 50, no. 8, pp. 698–703, 1982.

Guandalin, C., Adamek, J., Bull, P., Clarkson, C., Abramo, L. R. and Coates, L., Observing relativistic features in large-scale structure surveys – I. Multipoles of the power spectrum. *Monthly Notices of the Royal Astronomical Society*, vol. 501, no. 2, p. 2547–2561, 2020.

Guandalin, C., Carucci, I. P., Alonso, D. and Moodley, K., Clustering redshifts with the 21cm-galaxy cross-bispectrum. *arXiv preprint: 2112.05034*, (*submitted to MNRAS*), 2021.

Guth, A. H., Inflationary universe: A possible solution to the horizon and flatness problems. *Physical Review D*, vol. 23, no. 2, p. 347, 1981.

Guth, A. H., *The inflationary universe: the quest for a new theory of cosmic origins*. Basic Books, 1997.

Hamaus, N., Seljak, U., Desjacques, V., Smith, R. E. and Baldauf, T., Minimizing the stochasticity of halos in large-scale structure surveys. *Physical Review D*, vol. 82, no. 4, p. 043515, 2010.

Harrison, E., *Cosmology: the science of the universe*. Cambridge University Press, New York, 2000.

Harrison, E., *Masks of the Universe: changing ideas on the nature of the cosmos*. Cambridge University Press, New York, 2003.

Harrison, E. R., Fluctuations at the threshold of classical cosmology. *Physical review D*, vol. 1, no. 10, p. 2726, 1970.

Hearin, A. P., Zentner, A. R., van den Bosch, F. C., Campbell, D. and Tollerud, E., Introducing decorated HODs: modelling assembly bias in the galaxy–halo connection. *Monthly Notices of the Royal Astronomical Society*, vol. 460, no. 3, p. 2552–2570, 2016.

Heavens, A. F., Estimating non-Gaussianity in the microwave background. *Monthly Notices of the Royal Astronomical Society*, vol. 299, no. 3, pp. 805–808, 1998.

Heinesen, A., Multipole decomposition of the general luminosity distance Hubble law — a new framework for observational cosmology. *Journal of Cosmology and Astroparticle Physics*, vol. 2021, no. 05, p. 008, 2021.

Hubble, E., A relation between distance and radial velocity among extra-galactic nebulae. *Proceedings of the national academy of sciences*, vol. 15, no. 3, pp. 168–173, 1929.

Hubble, E. and Humason, M. L., The velocity-distance relation among extra-galactic nebulae. *The Astrophysical Journal*, vol. 74, p. 43, 1931.

Jain, B. and Bertschinger, E., Second order power spectrum and nonlinear evolution at high redshift. *The Astrophysical Journal*, vol. 431, pp. 495–505, 1994.

James, F., *Statistical methods in experimental physics*. World Scientific Publishing Company, 2006.

Jeong, D., *Cosmology with high ($z > 1$) redshift galaxy surveys*. Ph.D. thesis, Univ. Texas at Austin, 2010.

Kaiser, N., On the spatial correlations of Abell clusters. *The Astrophysical Journal: Letters*, vol. 284, pp. L9–L12, 1984.

Karagiannis, D., Lazanu, A., Liguori, M., Raccanelli, A., Bartolo, N. and Verde, L., Constraining primordial non-Gaussianity with bispectrum and power spectrum from upcoming optical and radio surveys. *Monthly Notices of the Royal Astronomical Society*, vol. 478, no. 1, pp. 1341–1376, 2018.

Kay, S. M., *Fundamentals of statistical signal processing*. Prentice Hall PTR, 1993.

Khoury, J., Ovrut, B. A., Steinhardt, P. J. and Turok, N., Ekpyrotic universe: colliding branes and the origin of the hot big bang. *Physical Review D*, vol. 64, no. 12, p. 123522, 2001.

Komatsu, E., Afshordi, N., Bartolo, N., Baumann, D., Bond, J. R., Buchbinder, E. I., Byrnes, C. T., Chen, X., Chung, D. J. H., Cooray, A., Creminelli, P., Dalal, N., Dore, O., Easther, R., Frolov, A. V., Gorski, K. M., Jackson, M. G., Khoury, J., Kinney, W. H., Kofman, L., Koyama, K., Leblond, L., Lehners, J. L., Lidsey, J. E., Liguori, M., Lim, E. A., Linde, A., Lyth, D. H., Maldacena, J., Matarrese, S., McAllister, L., McDonald, P., Mukohyama, S., Ovrut, B., Peiris, H. V., Raeth, C., Riotto, A., Rodriguez, Y., Sasaki, M., Scoccimarro, R., Seery, D., Sefusatti, E., Seljak, U., Senatore, L., Shandera, S., Shellard, E. P. S., Silverstein, E., Slosar, A., Smith, K. M., Starobinsky, A. A., Steinhardt, P. J., Takahashi, F., Tegmark, M., Tolley, A. J., Verde, L., Wandelt, B. D., Wands, D., Weinberg, S., Wyman, M., Yadav, A. P. S. and Zaldarriaga, M., Non-Gaussianity as a probe of the physics of the primordial universe and the astrophysics of the low redshift universe. *arXiv preprint: 0902.4759*, 2009.

Komatsu, E. and Spergel, D. N., Acoustic signatures in the primary microwave background bispectrum. *Physical Review D*, vol. 63, p. 063002, 2001.

Kovetz, E. D., Viero, M. P., Lidz, A., Newburgh, L., Rahman, M., Switzer, E., Kamionkowski, M., Aguirre, J., Alvarez, M., Bock, J., Bond, J. R., Bower, G., Bradford, C. M., Breysse, P. C., Bull, P., Chang, T.-C., Cheng, Y.-T., Chung, D., Cleary, K., Corray, A., Crites, A., Croft, R., Doré, O., Eastwood, M., Ferrara, A., Fonseca, J., Jacobs, D., Keating, G. K., Lagache, G., Lakhlani, G., Liu, A., Moodley, K., Murray, N., Pénin, A., Popping, G., Pullen, A., Reichers, D., Saito, S., Saliwanchik, B., Santos, M., Somerville, R., Stacey, G., Stein, G., Villaescusa-Navarro, F., Visbal, E., Weltman, A., Wolz, L. and Zemcov, M., Line-Intensity Mapping: 2017 Status Report, 2017.

Kragh, H., *Cosmology and controversy: the historical development of two theories of the Universe*. Princeton, Princeton University Press, 1996.

Kragh, H., Hubble Law or Hubble-Lemaître Law? The IAU Resolution, 2018.

Lazeyras, T., Wagner, C., Baldauf, T. and Schmidt, F., Precision measurement of the local bias of dark matter halos. *Journal of Cosmology and Astroparticle Physics*, vol. 2016, no. 02, p. 018–018, 2016.

Lee, S., Huff, E. M., Choi, A., Elvin-Poole, J., Hirata, C., Honscheid, K., MacCrann, N., Ross, A. J., Troxel, M. A., Eifler, T. F., Kong, H., Ferté, A., Blazek, J., Huterer, D., Campos, A., Chen, A., Dodelson, S., Lemos, P., Leonard, C. D., Miranda, V., Muir, J., Raveri, M., Secco, L. F., Weaverdyck, N., Zuntz, J., Bridle, S. L., Davis, C., DeRose, J., Gatti, M., Prat, J., Rau, M. M., Samuroff, S., Sánchez, C., Vielzeuf, P., Aguena, M., Allam, S., Amon, A., Andrade-Oliveira, F., Bernstein, G. M., Bertin, E., Brooks, D., Burke, D. L., Rosell,

A. C., Kind, M. C., Carretero, J., Castander, F. J., Cawthon, R., Conselice, C., Costanzi, M., da Costa, L. N., Pereira, M. E. S., Vicente, J. D., Desai, S., Diehl, H. T., Dietrich, J. P., Doel, P., Everett, S., Evrard, A. E., Ferrero, I., Fosalba, P., Frieman, J., García-Bellido, J., Gaztanaga, E., Gerdes, D. W., Giannantonio, T., Gruen, D., Gruendl, R. A., Gschwend, J., Gutierrez, G., Hartley, W. G., Hinton, S. R., Hollowood, D. L., Hoyle, B., James, D. J., Kuehn, K., Kuropatkin, N., Lahav, O., Lima, M., Maia, M. A. G., March, M., Marshall, J. L., Menanteau, F., Miquel, R., Mohr, J. J., Morgan, R., Palmese, A., Paz-Chinchón, F., Pieres, A., Malagón, A. A. P., Roodman, A., Sanchez, E., Scarpine, V., Schubnell, M., Serrano, S., Sevilla-Noarbe, I., Sheldon, E., Smith, M., Suchyta, E., Swanson, M. E. C., Tarle, G., Thomas, D., To, C., Varga, T. N. and Weller, J., Probing gravity with the DES-CMASS sample and BOSS spectroscopy, 2021.

Lemaître, G., Un Univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques. *Annales de la Société scientifique de Bruxelles*, vol. 47, pp. 49–59, 1927.

Lemaître, G., A homogeneous Universe of constant mass and increasing radius accounting for the radial velocity of extra-galactic nebulae. *Monthly Notices of the Royal Astronomical Society*, vol. 91, no. 5, pp. 483–490, 1931.

Lepori, F., Adamek, J., Durrer, R., Clarkson, C. and Coates, L., Weak-lensing observables in relativistic N-body simulations. *Monthly Notices of the Royal Astronomical Society*, vol. 497, no. 2, pp. 2078–2095, 2020.

Lima, M. and Hu, W., Photometric redshift requirements for self-calibration of cluster dark energy studies. *Physical Review D*, vol. 76, no. 12, 2007.

Lombriser, L., Yoo, J. and Koyama, K., Relativistic effects in galaxy clustering in a parametrized post-Friedmann universe. *Physical Review D*, vol. 87, p. 104019, 2013.

Lopez-Honorez, L., Mena, O. and Rigolin, S., Biases on cosmological parameters by general relativity effects. *Physical Review D*, vol. 85, no. 2, p. 023511, 2012.

Luongo, O., Muccino, M., Colgáin, E. O., Sheikh-Jabbari, M. M. and Yin, L., On larger H_0 values in the CMB dipole direction, 2021.

Lyth, D. H., Particle physics models of inflation. In: *Inflationary cosmology*, Springer, 2008, pp. 81–118.

Lyth, D. H. and Riotto, A., Particle physics models of inflation and the cosmological density perturbation. *Physics Reports*, vol. 314, no. 1-2, pp. 1–146, 1999.

Ma, Z., Hu, W. and Huterer, D., Effects of photometric redshift uncertainties on weak-lensing tomography. *The Astrophysical Journal*, vol. 636, no. 1, p. 21–29, 2006.

Macpherson, H. J. and Heinesen, A., Luminosity distance and anisotropic sky-sampling at low redshifts: A numerical relativity study. *Physical Review D*, vol. 104, no. 2, 2021.

Maggiore, M. and Riotto, A., The Halo Mass Function from Excursion Set Theory. I. Gaussian fluctuations with non-Markovian dependence on the smoothing scale. *The Astrophysical Journal*, vol. 711, no. 2, p. 907, 2010.

Maldacena, J., Non-Gaussian features of primordial fluctuations in single field inflationary models. *Journal of High Energy Physics*, vol. 2003, no. 05, p. 013, 2003.

Manchak, J. B., Can we know the global structure of spacetime? *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, vol. 40, no. 1, pp. 53–56, 2009.

Martínez, V. J. and Saar, E., *Statistics of the galaxy distribution*. CRC press, 2001.

Matarrese, S., Verde, L. and Heavens, A. F., Large-scale bias in the universe: bispectrum method. *Monthly Notices of the Royal Astronomical Society*, vol. 290, no. 4, pp. 651–662, 1997.

Matsubara, T., The correlation function in redshift space: general formula with wide-angle effects and cosmological distortions. *The Astrophysical Journal*, vol. 535, no. 1, p. 1, 2000.

McDonald, P., Gravitational redshift and other redshift-space distortions of the imaginary part of the power spectrum. *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 11, p. 026, 2009.

McQuinn, M. and White, M., On using angular cross-correlations to determine source redshift distributions. *Monthly Notices of the Royal Astronomical Society*, vol. 433, no. 4, p. 2857–2883, 2013.

Ménard, B., Scranton, R., Schmidt, S., Morrison, C., Jeong, D., Budavari, T. and Rahman, M., Clustering-based redshift estimation: method and application to data. *arXiv preprint: 1303.4722*, 2013.

Merson, A. I., Baugh, C. M., Helly, J. C., Gonzalez-Perez, V., Cole, S., Bielby, R., Norberg, P., Frenk, C. S., Benson, A. J., Bower, R. G. and et al., Lightcone mock catalogues from semi-analytic models of galaxy formation – I. Construction and application to the BzK colour selection. *Monthly Notices of the Royal Astronomical Society*, vol. 429, no. 1, pp. 556–578, 2012.

Migkas, K., Schellenberger, G., Reiprich, T. H., Pacaud, F., Ramos-Ceja, M. E. and Lovisari, L., Probing cosmic isotropy with a new X-ray galaxy cluster sample through the LX–T scaling relation. *Astronomy & Astrophysics*, vol. 636, p. A15, 2020.

Mo, H., Van den Bosch, F. and White, S., *Galaxy formation and evolution*. Cambridge University Press, 2010.

Modi, C., White, M., Slosar, A. and Castorina, E., Reconstructing large-scale structure with neutral hydrogen surveys. *Journal of Cosmology and Astroparticle Physics*, vol. 2019, no. 11, p. 023–023, 2019.

Montero-Dorta, A. D., Abramo, L. R., Granett, B. R., de la Torre, S. and Guzzo, L., The Multi-Tracer Optimal Estimator applied to VIPERS. *Monthly Notices of the Royal Astronomical Society*, vol. 493, no. 4, pp. 5257–5272, 2020.

Mukhanov, V., *Physical foundations of cosmology*. Cambridge University Press, 2005.

Mukhanov, V. and Winitzki, S., *Introduction to quantum effects in gravity*. Cambridge University Press, 2007.

Mukhanov, V. F., Feldman, H. A. and Brandenberger, R. H., Theory of cosmological perturbations. *Physics reports*, vol. 215, no. 5-6, pp. 203–333, 1992.

Nan, Y., Yamamoto, K. and Hikage, C., Higher multipoles of the galaxy bispectrum in redshift space. *Journal of Cosmology and Astroparticle Physics*, vol. 2018, no. 07, p. 038–038, 2018.

Nelson, D., Springel, V., Pillepich, A., Rodriguez-Gomez, V., Torrey, P., Genel, S., Vogelsberger, M., Pakmor, R., Marinacci, F., Weinberger, R., Kelley, L., Lovell, M., Diemer, B. and Hernquist, L., The IllustrisTNG Simulations: Public Data Release, 2021.

Newburgh, L. B., Bandura, K., Bucher, M. A., Chang, T.-C., Chiang, H. C., Cliche, J., Davé, R., Dobbs, M., Clarkson, C., Ganga, K. M. and et al., HIRAX: a probe of dark energy and radio transients. *Ground-based and Airborne Telescopes VI*, 2016.

Newman, J. A., Calibrating redshift distributions beyond spectroscopic limits with cross-correlations. *The Astrophysical Journal*, vol. 684, no. 1, p. 88, 2008.

Obuljen, A. and Percival, W. J., Anisotropic effective redshift and evolving clustering amplitude, 2021.

Ostriker, J. P. and Steinhardt, P. J., The observational case for a low-density Universe with a non-zero cosmological constant. *Nature*, vol. 377, no. 6550, pp. 600–602, 1995.

Peebles, P. J. and Yu, J., Primeval adiabatic perturbation in an expanding universe. *The Astrophysical Journal*, vol. 162, p. 815, 1970.

Peebles, P. J. E., Primeval helium abundance and the primeval fireball. *Physical Review Letters*, vol. 16, no. 10, p. 410, 1966.

Penrose, R., Before the Big Bang: an outrageous new perspective and its implications for particle physics. In: *Proceedings of EPAC*, 2006, pp. 2759–2763.

Penrose, R., The Big Bang and its dark-matter content: whence, whither, and wherefore. *Foundations of Physics*, vol. 48, no. 10, pp. 1177–1190, 2018.

Percival, W. J., Verde, L. and Peacock, J. A., Fourier analysis of luminosity-dependent galaxy clustering. *Monthly Notices of the Royal Astronomical Society*, vol. 347, no. 2, pp. 645–653, 2004.

Perivolaropoulos, L. and Skara, F., Challenges for ΛCDM: an update, 2021.

Perlmutter, S., Aldering, G., Goldhaber, G., Knop, R. A., Nugent, P., Castro, P. G., Deustua, S., Fabbro, S., Goobar, A., Groom, D. E. and et al., Measurements of Ω and Λ from 42 high-redshift supernovae. *The Astrophysical Journal*, vol. 517, no. 2, p. 565–586, 1999.

Peter, P. and Uzan, J.-P., *Primordial cosmology*. Oxford University Press, 2013.

Pillepich, A., Porciani, C. and Hahn, O., Halo mass function and scale-dependent bias from N-body simulations with non-Gaussian initial conditions. *Monthly Notices of the Royal Astronomical Society*, vol. 402, no. 1, pp. 191–206, 2010.

Plebanski, J. and Krasinski, A., *An introduction to general relativity and cosmology*. Cambridge University Press, 2006.

Pourtsidou, A., Bacon, D. and Crittenden, R., HI and cosmological constraints from intensity mapping, optical and CMB surveys. *Monthly Notices of the Royal Astronomical Society*, vol. 470, no. 4, p. 4251–4260, 2017.

Press, W. H. and Schechter, P., Formation of galaxies and clusters of galaxies by self-similar gravitational condensation. *The Astrophysical Journal*, vol. 187, pp. 425–438, 1974.

de Putter, R., Primordial physics from large-scale structure beyond the power spectrum. *arXiv preprint: 1802.06762*, 2018.

de Putter, R., Gleyzes, J. and Doré, O., Next non-Gaussianity frontier: What can a measurement with $\sigma(f_{\rm NL}) \leq 1$ tell us about multifield inflation? *Physical Review D*, vol. 95, no. 12, 2017.

Reimberg, P., Bernardeau, F. and Pitrou, C., Redshift-space distortions with wide angular separations. *Journal of Cosmology and Astroparticle Physics*, vol. 2016, no. 01, p. 048, 2016.

Renk, J., Zumalacarregui, M. and Montanari, F., Gravity at the horizon: on relativistic effects, CMB-LSS correlations and ultra-large scales in Horndeski's theory. *Journal of Cosmology and Astroparticle Physics*, vol. 2016, no. 07, p. 040, 2016.

Riess, A. G., Filippenko, A. V., Challis, P., Clocchiatti, A., Diercks, A., Garnavich, P. M., Gilliland, R. L., Hogan, C. J., Jha, S., Kirshner, R. P. et al., Observational evidence from supernovae for an accelerating universe and a cosmological constant. *The Astronomical Journal*, vol. 116, no. 3, p. 1009, 1998.

Saga, S., Taruya, A., Breton, M.-A. and Rasera, Y., Modelling the asymmetry of the halo cross-correlation function with relativistic effects at quasi-linear scales. *Monthly Notices of the Royal Astronomical Society*, vol. 498, no. 1, pp. 981–1001, 2020.

Santos, M. G., Bull, P., Alonso, D., Camera, S., Ferreira, P. G., Bernardi, G., Maartens, R., Viel, M., Villaescusa-Navarro, F., Abdalla, F. B. et al., Cosmology with a SKA HI intensity mapping survey. *arXiv preprint: 1501.03989*, 2015.

Santos, M. G., Cluver, M., Hilton, M., Jarvis, M., Jozsa, G. I. G., Leeuw, L., Smirnov, O., Taylor, R., Abdalla, F., Afonso, J., Alonso, D., Bacon, D., Bassett, B. A., Bernardi, G., Bull, P., Camera, S., Chiang, H. C., Colafrancesco, S., Ferreira, P. G., Fonseca, J., van der Heyden, K., Heywood, I., Knowles, K., Lochner, M., Ma, Y.-Z., Maartens, R., Makhathini, S., Moodley, K., Pourtsidou, A., Prescott, M., Sievers, J., Spekkens, K., Vaccari, M., Weltman, A., Whittam, I., Witzemann, A., Wolz, L. and Zwart, J. T. L., MeerKLASS: MeerKAT Large Area Synoptic Survey, 2017.

Schuller, F., Lecture 20: Cosmology – The early epoch (International Winter School on Gravity and Light 2015), 2015.

Scoccimarro, R., The bispectrum: from theory to observations. *The Astrophysical Journal*, vol. 544, no. 2, p. 597, 2000.

Scoccimarro, R., Fast estimators for redshift-space clustering. *Physical Review D*, vol. 92, no. 8, p. 083532, 2015.

Scoccimarro, R., Colombi, S., Fry, J. N., Frieman, J. A., Hivon, E. and Melott, A., Nonlinear evolution of the bispectrum of cosmological perturbations. *The Astrophysical Journal*, vol. 496, no. 2, p. 586, 1998.

Sefusatti, E. and Komatsu, E., Bispectrum of galaxies from high-redshift galaxy surveys: Primordial non-Gaussianity and nonlinear galaxy bias. *Physical Review D*, vol. 76, no. 8, p. 083004, 2007.

Seljak, U., Extracting primordial non-Gaussianity without cosmic variance. *Physical Review Letters*, vol. 102, no. 2, p. 021302, 2009.

Senatore, L., Smith, K. M. and Zaldarriaga, M., Non-Gaussianities in single field inflation and their optimal limits from the WMAP 5-year data. *Journal of Cosmology and Astroparticle Physics*, vol. 2010, no. 01, p. 028, 2010.

Seo, H.-J. and Hirata, C. M., The foreground wedge and 21-cm BAO surveys. *Monthly Notices of the Royal Astronomical Society*, vol. 456, no. 3, p. 3142–3156, 2016.

Sheth, R. K. and Tormen, G., Large-scale bias and the peak background split. *Monthly Notices of the Royal Astronomical Society*, vol. 308, no. 1, pp. 119–126, 1999.

Silva, C. Q. A., *Redshifts fotométricos de quasares em levantamentos de bandas estreitas*. Ph.D. thesis, Instituto de Física, Universidade de São Paulo, São Paulo, 2021.

da Silva, M. A., *Cosmologia usando aglomerados de galáxias no Dark Energy Survey*. Master's thesis, Instituto de Física, Universidade de São Paulo, São Paulo, 2017.

Slosar, A., Hirata, C., Seljak, U., Ho, S. and Padmanabhan, N., Constraints on local primordial non-Gaussianity from large-scale structure. *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 08, p. 031, 2008.

Smeenk, C., Philosophy of Cosmology. In: R. Batterman (Editor), *Oxford Handbook of Philosophy of Physics*, Oxford: Oxford University Press, 2013, pp. 607–652.

Smith, A., Cole, S., Baugh, C., Zheng, Z., Angulo, R., Norberg, P. and Zehavi, I., A lightcone catalogue from the Millennium-XXL simulation. *Monthly Notices of the Royal Astronomical Society*, vol. 470, no. 4, pp. 4646–4661, 2017.

Smith, R. W., *The expanding universe: Astronomy's 'great debate', 1900-1931*. Cambridge University Press, 1982.

Sparke, L. S. and Gallagher, J. S., *Galaxies in the universe: an introduction*. Cambridge University Press, 2007.

Stewart, J. M., *Advanced general relativity*. Cambridge University Press, 1993.

Stoeger, W. R., Maartens, R. and Ellis, G., Proving almost-homogeneity of the universe: an almost Ehlers-Geren-Sachs theorem. *The Astrophysical Journal*, vol. 443, pp. 1–5, 1995.

Strauss, M. A. and Willick, J. A., The density and peculiar velocity fields of nearby galaxies. *Physics Reports*, vol. 261, no. 5-6, p. 271–431, 1995.

Szapudi, I., Introduction to higher order spatial statistics in cosmology. In: *Data Analysis in Cosmology*, Springer, 2008, pp. 457–492.

Tansella, V., Bonvin, C., Durrer, R., Ghosh, B. and Sellentin, E., The full-sky relativistic correlation function and power spectrum of galaxy number counts. Part I: theoretical aspects. *Journal of Cosmology and Astroparticle Physics*, vol. 2018, no. 03, p. 019, 2018.

Tegmark, M., Hamilton, A. J., Strauss, M. A., Vogeley, M. S. and Szalay, A. S., Measuring the galaxy power spectrum with future redshift surveys. *The Astrophysical Journal*, vol. 499, no. 2, p. 555, 1998.

Tegmark, M., Taylor, A. N. and Heavens, A. F., Karhunen-Loève eigenvalue problems in cosmology: how should we tackle large data sets? *The Astrophysical Journal*, vol. 480, no. 1, pp. 22–35, 1997.

Tinker, J., Kravtsov, A. V., Klypin, A., Abazajian, K., Warren, M., Yepes, G., Gottlöber, S. and Holz, D. E., Toward a halo mass function for precision cosmology: the limits of universality. *The Astrophysical Journal*, vol. 688, no. 2, pp. 709–728, 2008.

Tinker, J. L., Robertson, B. E., Kravtsov, A. V., Klypin, A., Warren, M. S., Yepes, G. and Gottlöber, S., The large scale bias of dark matter halos: numerical calibration and model tests. *The Astrophysical Journal*, vol. 724, no. 2, pp. 878–886, 2010.

Umeh, O., Jolicoeur, S., Maartens, R. and Clarkson, C., A general relativistic signature in the galaxy bispectrum: the local effects of observing on the lightcone. *Journal of Cosmology and Astroparticle Physics*, vol. 2017, no. 03, p. 034, 2017.

Voivodic, R., Lima, M. and Abramo, L. R., Excursion Set Halos – ExSHalos: a new parameter free method for fast generation of halo catalogues. *arXiv preprint:1906.06630*, 2019.

Wang, M. S., Beutler, F. and Bacon, D., Impact of relativistic effects on the primordial non-Gaussianity signature in the large-scale clustering of quasars. *Monthly Notices of the Royal Astronomical Society*, staa2998, 2020.

Watkinson, C. A., Majumdar, S., Pritchard, J. R. and Mondal, R., A fast estimator for the bispectrum and beyond–a practical method for measuring non-Gaussianity in 21-cm maps. *Monthly Notices of the Royal Astronomical Society*, vol. 472, no. 2, pp. 2436–2446, 2017.

Weinberg, S., *Gravitation and cosmology: principles and applications of the general theory of relativity*. Wiley, 1972.

Wilson, M. J., Peacock, J. A., Taylor, A. N. and de la Torre, S., Rapid modelling of the redshift-space power spectrum multipoles for a masked density field. *Monthly Notices of the Royal Astronomical Society*, vol. 464, no. 3, pp. 3121–3130, 2016.

Witzemann, A., Alonso, D., Fonseca, J. and Santos, M. G., Simulated multitracer analyses with HI intensity mapping. *Monthly Notices of the Royal Astronomical Society*, vol. 485, no. 4, pp. 5519–5531, 2019.

Woodbury, M. A., *Inverting modified matrices*. Statistical Research Group, 1950.

Yadav, A. P. S., Komatsu, E. and Wandelt, B. D., Fast estimator of primordial non-Gaussianity from temperature and polarization anisotropies in the cosmic microwave background. *The Astrophysical Journal*, vol. 664, no. 2, p. 680–686, 2007.

Yamamoto, K., Nakamichi, M., Kamino, A., Bassett, B. A. and Nishioka, H., A measurement of the quadrupole power spectrum in the clustering of the 2dF QSO survey. *Publications of the Astronomical Society of Japan*, vol. 58, no. 1, pp. 93–102, 2006.

Yamamoto, K., Nishioka, H. and Taruya, A., The effect of bias and redshift distortions on a geometric test for the cosmological constant. *arXiv preprint: astro-ph/0012433*, 2000.

Yamauchi, D., Yokoyama, S. and Takahashi, K., Multitracer technique for galaxy bispectrum: An application to constraints on nonlocal primordial non-Gaussianities. *Physical Review D*, vol. 95, no. 6, p. 063530, 2017.

Yoo, J., General relativistic description of the observed galaxy power spectrum: do we understand what we measure? *Physical Review D*, vol. 82, no. 8, p. 083508, 2010.

Yoo, J., Relativistic effect in galaxy clustering. *Classical and Quantum Gravity*, vol. 31, no. 23, p. 234001, 2014.

Zeldovich, Y. B., A hypothesis, unifying the structure and the entropy of the Universe. *Monthly Notices of the Royal Astronomical Society*, vol. 160, no. 1, pp. 1P–3P, 1972.

Zhang, Y., Pullen, A. R. and Maniyar, A. S., Joint analyses of 2D CMB lensing and 3D galaxy clustering in spherical Fourier-Bessel basis, 2021.

Zhao, G.-B., Wang, Y., Taruya, A., Zhang, W., Gil-Marín, H., de Mattia, A., Ross, A. J., Raichoor, A., Zhao, C., Percival, W. J., Alam, S., Bautista, J. E., Burtin, E., Chuang, C.-H., Dawson, K. S., Hou, J., Kneib, J.-P., Koyama, K., du Mas des Bourboux, H., Mueller, E.-M., Newman, J. A., Peacock, J. A., Rossi, G., Ruhlmann-Kleider, V., Schneider, D. P. and Shafieloo, A., The completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: a multitracer analysis in Fourier space for measuring the cosmic structure growth and expansion rate. *Monthly Notices of the Royal Astronomical Society*, vol. 504, no. 1, p. 33–52, 2021.

Zwicky, F., Die rotverschiebung von extragalaktischen nebeln. *Helvetica physica acta*, vol. 6, pp. 110–127, 1933.

Zwicky, F., On the masses of nebulae and of clusters of nebulae. *The Astrophysical Journal*, vol. 86, p. 217, 1937.

Zwicky, F., Republication of: The redshift of extragalactic nebulae. *General Relativity and Gravitation*, vol. 41, no. 1, pp. 207–224, 2009.

APPENDICES

Appendix A

Derivation of Fundamental Equations

In this appendix we derive some of the main equations of cosmology, after the cosmological principle is assumed. We go through a basic overview of general relativity, then we discuss the symmetries the cosmological principle imposes, which leads us to the FLRW metric and, within this class of models, we state the different distances used in cosmology.

A.1 Geometric description of gravity

Cosmology, as dealt in this work, was able to flourish with the development of Einstein's theory of general relativity. While classical physics made allusion to absolute motions, Einstein's views that we are only able to observe the relative motion of physical objects was a crucial step for him to renounce the notions of *absolute velocity* and *absolute rest* (Friedman, 1983), which eventually led him to introduce the special principle of relativity:

> "*If, relative to K, K′ is a uniformly moving coordinate system devoid of rotation, then natural phenomena run their course with respect to K′ according to exactly the same general laws as with respect to K.*"
>
> (Einstein, 1920, p. 15)

This principle was still incomplete, as it discriminated among inertial and accelerated systems. Therefore, since we perceive acceleration in relation to the motion of physical entities, this notion should also be abandoned. Dropping this distinction leads us to the general principle of relativity, in which all states of motion are equivalent (Friedman, 1983, p. 5):

> "*All bodies of reference K, K′, etc., are equivalent for the description of natural phenomena (formulation of the general laws of nature), whatever may be their state of motion.*"
>
> (Einstein, 1920, p. 72)

More formally, this is called the *principle of general covariance* or, from a mathematical perspective, *diffeomorphism invariance*. This last definition is the mathematical statement that physical laws should be the same regardless of the coordinate system chosen and we shall return to it in Appendix B.

To generalise the special principle, Einstein's theory of general relativity is based upon the generalisation of Euclidean spaces commonly used in almost every sub-field of physics. We quote the following definition of the underlying spacetime assumed in general relativity:

> "*Space-time is a four-dimensional topological manifold with a smooth atlas carrying a torsion-free connection compatible with a Lorentzian metric and a time orientation satisfying the Einstein equations*."
>
> (Schuller, 2015)

A topology $\mathcal{O}$ is the simplest structure one can add to a set $\mathcal{M}$ allowing for the definition of continuity. A topological space $(\mathcal{M}, \mathcal{O})$, like the surface of a sphere $\mathbb{S}^2$, can be charted trough a map $\phi : \mathcal{M} \to \mathbb{R}^n$, just like we set latitude and longitude to map $\mathbb{S}^2$ into a plane piece of paper $\mathbb{R}^2$ (Stewart, 1993). The chart ϕ is known as a coordinate system. A collection of many C^∞-charts[1] is known as a smooth atlas $\mathcal{A}$. Different smooth manifolds carry different connections ∇, which allows us to study the parallel transport of vector fields defined on the manifold. For example, the surface of an ellipsoid has the same structure as $\mathbb{S}^2$, but they have different connections, i.e. vector fields change differently along their surface. Although ∇ is not a tensorial quantity, depending upon the basis chosen to parametrize the vector fields through its *components* Γ, it carries two tensorial information: the torsion and the Riemann curvature tensor $R^\alpha{}_{\beta\gamma\delta}$. In the case of a torsion-free connection, $R^\alpha{}_{\beta\gamma\delta}$ suffices to determine the connection. However, what determines the curvature tensor will be the matter content, described by the energy-momentum tensor $T_{\mu\nu}$, through the spacetime metric $g_{\mu\nu}$. The latter allows us to assign magnitudes to vectors in the tangent spaces of the 4-dimensional topological manifold $(\mathcal{M}, \mathcal{O})$.

Just to complete the definitions, the time orientation is determined by a non-vanishing vector field T, for which $g(T, T) \equiv T \cdot T > 0$, defining events p that are future directed and past directed on the spacetime manifold. Finally, a Lorentzian metric $g_{\mu\nu}$ has signature $(-,+,+,+)$ or $(+,-,-,-)$. This means that defining distances on a generic manifold can be done via the line element

$$\mathrm{d}s^2 = \sum_{\mu\nu} g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu \equiv g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu. \tag{A.1}$$

For the Euclidean (flat) manifold, $g_{\mu\nu} = \mathrm{diag}(1, 1, \ldots, 1)$ and g is called a Riemannian metric. Concentrating only on 3 dimensions, this reduces to the Pythagorean relation $\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}y^2 +$

[1] For a C^∞-function, all of its derivatives exits and are continuous.

dz^2, and $g_{\mu\nu} = \text{diag}(1, 1, 1)$. If we change to a non-Euclidean structure, allowing for curvature, then the new line element $d\tilde{s}^2 = \tilde{g}_{\mu\nu} dx^\mu dx^\nu$ is such that there are no coordinates whatsoever that can put $\tilde{g}_{\mu\nu}$ in the form $\text{diag}(1, 1, \ldots, 1)$ (Friedman, 1983, p. 10). Lorentzian metrics will determine the causal structure of spacetime by defining the so-called light cones.

We will not go through all the derivations involved; instead, we merely state that Einstein's equations,

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = \frac{8\pi G}{c^4} T_{\mu\nu}, \tag{A.2}$$

which describes the dynamics of relativistic matter and gravitation, satisfies the general principle of relativity. Because of the interplay between the energy-momentum tensor of matter $T_{\mu\nu}$ and the metric, we say that general relativity is a geometric theory of gravitation as the geometry of spacetime itself will be subjected to the dynamics induced by matter.

It is interesting to point out that the geometric description is not exclusive of general relativity. One can think about spacetime from a Newtonian perspective: as a 4-dimensional manifold $(\mathcal{M}, \mathcal{O}, \mathcal{A}, \nabla)$ with a smooth map $t : \mathcal{M} \to \mathbb{R}$ with a non-vanishing gradient $dt \neq 0$, for all points of spacetime, that we call absolute time (Stewart, 1993, Section 1.11).

A.2 Symmetries

Einstein's equations (A.2) are highly non-trivial to solve. In order to obtain analytical solutions thus, it is possible to rely on symmetry properties of the metric to solve them. There are two routes to simplify the Equations (A.2): one either knows the metric and look for solutions, or one imposes the desired symmetries and find what is the metric that has those properties, and then solves the equations. This second avenue is the one usually adopted in cosmology: some set of fundamental principles, such as the cosmological one (as done in Appendix A.3), tell us what are the desired symmetries, we find the metric and energy-momentum tensor that satisfy those properties, and then solve the equations.

Therefore, we aim at describing the symmetries of the metric. To do that, one needs a set of technical definitions that are beyond the scope of this work. We address the reader to Sections 1.5 and 1.6 of Stewart (1993) for a complete discussion. In this section, we give an overall qualitative picture of the so-called Lie derivative $\mathcal{L}$, used to define the flows of some vector field X and by pulling back the metric along this flow, i.e. $\mathcal{L}_X g$. If the metric does not change, then it will be conserved. Finally, if the vector field is associated with some symmetry property, then "conservation" of the metric with respect to this *Killing vector field* implies that the metric has the underlying symmetry (Schuller, 2015).

What are these Killing vector fields and how they are related to a set of symmetry properties will become clearer in the next section, where we apply this idea to find what metric satisfies

a set of symmetries described by the vector fields that generate rotations and translations, that is, what metric carries the symmetries spatial isotropy and homogeneity.

A.3 Friedmann–Lemaître–Robertson–Walker universes

The Friedmann–Lemaître–Robertson–Walker (FLRW) metric is based on the cosmological principle. In order to derive it, we require two spatial symmetries: translation and rotation. The translational symmetry is associated with the homogeneity assumption of the principle, whereas the rotational symmetry is related to isotropy. It is interesting to mention the theorem which states that if all points in a spacelike hypersurface are spherically symmetric, meaning that the metric structure is preserved under rotations about every point, then homogeneity follows as a consequence (Weinberg, 1972; Butterfield, 2014).

These symmetries are described by the corresponding algebras, and in this case they lead us to a maximally symmetric 3-dimensional space. Hence, there will be one Killing vector field for each spatial dimensional, for both rotations and translations, named J_a and P_a, respectively. These vector fields satisfy the commutation relations (Schuller, 2015)

$$\begin{aligned} [J_a, J_b] &= \varepsilon_{abc}\, J_c, \\ [P_a, P_b] &= 0, \\ [J_a, P_b] &= \varepsilon_{abc}\, P_c, \end{aligned} \tag{A.3}$$

where J_a and P_a can also be identified with the generators of rotations and translations, and a, b, c run over the three spatial dimensions (for example, see Plebanski and Krasinski, 2006, Section 10.7).

It is by looking at these Killing vector fields that we can select the hypersurfaces that look homogeneous and isotropic. Therefore, observers outside these vector fields may not observe an universe that follows the cosmological principle. This is crucial to understand the so-called CMB frame, and how it can be used as a hypothesis check for the principle (Schuller, 2015).

Preserving the metric structure under some transformation is equivalent of saying that there will be Killing vectors J_a, or P_a, for which

$$\mathcal{L}_{J_a} g = \mathcal{L}_{P_a} g = 0. \tag{A.4}$$

Therefore, after the symmetries are fixed, we can find the corresponding metric of the underlying space. Analogously, if the metric is known, we can find the corresponding symmetries. Therefore, the FLRW metric can be obtained by solving the Equations (A.4) for the vector fields of Equation (A.3), which impose the desired symmetries (see Chapter 8 of Plebanski and

Krasinski, 2006, for a complete discussion).

After finding the *spatial* metric obtained from the *spatial* symmetries imposed by Equation (A.3), the spatial part of the Riemann tensor is (Peter and Uzan, 2013; Durrer, 2020)

$$R^{i}{}_{jk\ell} = \kappa\,(\delta^{i}{}_{k}\gamma_{j\ell} - \delta^{i}_{\ell}\gamma_{jk}), \tag{A.5}$$

where

$$\gamma_{ij}(r,\theta,\phi) = \begin{pmatrix} \frac{1}{1-\kappa r^2} & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2\sin\theta \end{pmatrix} \tag{A.6}$$

is the 3-dimensional metric describing an open (hyperbolic space, $\kappa < 0$), flat (Euclidean space, $\kappa = 0$) or closed (spherical space, $\kappa > 0$) universe, after we chose a chart: in this case the one that represents the spherical coordinate system described by the variables r, θ and ϕ.

For the 4-dimensional manifold describing spacetime, the left freedom is to add a temporal function, say $a(t)$, to the spatial part, which allows the spacelike hypersurfaces to expand or contract equally in each dimension[2]:

$$g_{\mu\nu}(t,r,\theta,\phi) = \begin{pmatrix} -c^2 & 0 & 0 & 0 \\ 0 & \frac{a^2(t)}{1-\kappa r^2} & 0 & 0 \\ 0 & 0 & a^2(t)r^2 & 0 \\ 0 & 0 & 0 & a^2(t)r^2\sin\theta \end{pmatrix}. \tag{A.7}$$

It is a common convention in cosmology to set the scale factor $a(t)$ to unity today, i.e. $a(t_0) = 1$. This can be done thanks to the rescaling symmetries of the FLRW metric: changing $a \to \lambda\, a$, $r \to \lambda^{-1}\, r$ and $k \to \lambda^2\, k$ simultaneously leave the metric unchanged (Baumann, 2014). However, this means that by setting $a(t_0) = 1$, we cannot force $\kappa = \pm 1$. We will adopt this convention and treat $a(t)$ as a dimensionless quantity. Therefore, the constant κ inherits dimension of $1/L^2$ and it is given by the spatial Ricci scalar ${}^3R_0$, which defines the curvature scale at the present time:

$$\kappa = \frac{\pm 1}{{}^3R_0^2}, \qquad {}^3R_0 \equiv \frac{c}{a_0\, H_0\, \sqrt{|\Omega_k^0|}}, \tag{A.8}$$

with $a_0 = a(t_0) = 1$, H_0 the Hubble constant today and Ω_k^0 the present energy density in the form of curvature (Ellis et al., 2012, p. 206).

[2] An interesting metric is the Bianchi type I (Plebanski and Krasinski, 2006, Chapter 10), where $ds^2 = -dt^2 + a_x^2(t)\,\mathrm{d}x^2 + a_y^2(t)\,\mathrm{d}y^2 + a_z^2(t)\,\mathrm{d}z^2$, that is, each spatial dimension can expand in a different rate. This model is homogeneous, but anisotropic.

We can also write Equation (A.7) in terms of the conformal time,

$$\tau \equiv \int \frac{\mathrm{d}t}{a(t)}, \tag{A.9}$$

which is used to make the metric conformally Minkowski when the spatial curvature κ vanishes:

$$g_{\mu\nu}(\tau, r, \theta, \phi) = a^2(\tau) \begin{pmatrix} -c^2 & 0 & 0 & 0 \\ 0 & \frac{1}{1-\kappa r^2} & 0 & 0 \\ 0 & 0 & r^2 & 0 \\ 0 & 0 & 0 & r^2 \sin\theta \end{pmatrix}. \tag{A.10}$$

With this metric, the line element of Equation (A.1) is written as

$$\mathrm{d}s^2 = a^2(\tau) \left[-c^2 \mathrm{d}\tau^2 + \frac{1}{1-\kappa r^2} \mathrm{d}r^2 + r^2 \mathrm{d}\theta^2 + \sin^2(\theta)\, \mathrm{d}\phi^2 \right]. \tag{A.11}$$

To illustrate the choice of charts, we show the metric defined by the so-called hyperspherical coordinates:

$$\mathrm{d}s^2 = a^2(\tau) \left[-c^2 \mathrm{d}\tau^2 + \mathrm{d}r^2 + S_k^2(r) \left(\mathrm{d}\theta^2 + \sin^2(\theta)\, \mathrm{d}\phi^2 \right) \right], \tag{A.12}$$

with

$$S_k(r) = \begin{cases} \kappa^{-1/2} \sin(r\sqrt{\kappa}), & \kappa > 0 \\ r, & \kappa = 0 \\ |\kappa|^{-1/2} \sinh(r\sqrt{|\kappa|}), & \kappa < 0 \end{cases}. \tag{A.13}$$

From now on we consider natural units where $c = 1$. The non-vanishing connection coefficients for the metric in Equation (A.7) are

$$\Gamma^i{}_{j0} = \frac{\dot{a}}{a} \delta^i{}_j \equiv H\, \delta^i{}_j, \quad \Gamma^0{}_{ij} = \dot{a}\, a\, \gamma_{ij}, \quad \Gamma^k{}_{ij} = \frac{1}{2} \gamma^{km} \left(\gamma_{im,j} + \gamma_{jm,i} - \gamma_{ij,m} \right), \tag{A.14}$$

These coefficients change trivially to conformal time by employing the relation $a(t)\, \mathrm{d}\tau = \mathrm{d}t$. The Ricci curvature tensor $R_{\mu\nu}$ and Ricci scalar R are, respectively,

$$R_{00} = -3\frac{\ddot{a}}{a}, \quad R_{ij} = \left[\ddot{a}\, a + 2(\dot{a}^2 + \kappa) \right] \gamma_{ij}, \tag{A.15}$$

and

$$R = 6 \left(\frac{\ddot{a}}{a} + H^2 + \frac{\kappa}{a^2} \right). \tag{A.16}$$

Finally, we can compute the Einstein tensor $G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R$, whose only non-

vanishing components are

$$G_{00} = 3\left(H^2 + \frac{k}{a^2}\right), \quad G_{ij} = -\left(\dot{a}^2 + 2\,\ddot{a}\,a + \kappa\right)\gamma_{ij}. \tag{A.17}$$

These quantities have been obtained through the symmetry properties of the FLRW metric. The only undefined variable is the function $a(t)$, which will be the solution of Einstein's equations after we find the matter content that satisfies the underlying symmetries. We address the allowed energy-momentum tensor and the final form of Einstein's equations for a homogeneous and isotropic universe in what follows.

A.3.1 Friedmann equations

Besides the assumptions of homogeneity and isotropy, which led us to the FLRW metric, shown in Equations (A.7) and (A.10), to finally derive the equations used within the ΛCDM model we need to consider a final ingredient: the matter/energy content of the universe that, in order to keep self-consistency with the FLRW metric, will be described by isotropic perfect fluids, that is, fluids whose energy-momentum tensor is completely described by its rest frame energy density ρ and an isotropic pressure P:

$$T_{\mu\nu} = [\rho(t) + P(t)]\,u_\mu\,u_\nu - P(t)\,g_{\mu\nu} \tag{A.18}$$

As required by the isotropy assumption, ρ and P are functions of time only. Also, the timelike vector field associated to the fluid element at each point of spacetime is $u^\mu = (-1, 0, 0, 0)$, again from the assumption of an universe that is isotropic everywhere.

With Equation (A.17), Einstein's equations become

$$H^2 + \frac{\kappa}{a^2} = \frac{8\pi G}{3}\rho \tag{A.19}$$

and

$$H^2 + 2\,\frac{\ddot{a}}{a} + \frac{\kappa}{a^2} = 8\pi G\,P \tag{A.20}$$

Combining Equations (A.19) and (A.20), we arrive at

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}[\rho + 3P]. \tag{A.21}$$

Equations (A.19) and (A.21) are commonly known as Friedmann's equations.

Friedmann equations in conformal time Recalling that $\mathrm{d}\tau = a^{-1}\mathrm{d}t \Rightarrow \mathrm{d}/\mathrm{d}t = a^{-1}\mathrm{d}/\mathrm{d}\tau$, we define the conformal Hubble factor:

$$\mathcal{H} \equiv aH, \quad \mathcal{H} = \frac{a'(\tau)}{a(\tau)}. \tag{A.22}$$

Notice that $\mathcal{H}^{-1} = (aH)^{-1}$ is the so-called *comoving Hubble horizon*. Hence, the first Friedmann equation (Equation A.19) becomes

$$\mathcal{H}^2 = \frac{a^2}{3}\rho - k. \tag{A.23}$$

The second Friedmann equation (Equation A.21) can be written as

$$\ddot{a} = \frac{1}{a}\frac{\mathrm{d}}{\mathrm{d}\tau}\left(\frac{1}{a}\frac{\mathrm{d}a}{\mathrm{d}\tau}\right) = \frac{1}{a(\tau)}\left[\frac{1}{a(\tau)}\frac{\mathrm{d}^2a}{\mathrm{d}\tau^2} - \frac{1}{a^2(\tau)}\left(\frac{\mathrm{d}a}{\mathrm{d}\tau}\right)^2\right]. \tag{A.24}$$

Because

$$\mathcal{H}' = \frac{\mathrm{d}}{\mathrm{d}\tau}\left(\frac{a'}{a}\right) = \frac{a''(\tau)}{a(\tau)} - \frac{a'(\tau)^2}{a^2(\tau)} = \frac{a''(\tau)}{a(\tau)} - \mathcal{H}^2, \tag{A.25}$$

and hence

$$\ddot{a}(t) = \frac{\mathcal{H}'}{a(\tau)}, \tag{A.26}$$

we arrive at

$$\mathcal{H}' = \frac{a^2(\tau)}{2}\left[P(\tau) - \frac{1}{3}\rho(\tau)\right], \tag{A.27}$$

or equivalently

$$\frac{a''}{a} - \mathcal{H}^2 = \frac{a^2}{2}\left(P - \frac{1}{3}\rho\right). \tag{A.28}$$

Energy-momentum conservation Finally, conservation of the energy-momentum tensor,

$$\nabla_\mu T^{\mu\nu} = 0, \tag{A.29}$$

leads to the continuity equation

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = -3H(\rho + P). \tag{A.30}$$

Equation of state Equation (A.18) can be further simplified if one gives an equation of state (EoS), given by thermodynamical considerations, that relates the pressure and the density of a physical system. In turn, this will also allow us to simplify Equation (A.21). For a fluid with trace-free energy-momentum tensor, as it is the case for ultra-relativistic particles (*radiation*), $P = \rho/3$; for a scalar-field ϕ dominated by its potential energy, $P \approx -\rho$. Therefore, we usually write the linear relation $P = w\rho$, where w is called the EoS for the fluid. For pressureless matter, also called *dust*, $w = 0$. Finally, when dark energy can be treated as a cosmological constant,

$w = -1$ exactly.

A.3.2 Single-component solutions

If we assume the universe to be dominated by different types of "matter" (e.g. radiation, cold dark matter or dark energy) such that at different epochs one dominates the evolution history over the others, we are able to obtain analytical solutions for the scale factor.

In this single-component limit and using the relation $P = w\rho$, the Equation (A.30) gives

$$\rho(t) = \rho_0 \left(\frac{a(t)}{a_0} \right)^{-3(1+w)}, \tag{A.31}$$

where ρ_0 is the energy density of some component whose EoS is w today. In terms of the scale factor, we can also write

$$\rho(a) \propto a^{-3(1+w)}. \tag{A.32}$$

This change allows us to simplify, for the case of a flat ($k = 0$) universe, the first Friedmann equation in conformal time:

$$a' \propto a^{\frac{1}{2}(1-3w)}. \tag{A.33}$$

In what follows, we analyse the solution of the Einstein's equations for a universe that obeys the cosmological principle, dominated by a single type of energy density whose energy density evolves as in Equation (A.31).

Dust (matter) We start by considering how the energy density of dust evolves. For dust ($w = 0$), Equation (A.32) implies that $\rho_\mathrm{m} \propto a^{-3}$. This means that with the cosmic expansion, the energy density of matter gets diluted in proportion to the "volume" of the universe: if we consider the Newtonian analogue of an expanding sphere whose radius expand according to $R(t) = a(t)R_0$, where R_0 is the initial radius of the sphere, than the matter contained inside the volume $V(t) \propto R^3(t) = a^3 R_0$ dilutes as a^{-3}. To find the scale factor, we integrate Equation (A.33) for the $w = 0$ case:

$$\int a^{-1/2} \frac{\mathrm{d}a}{\mathrm{d}\tau} \mathrm{d}\tau \sim \int \mathrm{d}\tau \Rightarrow a^{1/2} \sim \tau. \tag{A.34}$$

Therefore, the solution for the scale factor as a function of conformal time, in a matter dominated universe, is given by

$$a(\tau) \sim \tau^2. \tag{A.35}$$

In terms of the cosmic time t, it is straightforward to show that $a(t) \sim t^{2/3}$, and hence $\rho(t) \sim t^{-2}$. Therefore, as $t \to 0, \rho \to \infty$. At this point, Einstein's equations are ill-defined and we reach an

initial singularity. We shall call this the "Big Bang".

Radiation For radiation ($w = 1/3$), $\rho_\gamma \propto a^{-4}$. The extra factor of $a(t)$ diluting the radiation energy density comes from the fact that, as the universe expand, the photon's wavelength also stretches, following the relation $\lambda \propto a$. Therefore, the number of photons decay proportionally to the "volume", while its energy decreases due to the extra wavelength factor. Analogous to what has been done for matter, it is straightforward to show that $a(\tau) \sim \tau$, or equivalently $a(t) \sim t^{1/2}$. Again, $\rho(t) \sim t^{-2}$ and we reach the same Big Bang singularity as $t \to 0$.

Dark energy For the case of dark energy, $w = -1$ and $\rho_\Lambda = \rho_{\Lambda,0}$ is constant in time. The scale factor shows a distinct behaviour than what has been obtained for matter and radiation, which had a decelerated expansion: in this case, $a(t) \sim \mathrm{e}^{Ht}$ and no singularity appears for $t \to 0$. In terms of the conformal time, $a(\tau) = -1/\tau$. We summarise what we obtained so far in Figure A.1.

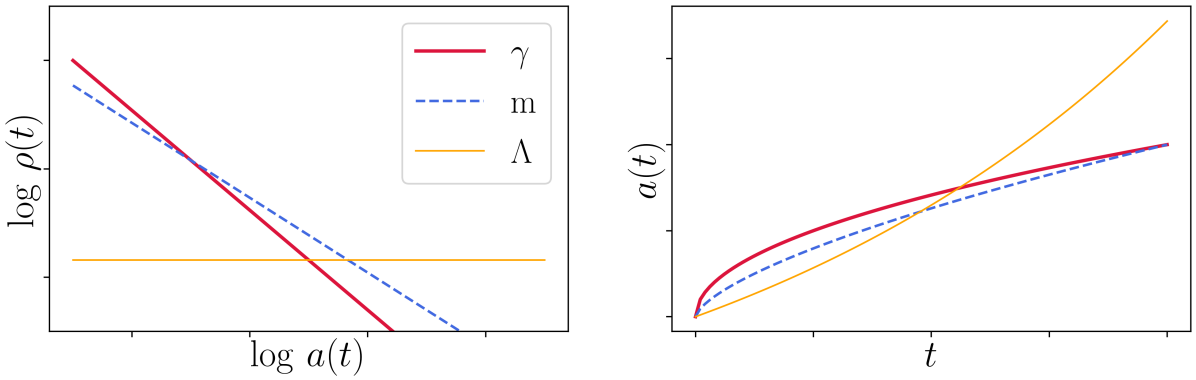


Figure A.1: (*Left*) The evolution of energy density $\rho(t)$ of radiation (*red*), dust/matter (*dashed blue*) and a cosmological constant (*orange thin*) as a function of the scale factor $a(t)$ (in log-scale). With the passing of time, radiation decays and there is a period in the evolution of the universe that matter dominates; a dark energy component described by a cosmological constant has no time evolution, and starts dominating the standard forms of matter at more recent times. (*Right*) Evolution of the scale factor as a function of cosmic time t. For matter and radiation, the expansion is decelerated. A universe dominated by dark energy, however, expands faster as time goes by. Arbitrary units are employed in these plots.

A.3.3 ΛCDM class of models

Consider the first Friedmann equation, (A.19). It is straightforward to see that there is a particular value of energy density $\rho_{c,0}$ that corresponds to a flat spacetime ($\kappa = 0$) today:

$$\rho_{c,0} = \frac{3H_0^2}{8\pi G} \approx 2.77 \times 10^{11}\ [h^2 M_\odot/\mathrm{Mpc}^3], \tag{A.36}$$

for $H_0 \approx 67$ [km/s/Mpc]. This quantity allows us to define the dimensionless density parameters

$$\Omega_X \equiv \frac{\rho_X}{\rho_{\mathrm{c},0}}, \tag{A.37}$$

where X refers to different species (matter, radiation, etc).

As discussed in Chapter 1, the ΛCDM model carries a late-time evolution dominated by a cosmological expansion. To properly describe this model, we rewrite Equation (A.19) explicitly in terms of all the allowed energy density components in the model:

$$H^2 = \frac{8\pi G}{3}(\rho_{\mathrm{m}} + \rho_\gamma + \rho_\Lambda) - \frac{\kappa}{a^2}. \tag{A.38}$$

In terms of the temporal evolution of each single component, we can cast the above expression to

$$H^2(a) = H_0^2\left[\Omega_m^0\left(\frac{a_0}{a}\right)^3 + \Omega_\gamma^0\left(\frac{a_0}{a}\right)^4 + \Omega_k^0\left(\frac{a_0}{a}\right)^2 + \Omega_\Lambda\right], \tag{A.39}$$

where $\Omega_k^0 \equiv -\kappa/(a_0 H_0)^2$ is the "density in the form of curvature" today. Its a^2 dependence can be shown by solving the Friedmann equations in the presence of curvature only.

For each Ω_X^0 value we have a ΛCDM model. Current constraints set these values to $\Omega_m^0 \approx 0.31, \Omega_\gamma^0 \sim O(10^{-5}), |\Omega_k| \leq 0.01$, and $\Omega_\Lambda \approx 0.69$.

A.3.4 Redshift

Expression (1.1) defines the redshift suffered by a photon's emitted frequency ν_e, which is observed as ν_o. From a relativistic perspective, in which photons follow null geodesics $k^\mu k^\nu{}_{;\mu} = 0$, for the background FLRW metric the photon's trajectory is given by

$$\frac{d}{d\lambda}\bar{n}^\mu + \Gamma^\mu{}_{\alpha\beta}\bar{n}^\alpha\,\bar{n}^\beta = 0, \tag{A.40}$$

$$x^\mu(\lambda) = n^\mu(\lambda - \lambda_e) + x_e^\mu, \tag{A.41}$$

where n^μ is the photon's 4-vector subjected to the constraints of unitarity, i.e. $n^2 = \delta_{ij}n^i n^j = n^0 = 1$, and the index e refers to the photon's emission coordinates. To get (A.41), the geodesic equation for the photon is solved in Minkowski space; thus, the affine parameter in the FLRW background, $\bar{\lambda}$, is related to the Minkowski affine parameter λ through the relation $\bar{\lambda} = a^2\lambda$.

The photon 4-momentum can be written as $k^\mu = \bar{\omega}\,n^\mu = \bar{\omega}\,(n^0, -n^i)$, with the frequency $\bar{\omega} = 2\pi\bar{\nu}$ being measured by an observer at its background (locally Minkowski) rest frame. Since the observed redshift of the source can be defined as the ratio between the emitted and

received frequencies, we have:

$$1 + z = \frac{\nu_e}{\nu_o} = \frac{(\tilde{n}^\mu \, \tilde{u}_\mu)_e}{(\tilde{n}^\mu \, \tilde{u}_\mu)_o} = \left(\frac{a_o}{a_e}\right) \frac{(n \cdot u_o)_o}{(n \cdot u_s)_e} = \frac{a_o}{a_e}, \tag{A.42}$$

where u_o and u_s are the observed and source 4-velocities. In the above expression it was taken into account that $\tilde{n} \sim a^{-2} n$, $\tilde{u} \sim a^{-1} u$ and the scalar product for the metric $\tilde{g}_{\mu\nu}$ is the same for the conformal $g_{\mu\nu}$ one, apart from a^2. Therefore, the redshift and scale factors are related as

$$1 + z = \frac{1}{a(t_e)}, \tag{A.43}$$

when we take $a(t_o) = a(t_0) = 1$.

A.3.5 Distances

Given the spatial part of the FLRW line element (Equation A.12),

$$\mathrm{d}s^2 = a^2(t) \left[\mathrm{d}r^2 + S_k^2(r) \left(\mathrm{d}\theta^2 + \sin^2(\theta) \, \mathrm{d}\phi^2 \right) \right], \tag{A.44}$$

we are naturally led to two distinct types of distances: one associated with the radial part $\mathrm{d}r^2$ (called comoving distance), and the other associated with the angular part of the metric $\mathrm{d}\Omega^2 \equiv \mathrm{d}\theta^2 + \sin^2(\theta) \, \mathrm{d}\phi^2$.

Comoving distance By factoring out the cosmic expansion from the physical (observed) positions in three-dimensions, $\boldsymbol{x}(t)$, we arrive at the comoving coordinates $\boldsymbol{r}(t) = a^{-1}(t) \, \boldsymbol{x}(t)$. They can be thought of as the Lagrangian coordinates, but if there are inhomogeneities present, which is actual the case, then $\boldsymbol{r}$ will inherit all the coordinates evolution, but at a "fixed grid" size: if the universe expands by an amount Δa, then "we zoom out" the same amount, keeping a constant spatial position between objects.

Distance-redshift relation We do not have access to the comoving coordinates. A direct observable in cosmology is the redshift of objects. Therefore, we can define the comoving distance r up to some redshift z through the so-called distance-redshift relation:

$$r(z) = c \int_0^z \frac{\mathrm{d}z}{H(z)} \tag{A.45}$$

where $H(z)$ is the Hubble parameter written in terms of redshift:

$$H^2(z) = H_0^2 \left[\Omega_{\mathrm{m}}^0 \, (1 + z)^3 + \Omega_\gamma^0 \, (1 + z)^4 + \Omega_k^0 \, (1 + z)^2 + \Omega_\Lambda^0 (1 + z)^{3(1+w)} \right]. \tag{A.46}$$

In this equation we extended the ΛCDM assumption of $w = -1$, and parametrized the dark energy evolution in terms of its equation of state. Notice that whenever a comoving distance is given, some cosmological model is implied.

Angular diameter distance Consider an observer looking at a source of physical size D (Figure A.2). The angular diameter distance d_A is defined as

$$d_A = \frac{D}{\delta\theta}. \tag{A.47}$$

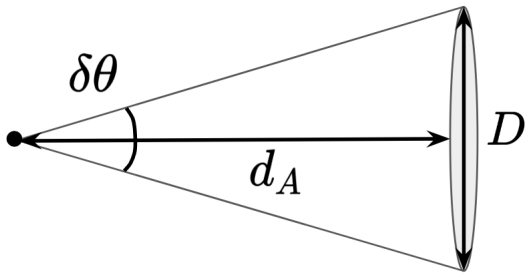


Figure A.2: Sketch representing the physical situation of an observer located at the vertex of the triangle (black dot) looking an object (shaded region) of physical size D. The observed angle allows us to determine the angular diameter distance d_A between the observer and the source.

From the metric of Equation (A.12),

$$\mathrm{d}s^2 = a^2(t) S_k^2(r) \left(\mathrm{d}\theta^2 + \sin^2(\theta)\, \mathrm{d}\phi^2\right), \tag{A.48}$$

d_A can be written in terms of the comoving distance r:

$$a^2(t)\, S_k^2(r)\, \delta\theta^2 = D^2 = d_A^2\, \delta\theta^2 \Rightarrow d_A = a(t)\, S_k(r), \tag{A.49}$$

or, in terms of redshift and for a flat FLRW universe,

$$d_A = \frac{r}{1+z}. \tag{A.50}$$

Appendix B

Overview on Perturbation Theory

B.1 Motivation and limits of validity

Let us consider a homogeneous and isotropic universe composed of photons (γ), cold dark matter (m), and many other species we would like. There will be no structures in this model universe as the total energy density field $\bar{\rho}(t) = \bar{\rho}_\gamma(t) + \bar{\rho}_\mathrm{m}(t) + \ldots$ is constant in space, and evolves in time according to the scale factor, getting diluted as time goes by. This is illustrated in the left panel of Figure B.1. Hence, we need to go beyond this picture if we wish to understand the universe in more details. The problem with general relativity is that solving the Einstein's equations for a generic model universe is highly non-trivial due to its non-linear structure, and actually there are few known analytical solutions for them.

Perturbation theory provides a fruitful method to obtain analytical solutions that allow us to understand the process of structure formation in the universe. The overall idea consists in separating a problem into a part with known solutions, e.g. the FLRW background, and a second piece that can be treated perturbatively. As long as the solution can be expanded in a convergent series, perturbation theory is a good approximation. In the scope of cosmology, we will slightly perturb the metric and energy-momentum tensors around a background whose solutions are known, as depicted in the right panel of Figure B.1. However, gravitational evolution will generate collapsed regions where perturbation theory breaks down, and one must rely on other methods to analyse the process of structure formation. In particular, the halo model is discussed in Appendix C. Despite simplistic, looking for solutions around a background universe provides many insights into the process of structure formation at large scales and early times.

In Newtonian theory, the fundamental equations describing the dynamics of a system composed of some perfect fluid are the continuity, Euler, and Poisson equations:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \boldsymbol{v}) = 0, \tag{B.1}$$

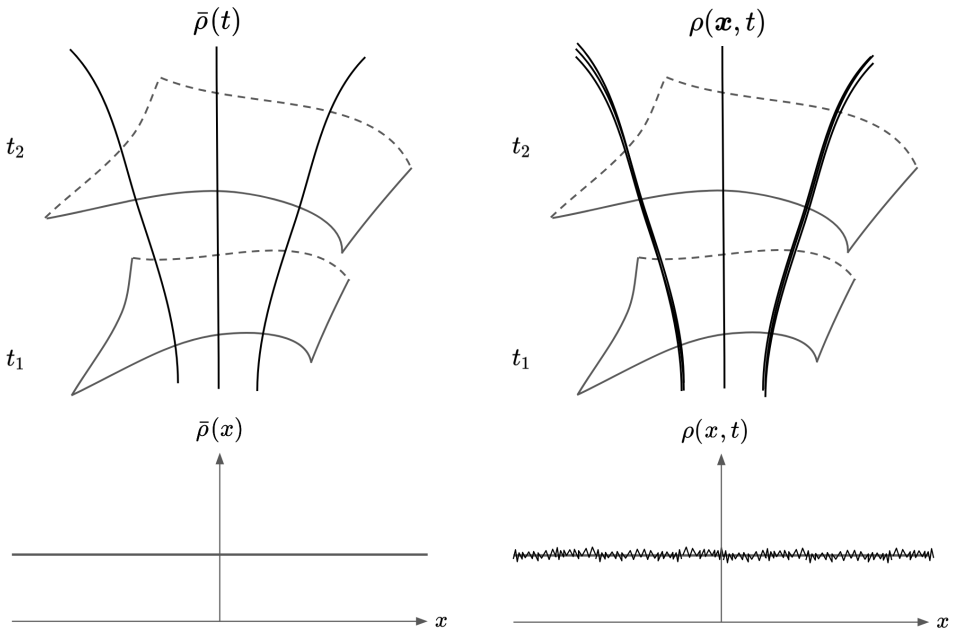


Figure B.1: (*Left*) Illustrative sketch of a background FLRW universe, where the three black curves represent the worldlines of fluid elements. In this picture, the energy density $\bar{\rho}(t)$ is spatially constant and simply gets diluted according to the time evolution obtained from the Einstein's equations (e.g. $\rho(t) \sim t^{-2}$ if we were talking about a fluid composed solely of dust or radiation). (*Right*) Schematic illustration of the perturbation theory programme: we add tiny fluctuations around the homogeneous FLRW $\bar{\rho}$, allowing for it to evolve not only on time, but spatially as well. As time passes, these small fluctuations grow, according to general relativity, to form structures such as galaxies and dark matter halos.

$$\frac{\partial \boldsymbol{v}}{\partial t} + (\boldsymbol{v} \cdot \boldsymbol{\nabla})\boldsymbol{v} + \frac{\boldsymbol{\nabla} P}{\rho} + \boldsymbol{\nabla}\phi = 0, \tag{B.2}$$

$$\nabla^2 \Phi = 4\pi G \rho, \tag{B.3}$$

where $\boldsymbol{v} = \boldsymbol{v}(\boldsymbol{x}, t)$ is the velocity field of the fluid in the Eulerian frame, $\rho = \rho(\boldsymbol{x}, t)$ its mass density, $P = P(\boldsymbol{x}, t)$ is the pressure and we neglect dissipative effects.

The theoretical scheme of the perturbation theory programme consists in perturbing the fluid variables, e.g. $\rho(\boldsymbol{x}, t)$, around a background solution $\bar{\rho}(t)$ which we will take to be the FLRW one[1]. Then we substitute the perturbed quantities in the set of dynamical equations, and solve for each order in the perturbations. For example, the Poisson equation (B.3) can be separated into the background $\nabla^2 \bar{\Phi} = 4\pi G\, \bar{\rho}$, and perturbation $\nabla^2 \delta\Phi = 4\pi G\, \delta\rho$ pieces.

This approach has been successful in describing scales smaller than the Hubble radius,

[1]Notice that there is no need to assume FLRW as the background around which we consider the perturbations. An example is the inhomogeneous Lemaître-Tolman-Bondi metric that is an exact solution of Einstein's equations, where we can perform the perturbation theory scheme (see, for example, Clarkson et al., 2009).

with weak gravitational fields and when the perturbations are not sourced by relativistic species such as photons, or neutrinos. Whenever we departure from this framework, we must account for the intrinsic limitations of the Newtonian theory by employing the complete geometric study allowed by general relativity. When there is an analogue between both frameworks, the Newtonian equations of motion acquire corrections; however, there are cases where no analogue can be drawn between both theories: in the case of gravitational waves (coming from the $ij-$part of the metric) and the frame-dragging effect (coming from the $0i$-part), as discussed in what follows, where we proceed by applying the perturbative programme to the proper equations of motion, namely the Einstein's equations.

B.2 Perturbation theory in general relativity

When applying perturbation theory to the Einstein's equations (A.2),

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = 8\pi G T_{\mu\nu}, \tag{B.4}$$

we see that we need to consider perturbations in the energy-momentum tensor,

$$T_{\mu\nu}(t, \boldsymbol{x}) = \bar{T}_{\mu\nu}(t) + \delta T_{\mu\nu}(t, \boldsymbol{x}), \tag{B.5}$$

and in the metric,

$$g_{\mu\nu}(t, \boldsymbol{x}) = \bar{g}_{\mu\nu}(t) + \delta g_{\mu\nu}(t, \boldsymbol{x}), \tag{B.6}$$

around the FLRW solutions $\bar{T}^{\mu}_{\nu} = (\bar{\rho} + \bar{P})\, \bar{u}^{\mu}\, \bar{u}_{\nu} - \bar{P}\, \delta^{\mu}{}_{\nu}$, introduced in Equation (A.18), and

$$\bar{g}_{\mu\nu}\, \mathrm{d}x^{\mu} \mathrm{d}x^{\nu} = a^2(\tau)[-\mathrm{d}\tau + \gamma_{ij}\, \mathrm{d}x^{i} \mathrm{d}x^{j}]. \tag{B.7}$$

We call the reader's attention to the natural units with which $c = 1$, stressing that ρ will refer to the fluid's energy density, and $\boldsymbol{x} \equiv x^{i}$.

We can write the full perturbed metric as

$$\mathrm{d}s^2 = \bar{g}_{\mu\nu}\, \mathrm{d}x^{\mu}\, \mathrm{d}x^{\nu} + \delta g_{\mu\nu}\, \mathrm{d}x^{\mu}\, \mathrm{d}x^{\nu}, \tag{B.8}$$

and the energy-momentum perturbation as

$$\delta T^{\mu}_{\nu} = (\delta\rho + \delta P)\, \bar{u}^{\mu}\, \bar{u}_{\nu} + (\bar{\rho} + \bar{P})(\delta u^{\mu}\, \bar{u}_{\nu} + \bar{u}^{\mu}\, \delta u_{\nu}) - \delta P\, \delta^{\mu}{}_{\nu}. \tag{B.9}$$

Above, we are neglecting sources of anisotropic stress $\Pi^{i}{}_{j}$, such that $T^{i}{}_{j} = 0$ for $i \neq j$. For computations including a non-vanishing $\Pi^{i}{}_{j}$, see Chapter 2 of Durrer (2020) and Appendix D

of Giovannini (2008).

B.2.1 Scalar-Vector-Tensor (SVT) decomposition

For the FLRW background, a given hypersurface of constant (conformal) time τ will carry translation and isotropy symmetries. This allows the perturbations to be decomposed into scalar, vector and tensor modes. At linear order and in Fourier space (flat case), these 3 objects will evolve independently, i.e. they are decoupled (for example see Durrer, 2020, Section 2.2.2).

B.2.1.1 Perturbations in the metric sector

We start by considering the perturbed part of the metric:

$$\delta g_{\mu\nu}\,\mathrm{d}x^{\mu}\,\mathrm{d}x^{\nu} = \delta g_{00}\,\mathrm{d}\tau^2 + 2\,\delta g_{0i}\,\mathrm{d}\tau\,\mathrm{d}x^{i} + \delta g_{ij}\,\mathrm{d}x^{i}\,\mathrm{d}x^{j}. \tag{B.10}$$

We can see that the temporal part is exclusively scalar; hence, we parametrize it in terms of the scalar perturbation $\mathcal{A}$:

$$\delta g_{00}^{(s)} = -2\,a^2(\tau)\,\mathcal{A} \equiv -2\,a^2(\tau)\,\phi(\tau, x^i), \tag{B.11}$$

The off-diagonal piece is a vectorial quantity and thus we can introduce the vector perturbation $\mathcal{B}_i$,

$$\delta g_{0i} = -a^2(\tau)\,\mathcal{B}_i, \tag{B.12}$$

which, in turn, can be decomposed into scalar (s) and pure-vector (v) components:

$$\delta g_{0i} \equiv \delta g_{0i}^{(s)} + \delta g_{0i}^{(v)} = -a^2(\tau)\,[\partial_i B(\tau, x^j) + B_i(\tau, x^j)]. \tag{B.13}$$

In this section, ∂_i denotes the covariant derivative for a general curvature κ, when γ_{ij} is the spatial metric component; for a flat spacetime ($\gamma_{ij} = \delta_{ij}$) it denotes the usual derivatives.

Finally, we introduce the tensorial perturbation C_{ij} in $ij-$part of the metric:

$$\delta g_{ij} = 2\,a^2(\tau)\,C_{ij}. \tag{B.14}$$

Again, it can be separated into a pure scalar, a pure vector, and a pure tensor (t):

$$\delta g_{ij} \equiv \delta g_{ij}^{(s)} + \delta g_{ij}^{(v)} + \delta g_{ij}^{(t)}, \tag{B.15}$$

where

$$\delta g_{ij}^{(s)} = 2\,a^2(\tau)\,[\psi(\tau, x^k)\,\gamma_{ij} + \partial_i\partial_j E(\tau, x^k)],$$
$$\delta g_{ij}^{(v)} = a^2(\tau)\,[\partial_i F_j(\tau, x^k) + \partial_j F_i(\tau, x^k)],$$
$$\delta g_{ij}^{(t)} = 2\,a^2(\tau)\,h_{ij}.$$

Therefore, the full perturbed metric in terms of the SVT decomposition is given by

$$\mathrm{d}s^2 = a^2(\tau)\left[-(1+2\phi)\,\mathrm{d}\tau^2 - 2\,\mathcal{B}_i\,\mathrm{d}\tau dx^i + (\gamma_{ij} + 2C_{ij})\,\mathrm{d}x^i\,\mathrm{d}x^j\right]. \tag{B.16}$$

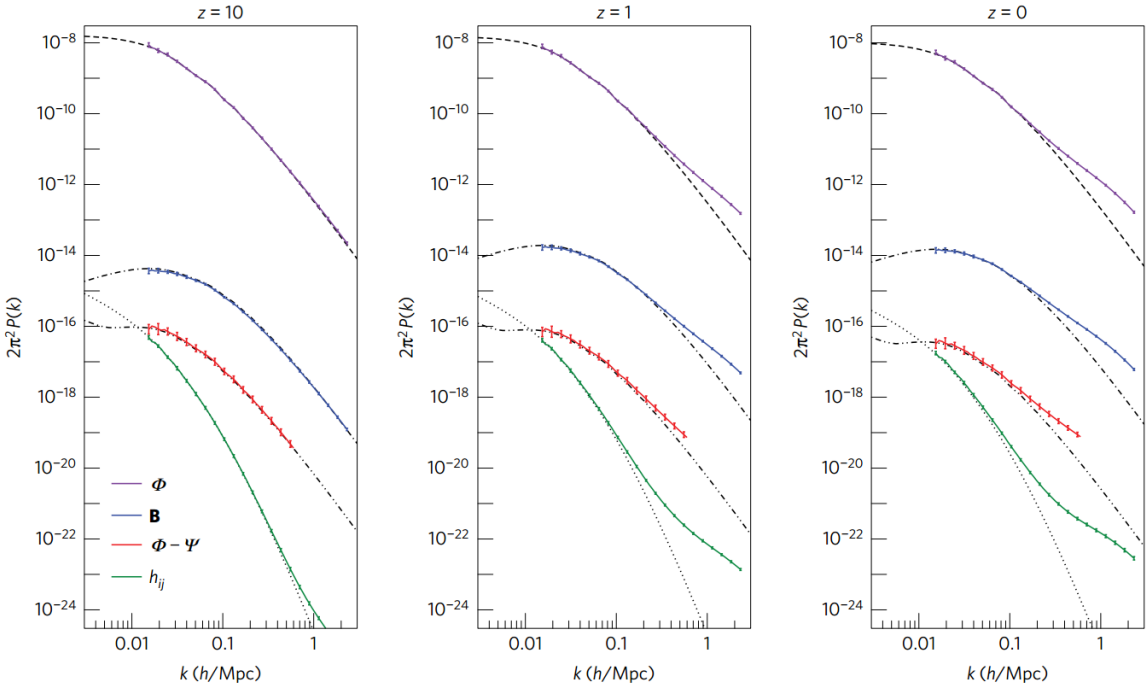


Figure B.2: Numerical power spectra of different perturbation theory quantities, at three different redshifts. The gravitational potential Φ (top curves, *violet*), the frame-dragging potential $\boldsymbol{B} \equiv B_i$ (second curves, *blue*), the difference between the scalar potentials $\Phi - \Psi$ (third curves from top, *red*) and the tensor perturbation h_{ij} (bottom curves, *green*) were obtained with eight simulations from the relativistic N-body code *gevolution*. Error bars indicate the random fluctuations from these different realisations. The dashed, dot-dashed, dot-dot-dashed, and dotted black curves shows the second-order perturbation theory results. As expected, perturbation theory breaks down at late times, where non-linearities are larger due to gravitational evolution, and at small scales. [Extracted from Adamek et al. (2016a)].

The vector modes, associated with rotational fluid flow, decay very quickly and will not be will not be relevant for our purposes. The tensor perturbations will give rise to gravitational waves. Finally, the scalar perturbations will exhibit instabilities responsible for the growth of inhomogeneities and are responsible for structure formation. A more accurate description, involving the perturbation theory programme up to second order, couples these different perturb-

ations, though. To see the magnitude these terms have on structure formation, we reproduce Figure 2 of Adamek et al. (2016a) in Figure B.2, which illustrates the S, V and T components after choosing the Poisson gauge that fixes $\mathcal{B}_i$ to B_i in a gauge-invariant fashion (see Section B.3).

B.2.1.2 Perturbations in the matter sector

Consider now the perturbations in the hydrodynamical matter (radiation or dark matter). As we saw, the energy-momentum tensor perturbations, Equation (B.9), depends on velocity perturbations: $u_\mu = \bar{u}_\mu + \delta u_\mu$. The normalisation condition for the full metric and the background, $g_{\mu\nu}u^\mu u^\nu = \bar{g}_{\mu\nu}\bar{u}^\mu\bar{u}^\nu = 1$, implies that, at first order in perturbations,

$$2\bar{u}_\mu \delta u^\mu + \delta g_{\mu\nu}\bar{u}^\mu\bar{u}^\nu = 0. \tag{B.17}$$

From the normalisation condition for the background, a comoving observer has 4-velocity $\bar{u}^\mu = a^{-1}(\tau)\delta^\mu{}_0$. From the temporal perturbations of the metric, $\delta g_{00} = -2a^2(\tau)\phi$, we conclude that

$$\delta u^0 = -\frac{\phi}{a(\tau)}. \tag{B.18}$$

Finally, defining $\delta u^i = \mathcal{V}^i/a(\tau)$, we obtain

$$u^\mu = \frac{1}{a(\tau)}\left(1-\phi, \mathcal{V}^i\right), \tag{B.19}$$

where the velocity perturbation $\mathcal{V}^i = v^{,i} + v^i$ is also decomposed into a pure scalar v and pure vector v^i. By lowering the index with the full metric, we obtain $u_\mu = a(\tau)\left(-(1+\phi), \mathcal{V}_i + \mathcal{B}_i\right)$

With these definitions, the energy-momentum tensor perturbations are:

$$\begin{aligned}
\delta T^0{}_0 &= -\delta\rho, \\
\delta T^i{}_0 &= (\bar{\rho} + \bar{P})\,\mathcal{V}^i, \\
\delta T^0{}_j &= -(\bar{\rho} + \bar{P})(\mathcal{V}_j + \mathcal{B}_j), \\
\delta T^i{}_j &= \delta P \delta^i{}_j.
\end{aligned} \tag{B.20}$$

B.3 Gauge transformations and its consequences

As discussed in Section A.1, general relativity is endowed with diffeomorphism invariance, which allows us to choose whatever coordinate system we would like, $x^\mu \to \tilde{x}^\mu(x^\mu)$, and still have equivalent physical laws (the principle of general relativity). In perturbation theory, when we separate the problem into two spacetimes, a fictitious and homogeneous background

$\overline{\mathcal{M}}$ and the physical perturbed $\mathcal{M}$, we must define a correspondence, or a map $\varphi : \overline{\mathcal{M}} \to \mathcal{M}$, between both spacetimes.

If we keep the background coordinate system fixed, but change the mapping between the two spacetimes, $\varphi \to \tilde{\varphi}$, we are performing a gauge transformation. This gauge freedom, i.e. the freedom to choose the correspondence between the physical and fictitious spacetimes, poses an issue in perturbation theory: one might generate non-physical perturbations and remove physical ones, that is, "the gauge transformation changes the event in $\mathcal{M}$ which is associated with a given event in the background $\overline{\mathcal{M}}$" (Ellis et al., 2012, p. 250).

For example, suppose we take the background density $\bar{\rho} = \bar{\rho}(t)$ at a coordinate system $x^\mu = (t, x^i)$. General covariance allows us to change to a new coordinate system $\tilde{x}^\mu = (\tilde{t}, x^i)$ (we keep the spatial coordinates fixed at the old system). We can relate $\tilde{t}$ and t as $\tilde{t} = t + \tilde{\delta}t(t, x)$. In this new chart, which defines the reference frame $\tilde{S}$

$$\tilde{\rho}(\tilde{t}, x^i) \equiv \bar{\rho}(t(\tilde{t}, x^i)), \tag{B.21}$$

which leads to

$$\tilde{\rho}(\tilde{t}, x^i) = \bar{\rho}(\tilde{t} - \tilde{\delta}t(t, x^i)) \approx \bar{\rho}(t) - \frac{\partial \bar{\rho}}{\partial t}\tilde{\delta}t \equiv \bar{\rho}(\tilde{t}) + \delta\rho(\tilde{t}, x^i) \tag{B.22}$$

if we assume that $\tilde{\delta}t \ll t$ (Mukhanov, 2005). Here, $\bar{\rho}(\tilde{t})$ may be interpreted as a new background density in the coordinate system $\tilde{S}$, whereas $\delta\rho(\tilde{t}, x^i)$ is a non-physical perturbation that appears through the coordinate change $x^\mu \to \tilde{x}^\mu$.

Any correspondence (gauge) $\varphi : \overline{\mathcal{M}} \to \mathcal{M}$, for a fixed coordinate system $\bar{x}^\mu$ at the background manifold $\overline{\mathcal{M}}$ with metric $\bar{g}_{\mu\nu}(\bar{x})$), maps some function $\bar{f}(\bar{x}^\mu) \in \overline{\mathcal{M}}$ into $f(\bar{x}^\mu) \in \mathcal{M}$. The perturbation of f is

$$\delta f(p) = f(p) - \bar{f}(\varphi^{-1}(p)), \tag{B.23}$$

where $p \in \mathcal{M}$ is some event at the physical manifold. However, by choosing $\tilde{\varphi} : \overline{\mathcal{M}} \to \mathcal{M}$, which induces another set of coordinates $\tilde{x}^\mu$,

$$\tilde{\delta f}(p) = \tilde{f}(p) - \bar{f}(\tilde{\varphi}^{-1}(p)), \tag{B.24}$$

we perform a gauge transformation

$$\delta f(p) \to \tilde{\delta f}(p). \tag{B.25}$$

Notice that the physical event p is kept fixed.

The quantity f is called gauge invariant if its functional dependence is the same for both coordinate systems: $f(x^\mu) = f(\tilde{x}^\mu)$. However, we cannot guarantee that $f(x^\mu) = \tilde{f}(\tilde{x}^\mu)$, meaning that δf can differ from one system to another. Because the metric $g_{\mu\nu}(x)$ is defined on a manifold

for a fixed coordinate system, we can alter the shape of the metric perturbations through a gauge transformation.

There are two routes to distinguish between physical and fictitious perturbations: **(1)** one can define gauge-invariant variables and solve Einstein's equations, whose solutions will be given in terms of gauge independent quantities, by construction; or **(2)** fix some gauge and solve the equations for this particular choice. Because route **(1)** is cumbersome, one usually works fixing the gauge. However, care must be taken: as we shall see, the power spectrum of overdensities δ depends on the gauge choice. It is possible, then, to construct the power spectrum of a gauge-independent quantity, such as the number counts worked out in Chapter 5.

B.3.1 Gauge-invariant variables: metric sector

Bardeen (1980) settled down a complete set of gauge-invariant variables that clarified the physical interpretation of the cosmological perturbations in the relativistic framework. This can be achieved by considering the most general coordinate transformation

$$\tilde{x}^\mu = x^\mu + \xi^\mu, \tag{B.26}$$

where ξ^μ is an infinitesimal shift in the coordinates. Under this change of coordinates, the metric $g_{\mu\nu}$ transforms as

$$\begin{aligned}
\tilde{g}_{\alpha\beta}(\tilde{x}) &= \frac{\partial x^\mu}{\partial \tilde{x}^\alpha}\frac{\partial x^\nu}{\partial \tilde{x}^\beta} g_{\mu\nu}(x), \\
\bar{g}_{\alpha\beta}(\tilde{x}) + \tilde{\delta g}_{\alpha\beta}(\tilde{x}) &= \left(\delta^\mu_\alpha - \frac{\partial \xi^\mu}{\partial \tilde{x}^\alpha}\right)\left(\delta^\nu_\beta - \frac{\partial \xi^\nu}{\partial \tilde{x}^\beta}\right)(\bar{g}_{\mu\nu}(x) + \delta g_{\mu\nu}(x)), \\
&\approx (\delta^\mu_\alpha \delta^\nu_\beta - \delta^\nu_\beta \xi^\mu_{,\alpha} - \delta^\mu_\alpha \xi^\nu_{,\beta})(\bar{g}_{\mu\nu}(x) + \delta g_{\mu\nu}(x)), \\
&\approx \bar{g}_{\alpha\beta}(x) + \delta g_{\alpha\beta}(x) - \bar{g}_{\mu\beta}\xi^\mu_{,\alpha} - \bar{g}_{\alpha\nu}\xi^\nu_{,\beta}.
\end{aligned}$$

Since

$$\bar{g}_{\alpha\beta}(x) = \bar{g}_{\alpha\beta}(\tilde{x}) - \bar{g}_{\alpha\beta,\rho}\xi^\rho, \tag{B.27}$$

perturbations in the metric will change, under the gauge transformation of Equation (B.26), as

$$\tilde{\delta g}_{\alpha\beta} = \delta g_{\alpha\beta} - \bar{g}_{\alpha\beta,\rho}\xi^\rho - \bar{g}_{\mu\beta}\xi^\mu_{,\alpha} - \bar{g}_{\alpha\nu}\xi^\nu_{,\beta}. \tag{B.28}$$

The dependence on ξ^μ lead us to the already reached conclusion that perturbations will depend on the background definition. Once again, we see that it is possible split ξ^μ into a scalar T and a vector part, $\xi^\mu = (T, L^i)$, with $L^i = \partial^i L + \hat{L}^i$, L a scalar and $\partial_i \hat{L}^i = 0$. In a FLRW

background, the transformations in Equation (B.28) for the full metric lead us to

$$\tilde{\delta g}_{00} = \delta g_{00} + 2a(aT)', \tag{B.29}$$

$$\tilde{\delta g}_{0i} = \delta g_{0i} - a^2(L_i' - T_{,i}), \tag{B.30}$$

$$\tilde{\delta g}_{ij} = \delta g_{ij} - a^2(2\mathcal{H}\gamma_{ij}T + L_{j,i} + L_{i,j}). \tag{B.31}$$

Scalar perturbations From Equation (B.29), we see that

$$\tilde{\phi} = \phi - \mathcal{H}T - T'. \tag{B.32}$$

For the scalar part of L_i, Equations (B.30) and (B.31) give us, respectively,

$$\tilde{B} = B + L' - T, \tag{B.33}$$

$$\tilde{\psi} = \psi + \mathcal{H}T, \tag{B.34}$$

and

$$\tilde{E} = E + L. \tag{B.35}$$

In Bardeen (1980) he defines the two gauge-invariant linear combinations of Equations (B.32), (B.33), (B.34) and (B.35):

$$\Phi = \phi - \frac{1}{a}\left[a(B - E')\right]', \tag{B.36}$$

and

$$\Psi = \psi - \mathcal{H}(B - E'), \tag{B.37}$$

called the Bardeen potentials.

Vector perturbations An analogous procedure can be done for the vector contributions, leading us to

$$\tilde{B}_i = B_i + \frac{\hat{L}_i'}{2}. \tag{B.38}$$

and

$$\tilde{F}_i = F_i + \hat{L}_i. \tag{B.39}$$

Again, these can be linearly combined into the gauge-invariant vector (Bardeen, 1980):

$$\Psi_i = B_i - F_i'. \tag{B.40}$$

Tensor perturbations From Equation (B.31), it is immediate to see that tensor perturbations are already gauge-invariant:

$$\tilde{h}_{ij} = h_{ij}. \tag{B.41}$$

B.3.2 Gauge-invariant variables: energy-momentum tensor

Gauge freedom also propagates to the energy-momentum tensor. Considering how T^{μ}_{ν} transforms under Equation (B.26),

$$\tilde{T}^{\alpha}_{\ \beta}(\tilde{x}) = \frac{\partial \tilde{x}^{\alpha}}{\partial x^{\mu}} \frac{\partial x^{\nu}}{\partial \tilde{x}^{\beta}} \tilde{T}^{\mu}_{\ \nu}(x), \tag{B.42}$$

we obtain

$$\bar{T}^{\alpha}_{\ \beta}(\tilde{x}) + \delta\tilde{T}^{\alpha}_{\ \beta}(\tilde{x}) \approx \bar{T}^{\alpha}_{\ \beta}(x) + \delta T^{\alpha}_{\ \beta}(x) - \bar{T}^{\alpha}_{\ \nu}\xi^{\nu}_{\ ,\beta} - \bar{T}^{\mu}_{\ \beta}\xi^{\alpha}_{\ ,\mu}, \tag{B.43}$$

which, at linear order in perturbations, results in the following expression for the gauge transformation (B.26):

$$\delta\tilde{T}^{\alpha}_{\ \beta} = \delta T^{\alpha}_{\ \beta} - \bar{T}^{\alpha}_{\ \beta,\rho}\xi^{\rho} - \bar{T}^{\alpha}_{\ \nu}\xi^{\nu}_{\ ,\beta} - \bar{T}^{\mu}_{\ \beta}\xi^{\alpha}_{\ ,\mu}. \tag{B.44}$$

In complete analogy to what has been done for the metric, we see how the energy-momentum tensor perturbations changes:

$$\delta\tilde{T}^{0}_{\ 0} = \delta T^{0}_{\ 0} - \bar{\rho}'T - 2\bar{\rho}T', \tag{B.45}$$

$$\delta\tilde{T}^{i}_{\ 0} = \delta T^{i}_{\ 0} - (\bar{\rho} - \bar{P})(\partial^{i}L + L^{i})', \tag{B.46}$$

$$\delta\tilde{T}^{0}_{\ i} = \delta T^{0}_{\ i} - (\bar{\rho} - \bar{P})T_{,i}, \tag{B.47}$$

$$\delta\tilde{T}^{i}_{\ j} = \delta T^{i}_{\ j} + \bar{P}'T\delta^{i}_{\ j} + 2\bar{P}(\partial^{i}L + L^{i})_{,j}. \tag{B.48}$$

Scalar perturbations Considering only the scalar part of the energy-momentum tensor, we have that

$$\delta\tilde{\rho} = \delta\rho + \bar{\rho}'T, \tag{B.49}$$

$$\delta\tilde{P} = \delta P + \bar{P}'T, \tag{B.50}$$

$$\tilde{v} = v - L'. \tag{B.51}$$

Therefore, the gauge-invariant combinations are

$$\Delta = \delta\rho + \bar{\rho}'(B - E') \tag{B.52}$$

and

$$V = v + E'. \tag{B.53}$$

Vector perturbations For the vector part we obtain

$$\tilde{v}_i = v_i - L_i', \tag{B.54}$$

with

$$V_i = v_i + B_i \tag{B.55}$$

being a gauge-invariant quantity.

Tensor perturbations Since we are not considering anisotropic stress, there are no tensor perturbations.

Summary For short, the gauge-invariant variables for the energy-momentum tensor can be written as (Mukhanov et al., 1992; Mukhanov, 2005)

$$\delta T^{0\,(\mathrm{gi})}_{\ 0} = \delta T^{0}_{\ 0} - (\bar{T}^{0}_{\ 0})'(B - E'), \tag{B.56}$$

$$\delta T^{0(\mathrm{gi})}_{\ i} = \delta T^{0}_{\ i} - \left(\bar{T}^{0}_{\ 0} - \frac{\bar{T}^{i}_{\ i}}{3}\right)(B - E')_{,i}, \tag{B.57}$$

$$\delta T^{i\ (\mathrm{gi})}_{\ j} = \delta T^{i}_{\ j} - (\bar{T}^{i}_{\ j})'(B - E'), \tag{B.58}$$

where $\bar{T}^{i}_{\ i}$ is the trace of the spatial components of $\bar{T}^{\mu}_{\ \nu}$.

B.3.3 Gauge choices

The 10 metric and energy-momentum perturbations set at Section B.2.1 can be arbitrarily chosen. In what follows we discuss a few of these choices.

Conformal-Newtonian gauge Setting the scalars $B = E = 0$, the metric perturbations defined in are equivalent to the Bardeen potentials of Equations (B.37) and (B.37):

$$\Phi = \phi, \quad \Psi = \psi. \tag{B.59}$$

Neglecting the vector and tensor perturbations, the metric in this "coordinate" system becomes

$$\mathrm{d}s^2 = a^2(\tau)[-(1 + 2\Phi)\,\mathrm{d}\tau^2 + (1 - 2\Psi)\,\gamma_{ij}\mathrm{d}x^i\,\mathrm{d}x^j]. \tag{B.60}$$

Synchronous gauge Choosing $\phi = \mathcal{B}_i = 0$, the metric takes the form

$$\mathrm{d}s^2 = a^2(\tau)\left[-\mathrm{d}\tau^2 + (\gamma_{ij} + 2C_{ij})\,\mathrm{d}x^i\,\mathrm{d}x^j\right]. \tag{B.61}$$

This is called the synchronous gauge. However, interpreting the results in this gauge is not straightforward due to the presence of spurious gauge modes. Nonetheless, by solving Einstein's equations in the Newtonian gauge, it is possible to relate the scalar perturbation modes ψ, E and the energy density in the synchronous gauge via the Bardeen potentials (Mukhanov, 2005, p.296).

Poisson gauge Generalising the Newtonian gauge to allow vector and tensor perturbations, one arrives at the Poisson gauge form of the metric:

$$\mathrm{d}s^2 = a^2(\tau)\left[-(1+2\Phi)\mathrm{d}\tau^2 - 2\,B_i\,\mathrm{d}\tau\mathrm{d}x^i + ((1-2\Psi)\gamma_{ij} + h_{ij})\,\mathrm{d}x^i\,\mathrm{d}x^j\right]. \tag{B.62}$$

The power spectrum for the potentials appearing in this metric are shown in Figure B.2.

B.4 Perturbed Einstein's equations

Now we turn to the Einstein's equations (A.2), which can be written in their perturbed form as

$$\bar{G}^{\mu}_{\ \nu} + \delta G^{\mu}_{\ \nu} = 8\pi G(\bar{T}^{\mu}_{\ \nu} + \delta T^{\mu}_{\ \nu}). \tag{B.63}$$

The perturbed connection coefficients $\Gamma^{\alpha}_{\ \beta\gamma}$, the Riemann $R^{\alpha}_{\ \beta\mu\nu}$ and Ricci $R_{\mu\nu}$ tensors and, finally, the Ricci scalar $R = R^{\mu}_{\ \nu}$ can be found in Peter and Uzan (2013). For a more general treatment, (Acquaviva et al., 2003) and (Bartolo et al., 2004) compute these quantities up to second order in perturbation theory.

B.4.1 Newtonian gauge

Here we write Einstein's equations after fixing the gauge $E = B = 0$. We also present the solutions for the scalar perturbations, highlighting the link with the Newtonian theory, that is, when the perturbation theory programmed is applied to the of Newtonian equations (B.1)-(B.3).

Scalar equations The scalar degrees of freedom obeys the following Einstein's equations:

$$4\pi G a^2 \delta T^{0\ (gi)}_{\ 0} = \nabla^2\Psi - 3\mathcal{H}(\Psi' + \mathcal{H}\Phi), \tag{B.64}$$

$$4\pi G a^2 \delta T^{0(gi)}_{\ \ i} = (\Psi' + \mathcal{H}\Phi)_{,i}, \tag{B.65}$$

$$-4\pi G a^2 \delta T^{i\ (gi)}_{\ \ j} = \left[\Psi'' + \mathcal{H}(2\Psi+\Phi)' + (2\mathcal{H}' + \mathcal{H}^2)\Phi + \frac{1}{2}\nabla^2(\Phi-\Psi)\right]\delta^i_{\ j} + \\ - \frac{1}{2}\partial^i\partial_j(\Phi-\Psi), \tag{B.66}$$

where the perturbations of the energy-momentum tensor are already the gauge-invariant quantities defined in Equations (B.56) - (B.58).

Equation (B.66), in the absence of anisotropic stress ($\delta T^{i}_{\ j} = 0$ whenever $i \neq j$) gives the standard relation $\Phi = \Psi$ between the Bardeen's potentials. This is a good approximation for the ΛCDM universe, as one can see in Figure B.2 where the contributions coming from $\chi \equiv \Phi - \Psi$ are orders of magnitude smaller than the ones from Φ. We can use this relation to write Equations (B.64) and (B.66) as:

$$4\pi G a^2 \delta\rho = \nabla^2\Phi - 3\mathcal{H}(\Phi' + \mathcal{H}\Phi), \tag{B.67}$$

$$-4\pi G a^2 \delta P = \Phi'' + 3\mathcal{H}\Phi' + (2\mathcal{H}' + \mathcal{H}^2)\Phi. \tag{B.68}$$

In Equation (B.67) we see a clear deviation from the standard Newtonian treatment: while it reduces to the Poisson equation (B.3) in a static universe ($\mathcal{H} = 0$), in an expanding space $-3\mathcal{H}(\Phi' + \mathcal{H}\Phi)$ appears as a relativistic correction, together with the fact that it is also couple to Equation (B.68).

For a universe dominated by dust ($\delta P = 0$), Equation (B.68) becomes

$$\Phi'' + 3\mathcal{H}\Phi' + (2\mathcal{H}' + \mathcal{H}^2)\Phi = 0. \tag{B.69}$$

Recalling that at this matter-dominated (MD) stage the scale factor is proportional to τ^2, hence $\mathcal{H} = 2\tau^{-1}$, it reduces to

$$\Phi'' + 6\,\tau^{-1}\,\Phi' = 0, \tag{B.70}$$

whose solution is

$$\Phi_{(\mathrm{MD})}(\mathbf{x}, \tau) = C_1(\mathbf{x}) + C_2(\mathbf{x})\tau^{-5}. \tag{B.71}$$

Neglecting the decaying mode, during the matter-dominated era the gravitational potential is constant,

$$\Phi(\mathbf{x}, \tau) = \Phi(\mathbf{x}) = C_1(\mathbf{x}). \tag{B.72}$$

From Equation (B.67) and using Friedmann equation (A.19), $4\pi G a^2 = 3\mathcal{H}^2/(2\bar{\rho})$, the matter perturbations follow

$$\frac{\delta\rho}{\bar{\rho}} = \frac{1}{6}\Big[(\nabla^2 C_1\tau^2 - 12C_1) + \tau^{-5}(\nabla^2 C_2\tau^2) + 18C_2\Big], \tag{B.73}$$

or, in terms of the density contrast in Fourier space,

$$\delta_{(\mathrm{MD})}(k, \tau) = -2C_1(\mathbf{k})\left[1 + \frac{(k\,\tau)^2}{12}\right] + \frac{3C_2(\mathbf{k})}{\tau^5}\left[1 - \frac{(k\,\tau)^2}{18}\right]. \tag{B.74}$$

Also neglecting the decaying mode, the density contrast depends on time and scale as

$$\delta_{(\mathrm{MD})}(k,\tau) = -2C_1(\mathbf{k})\left(1+\frac{k^2\tau^2}{12}\right). \tag{B.75}$$

For long-wavelength perturbations, i.e. $k\tau \ll 1 \Rightarrow k \ll \mathcal{H} \Rightarrow \lambda \gg \mathcal{H}$, called super-horizon scales (sp),

$$\delta^{\mathrm{sp}}_{(\mathrm{MD})}(k,\tau) \approx -2\Phi(k,\tau_i), \tag{B.76}$$

whereas for sub-horizon scales (sb) we obtain

$$\delta^{\mathrm{sb}}_{(\mathrm{MD})}(k,\tau) \approx -\Phi(k,\tau_i)\frac{k^2\tau^2}{6} \propto a(\tau). \tag{B.77}$$

Here $\Phi(k,\tau_i)$ is the gravitational potential when the universe enters the MD stage, which remains "frozen" until we have some phase transition (such as dark energy domination).

Equation (B.77) can be found within the Newtonian context by solving Equations (B.1)-(B.3)[2]. This highlights the Newtonian limit of the relativistic theory and expresses what is meant by saying that on sub-horizon scales the Newtonian approximation is valid within cosmology. Another interesting concept that becomes evident with these solutions is the one of "entering the horizon", i.e. the Hubble radius $1/\mathcal{H}$. While the matter perturbations are outside the horizon, they do not experience gravitation, remaining constant. After they enter the horizon, they start to feel gravity and the perturbations start to grow according to Equation (B.77). The presence of this "horizon" is not captured in the Newtonian treatment.

For completeness, we consider now what happens in an universe dominated by radiation, when the perturbations are adiabatic: considering the thermodynamic relation $P(\boldsymbol{x},t) = P(\rho(\boldsymbol{x},t), S(\boldsymbol{x},t)) = P(\bar{\rho}+\delta\rho, \bar{S}+\delta S)$, a pressure perturbation δP is adiabatic when there are now entropy perturbations, i.e. $\delta S = 0$ in

$$\delta P = \left(\frac{\partial P}{\partial \rho}\right)_S \delta\rho + \left(\frac{\partial P}{\partial S}\right)_\rho \delta S \equiv c_s^2\delta\rho + \chi_S\delta S. \tag{B.78}$$

Here, $c_s^2 \equiv \left(\frac{\partial P}{\partial \rho}\right)_S = w$ is the square of the speed of sound and $\chi_S \equiv \left(\frac{\partial P}{\partial S}\right)_\rho$ is the adiabatic compressibility coefficient.

For the radiation era, $p = \rho/3$, $c_s^2 = 1/3$, $a \propto \tau \Rightarrow \mathcal{H} = \tau^{-1}$, Equation (B.68) becomes

$$\Phi'' + \frac{4}{\tau}\Phi' + \frac{k^2}{3}\Phi = 0. \tag{B.79}$$

[2]For a complete derivation, see Mukhanov (2005), Chapter 6.

This equation can be solved by performing the change of variables $y = k\tau/\sqrt{3}$,

$$\frac{\mathrm{d}^2\Phi}{\mathrm{d}y^2} + \frac{4}{y}\frac{\mathrm{d}\Phi}{\mathrm{d}y} + \Phi = 0, \tag{B.80}$$

and after setting $f = y^{3/2}\Phi$, it takes the form of a Bessel equation,

$$\frac{\mathrm{d}^2 f}{\mathrm{d}y^2} + \frac{1}{y}\frac{\mathrm{d}f}{\mathrm{d}x} + \left[1 - \frac{(3/2)^2}{x^2}\right] f = 0, \tag{B.81}$$

whose solutions are:

$$f(y) = C_1 J_{3/2}(y) + C_2 Y_{3/2}(y). \tag{B.82}$$

Expanding around $y \ll 1$, the Bessel functions of first and second kind are, respectively,

$$J_{3/2}(y) = \sqrt{\frac{2}{\pi y}}\left(\frac{\sin y}{y} - \cos y\right) \approx \frac{1}{3}\sqrt{\frac{2}{\pi}} y^{3/2} + O\left(y^{5/2}\right) \tag{B.83}$$

and

$$Y_{3/2}(y) = \sqrt{\frac{2}{\pi y}}\left(-\sin y - \frac{\cos y}{y}\right) \approx -\frac{\sqrt{\frac{2}{\pi}}}{y^{3/2}} - \frac{\sqrt{y}}{\sqrt{2\pi}} + O\left(y^{5/2}\right). \tag{B.84}$$

Because $Y_{3/2}(y)$ diverges at the origin ($\tau \to 0$), we choose $C_2 = 0$. Switching back to the gravitational potential Φ,

$$\Phi(k, \tau) = \frac{C_1}{y^2}\left(\frac{\sin y}{y} - \cos y\right), \tag{B.85}$$

and the density contrast is

$$\delta(k, \tau) = 2C_1\left[\left(\frac{2 - y^2}{y^2}\right)\left(\frac{\sin y}{y} - \cos y\right) - \frac{\sin y}{y}\right]. \tag{B.86}$$

For long-wavelength perturbations ($y \ll 1$),

$$\Phi(k, y \ll 1) = \frac{1}{3}C_1 \equiv \Phi_k^0 \tag{B.87}$$

and

$$\delta(k, y \ll 1) = -2C_1 \equiv -6\Phi_k^0. \tag{B.88}$$

where we called $\Phi_k(y \ll 1) = \Phi_k^0$ to relate to the initial gravitational potential during inflation, determined in the next chapter.

For short-wavelength perturbations ($y \gg 1$),

$$\Phi_k(y \gg 1) \to -3\Phi_k^0 \frac{\cos y}{y^2} = -\frac{9\Phi_k^0}{(k\tau)^2}\cos\left(\frac{k\tau}{\sqrt{3}}\right) \tag{B.89}$$

and

$$\delta_k(y \gg 1) \to -6\Phi_k^0 \cos y. \tag{B.90}$$

In summary, during the radiation dominated era, the gravitational potential has a frozen amplitude long-wavelengths modes (super-horizon, $y \ll 1$), decaying at the moment it enters the horizon.

Vector equations For the vector contributions, we have

$$16\pi G a^2 \delta T^0_{\ i(V)} = \nabla^2 V_i, \tag{B.91}$$

$$-16\pi G a^2 \delta T^i_{\ j(V)} = (V_{i,j} + V_{j,i})' + 2\mathcal{H}(V_{i,j} + V_{j,i}), \tag{B.92}$$

where the subscript (V) refers to the vector part of the gauge-invariant energy-momentum tensor. The solution of the second equation yields $V_i \propto a^{-2}$ (Mukhanov, 2005, Section 7.3.2), decaying very quickly with time. Therefore, they will only be relevant if their amplitude, at primordial times, is very large. There are no known mechanism or models, so far, that can generate large primordial vector perturbations.

Tensor equations The tensor modes that describe gravitational waves obey the equation:

$$16\pi G a^2 \delta T^i_{\ j(T)} = h^i_{\ j}{}'' + 2\mathcal{H} h^i_{\ j}{}' - \nabla^2 h^i_{\ j}. \tag{B.93}$$

Notice that the left-hand side of the above equation is non-zero in case where there is anisotropic stress.

B.5 Observables in cosmology

Consider the power spectrum as a measure of the correlation of the density contrast or the gravitational potential:

$$P_\delta(k) = \frac{2\pi^2}{k^3}|\delta|^2. \tag{B.94}$$

The behaviour of perturbations can be analysed for two different regimes: $k\tau > 1$ and $k\tau < 1$. We saw that, during the radiation era, matter perturbations behave differently compared with the matter era. Because during inflation the comoving Hubble radius decrease with time, the majority of modes of a certain wavenumber k will eventually become super-horizon modes. With the standard Big Bang evolution, however, the Hubble radius starts to grow again, and these modes will eventually become sub-horizon.

For the modes that entered the horizon long after matter-radiation equality, the evolution is obtained during the matter era. In this case, perturbations grow with k^2 (see Equation B.77)

and $P(k) \propto k$. For those modes that entered the horizon well before equality, the evolution is what we have found for a radiation-dominated universe: they are frozen and hence $P(k) \propto k^{-3}$ for scales which were small enough to become sub-horizon at the time of matter-radiation equality (i.e. with modes $k \gg k_{eq}$). The turn-over point in Figure B.3 marks the matter-radiation transition.

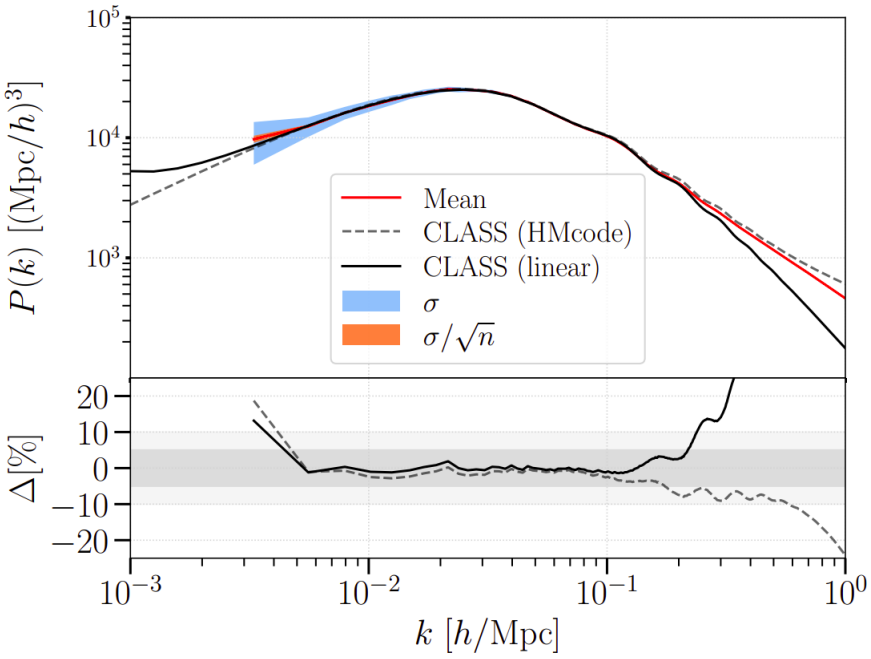


Figure B.3: (*Top*) Matter power spectrum (CDM and baryons) at $z \simeq 0$. The solid-red curve shows the mean value of 34 random realisations for the fiducial cosmology of the simulations presented in Coates et al. (2021). The standard deviation and error on the mean are shown in blue and orange (68% confidence intervals) respectively. The black-solid and dashed curves show the theoretical predictions using the `CLASS` linear power spectrum and non-linear (`HMcode`) model, respectively. HMcode uses the synchronous gauge while the other cases are in Poisson gauge which, for scalar perturbations, coincides with the Newtonian gauge. The different behaviour around the horizon scale comes from connecting the synchronous density contrast to its counterpart in the Newtonian gauge (Equation B.96). (*Bottom*) Fractional difference between the theoretical predictions and the mean of the simulations. The light and darker grey regions depict the 10% and 5% differences, respectively. [Extracted from Coates et al. (2021)].

However, when dealing with observables, such as the galaxy number density, we need to connect the density contrast obtained in theory with the number counts. The synchronous gauge is commonly used to relate the dark matter and the tracer overdensities, δ and δ_g respectively, via the standard linear biasing relation:

$$\delta_g^{\rm syn} = b\,\delta^{\rm syn}. \tag{B.95}$$

However, quantities such as the matter overdensity are usually obtained in the Newtonian gauge, as developed in Section B.4.1, or in some non-trivial gauges if they are obtained via numerical

simulations (Chisari and Zaldarriaga, 2011; Fidler et al., 2015, 2019).

The Newtonian-gauge number density is related to the synchronous one via

$$\delta_g^{\rm N} = b\,\delta^{\rm syn} + \left[\frac{{\rm d}\ln(a^3\,n_g)}{{\rm d}\tau} - 3\,\mathcal{H}\right]\frac{v}{k}, \tag{B.96}$$

where n_g is the comoving number density of galaxies (or sources) as a function of time and v is the Newtonian-gauge velocity of the sources (Challinor and Lewis, 2011). As one can see, the Newtonian overdensity $\delta^{\rm N}$ carries a k^{-1} term that diverges on large scales. We show the differences between the Newtonian and synchronous gauge power spectra in Figure (B.3).

Because in the power spectrum the divergence behaves as k^{-2}, there is a possible degeneracy with the scale-dependent bias that comes from local primordial non-Gaussianity (Chapter 4). This motivates the gauge-invariant description of the galaxy number density: either by computing it from a gauge-invariant perspective (Yoo, 2014), or by using some fixed gauge, but constructing the observed number density instead of performing a trivial (but naive at large scales) connection between the Newtonian-gauge δ and the galaxy overdensity δ_g.

Appendix C

Tracing the Large-Scale Structure of the Universe

Understanding the connection between the luminous tracers (e.g. quasars, Lyman-α forests, galaxies, cluster of galaxies, 21-cm line-intensity mapping, etc) of the dark matter and the matter density field is a crucial step to interpret cosmological surveys, which ultimately is used as a tool to comprehend the evolution of the universe from the perspective of modern physics. This appendix summarises how this link is achieved within the so-called halo model.

C.1 Halo model

The halo model is a way to predict and interpret the galaxy distribution under the assumption that every galaxy resides in a dark matter halo. Within this framework, the distribution of matter is assumed to be made up of distinct dark matter halos, with a variety of sizes (radius) and, therefore, masses. Initial approaches to link galaxies with the dark matter field accounted for the fact that galaxies are discrete objects to use statistical properties of a point sampling distribution. However, to do that it is required the knowledge of the distribution of cluster sizes, of points around the cluster centre, and a description of the clustering of the clusters (Cooray and Sheth, 2002). Hence, to understand how galaxies populate and are distributed within these dark matter halos should be enough to describe the statistics of the observed galaxy distribution (Hearin et al., 2016).

In order to describe the spatial statistics of dark matter halo, we separate between to distinct regions. Firstly, we require knowledge of internal properties (small scales) of the halos, such as its profile (how dark matter particles are distributed around the halo centre) and abundance (the number density of halos of a certain mass). This is encoded in the so-called 1-halo term. Secondly, the large scales will be described by their clustering properties, which gives rise to the bias. At these scales, encoded in the 2-halo term, the 1-halo properties will not matter too

much.

For short, the halo model programme consists of knowing **(1)** how many high and low mass halos we have (halo abundance); **(2)** how the mass is distributed around the centre of the halo (halo profile), and **(3)** how the halos are clustered in the universe (halo clustering). Once we have the halo mass and its profile, we can ask how many galaxies can be placed inside a halo. This last point, which brings up the halo-galaxy connection, is described via what is called a halo occupation distribution (HOD).

The properties and functions defined in this appendix neglects the halo dependence on properties beyond their mass. However, such "universal" properties do not hold in general and the inclusion of more parameters, or a more complete modelling, is required to reproduce some quantities such as the evolution (Beutler and Dio, 2020; Guandalin et al., 2020) and higher-order biases (Dizgah et al., 2021).

C.1.1 Halo mass function

We shall assume that all mass in the universe is in the form of dark matter halos, which are distinct bounded units of some mass M resulting from a collapse that virialised. First we need to define what is a halo. The simplest approach is to consider the spherical collapse (see Chapter 2 of da Silva, 2017, for a review) to model structure formation in the non-linear regime. This approach carries no approximations except for the (spherical) shape of the collapsed regions and the cosmological assumptions considered (FLRW universe). Within this set-up, halos can be modelled as spherical regions that collapsed after the local overdensity δ passed the critical threshold $\delta_c = 1.686$, established by the spherical collapse theory in an Einstein-de Sitter (matter only) universe, and finally virialised.

The variance $\sigma^2(z, R)$ of the linear density field filtered at some scale R by a window function, $W(x, R)$, which we consider to be a top-hat in real space,

$$W(x, R) = \begin{cases} 1, & r < R \\ 0, & r > R \end{cases}, \tag{C.1}$$

is given by

$$\sigma^2(z, R) = \int k^2 \frac{\mathrm{d}k}{2\pi^2} |\tilde{W}(k, R)|^2 P(z, k), \tag{C.2}$$

where $P(z, k)$ is the linear matter power spectrum at redshift z, and

$$\tilde{W}(x) = \frac{3}{x^2}\left(\frac{\sin x}{x} - \cos x\right).$$

is the Fourier transform of Equation (C.1).

Because the window function in Equation (C.1) is in configuration space, we can relate the radius R to the mass M that is enclosed at this scale, via the relation $R^3 = 3M/(4\pi\bar{\rho}_{m,0})$, where $\bar{\rho}_{m,0} = \Omega_{m,0}\,\rho_{c,0}$ and $\rho_{c,0}$ is the critical density of the universe today, also given by the relation $\rho_{c,0} = 3H_0^2/(8\pi G)$. For other window functions, however, this relation is not straightforward (Maggiore and Riotto, 2010).

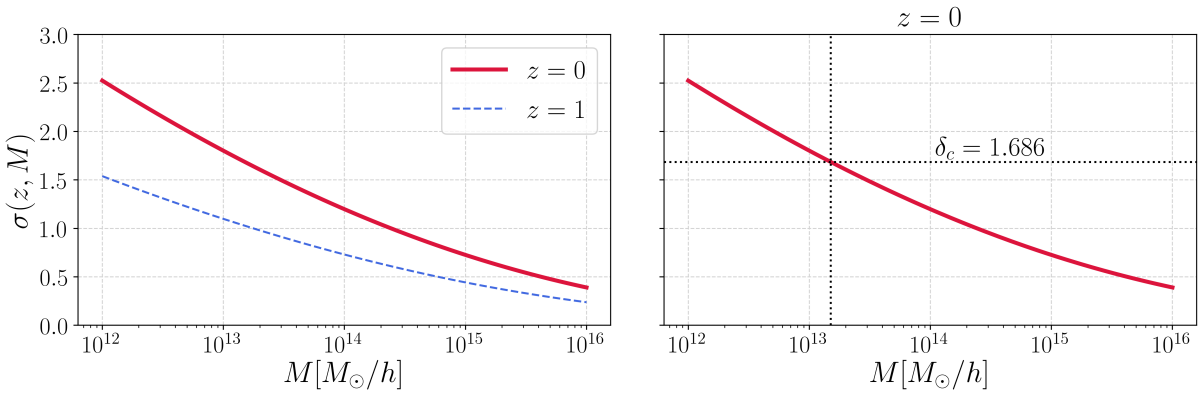


Figure C.1: (*Left*) Standard deviation $\sigma(z, M)$ at redshifts $z = 0$ (*thick red*) and $z = 1$ (*dashed blue*), for the mass range $10^{12} < M < 10^{16}\,[M_\odot/h]$. (*Right*) The standard deviation at $z = 0$. Horizontal dotted line shows the value $\sigma = 1.686$, corresponding to the linear extrapolation of the density perturbation δ_c for the spherical collapse model. Vertical dotted line corresponds to the mass $M_\star \approx 1.5 \times 10^{13}[M_\odot/h]$ (scale $R_\star \approx 3.5$ Mpc/h) for the cosmology utilised in this example: $h = 0.68, \omega_\mathrm{b} = 0.022$, $\omega_\mathrm{cdm} = 0.12, A_s = 2.1 \times 10^{-9}$, and $n_s = 0.97$.

In Figure C.1 we show the standard deviation (matter overdensity field) as a function of the mass enclosed at a scale R, for two redshifts (left), and at $z = 0$ with $\sigma(z = 0, M) = \delta_c = 1.686$ and the corresponding mass $M_\star \approx 1.88 \times 10^{13}[M_\odot/h]$. Since $R \propto M^{1/3}$, larger masses imply larger scales and the variance decreases: if $M \to \infty$, $R \to \infty$, which means that we are looking at the entire universe. From the hypothesis of homogeneity assumed, we expect $\sigma \to 0$.

The halo mass function gives the abundance of halos, of a certain mass M, that has virialised at redshift z. It is given by Equation (6.3), which we reproduce below:

$$\frac{\mathrm{d}n(M,z)}{\mathrm{d}\ln M} = \frac{\bar{\rho}_m}{M}\frac{\mathrm{d}\ln\sigma^{-1}}{\mathrm{d}\ln M} f(\sigma, z). \tag{C.3}$$

Here, n is the number density of halos, $\nu = \delta_c^2(z)/\sigma^2(M)$, and

$$f(\sigma) = \frac{M}{\bar{\rho}_m}\frac{\mathrm{d}n(z,M)}{\mathrm{d}\ln\sigma^{-1}}, \tag{C.4}$$

is called the multiplicity function. It is usually fitted from numerical simulations and, sometimes, $f(\sigma)$ is called the mass-function.

There are many $f(\sigma)$ available in the literature: the phenomenological fit of Tinker et al. (2008) ($f(\sigma)$ and $g(\sigma)$, with the latter normalised to unity when integrated over $\ln\sigma^{-1}$), the

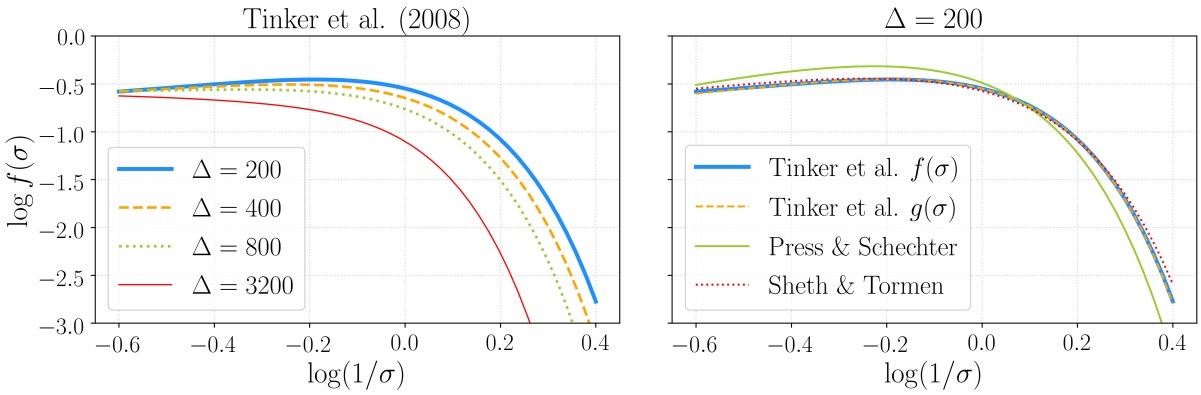


Figure C.2: (*Left*) Multiplicity function $f(\sigma)$, and the normalised $g(\sigma)$ for different values of Δ. (*Right*) Tinker et al. (2008) multiplicity functions $f(\sigma)$ and $g(\sigma)$, for $\Delta = 200$, and the Press-Schechter and Sheth-Tormen multiplicities (Equations 2.29 and 2.30 of da Silva, 2017, respectively).

expression obtained by Press and Schechter (1974), based on spherical collapse and Gaussian initial conditions, and the Sheth and Tormen (1999) function, which considers an ellipsoidal collapse. The Tinker et al. (2008) multiplicity functions have free parameters that are fitted from dark matter only N-body simulations. For example,

$$f(\sigma) = A\left[\left(\frac{\sigma}{b}\right)^{-a} + 1\right]\exp\left(-\frac{c}{\sigma^2}\right), \tag{C.5}$$

with A, a, b, c varying according to the overdensity of matter perturbations at virialisation $\Delta = 3\,M/4\pi R_v^3\bar{\rho}_m(z_v)$; R_v and z_v are, respectively, the radius and redshift at virilisation. In Figure C.2, we show the four aforementioned functions, with the Tinker et al. (2008) being computed for different values of Δ.

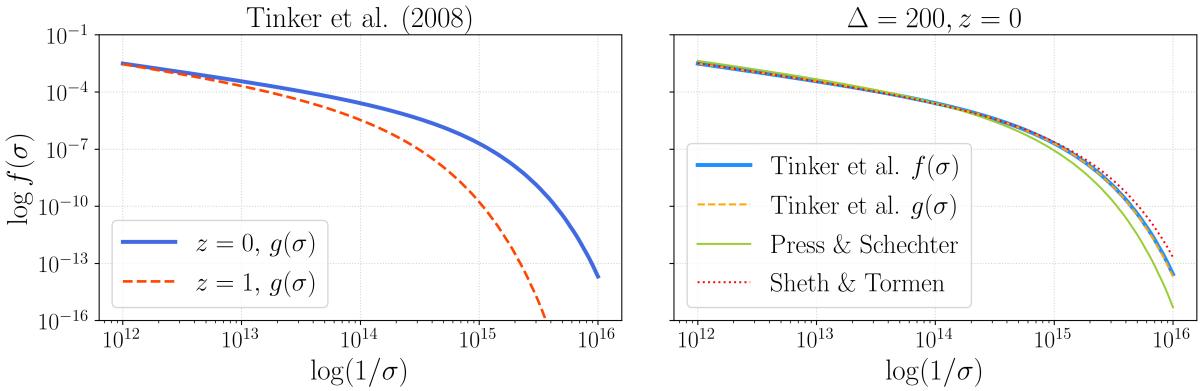


Figure C.3: (*Left*) Halo mass function using the multiplicity function fit from Tinker et al. (2008) at redshifts $z = 0$ (*thick blue*), and $z = 1$ (*thin red*). This gives us the expected number of halos of mass M, within the mass range $M_i < M < M_{i+1}$, at redshift z. (*Right*) Mass function obtained with the different multiplicity functions shown in Figure C.2.

With closed expressions for the multiplicity function, we can compute the mass func-

tion, shown in Figure C.3. As we can see on the left panel, for larger redshifts we observe a suppression in the abundance of large-mass halos. This is expected, since more massive objects require a larger formation time.

Finally, what simulations give us is the number density $\bar{n}_i$ of halos inside a mass bin $[M_i, M_{i+1}]$. This can be obtained from the mass function, Equation (C.3), after integrating over all masses within that bin:

$$\bar{n}_i(z) = \int_{M_i}^{M_{i+1}} \mathrm{d}\ln M \, \frac{\mathrm{d}n(z, M)}{\mathrm{d}\ln M}. \tag{C.6}$$

C.1.2 Halo profile

Finally, we compute the NFW halo profile, which describes an universal distribution for matter (e.g. galaxies) inside halos:

$$\rho(r|M_{\rm vir}, z) = \frac{\rho_s}{c(z, M)\, r\, r_{\rm vir}^{-1}(1 + c(z, M)\, r/r_{\rm vir})^2}, \tag{C.7}$$

where

$$\rho_s = \frac{4\pi r_{\rm vir}^3}{M_{\rm vir}} \frac{1}{c^3(z, M)} \left[\ln[]1 + c(z, M)] - \frac{c(z, M)}{1 + c(z, M)}\right] \tag{C.8}$$

is a scale density of the halo, $r_{\rm vir}$ is the virial radius of a halo with virial mass $M_{\rm vir}$, and

$$c(z, M) = \frac{9}{1 + z}\left[\frac{M_{\rm vir}}{M_\star}\right]^{-0.13} \tag{C.9}$$

is the concentration parameter, which depends on cosmology via $M_\star$, the non-linear mass scale determined by the condition $\delta_c = \sigma$. The NFW profile is shown in the left panel of Figure C.4. We call the reader's attention to the unconventional units.

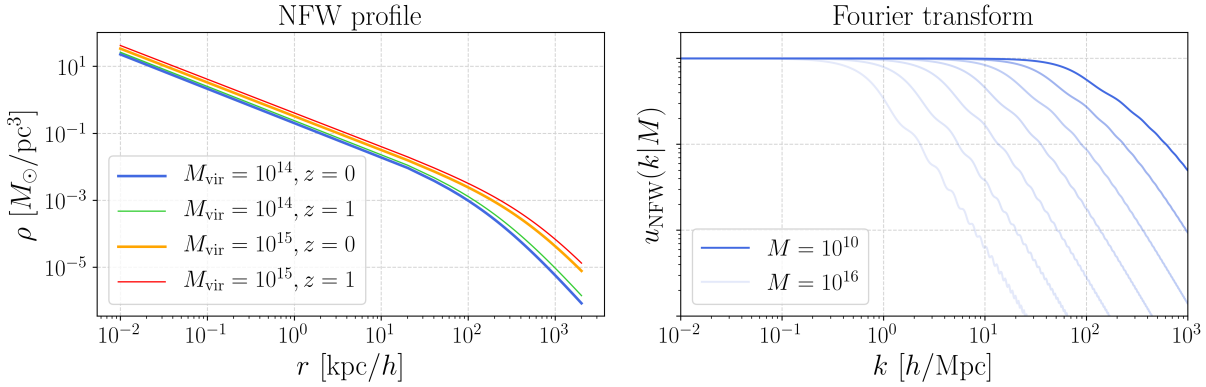


Figure C.4: Profiles at $z = 0$ and $z = 1$ for two virial masses, $M_{\rm vir} = 10^{14}\ M_\odot/h$ and $M_{\rm vir} = 10^{15}\ M_\odot/h$, in units of $M_\odot/\mathrm{pc}^3$ for r, the distance from the centre of the halo, in units of kpc/h.

For completeness we show, in the right panel of Figure C.4, the Fourier transform of

Equation (C.7) obtained from the analytical function given in Cooray and Sheth (2002):

$$u_{\rm NFW}(k|M_{\rm vir}) = \frac{4\pi\rho_s r_s^3}{M_{\rm vir}}\left(\mathcal{T}_1 - \mathcal{T}_2 + \mathcal{T}_3\right), \tag{C.10}$$

where Si and Ci are the sine and cosine integrals,

$$\mathrm{Ci} = -\int_x^\infty \mathrm{d}t\ \frac{\cos t}{t}, \quad \mathrm{Si} = \int_x^\infty \mathrm{d}t\ \frac{\sin t}{t}, \tag{C.11}$$

and

$$\mathcal{T}_1 = \sin(k\,r_s)[\mathrm{Si}([1+c]\,k\,r_s) - \mathrm{Si}(k\,r_s)], \tag{C.12}$$

$$\mathcal{T}_2 = \frac{\sin(c\,k\,r_s)}{(1+c)\,k\,r_s}, \tag{C.13}$$

$$\mathcal{T}_3 = \cos(k\,r_s)[\mathrm{Ci}([1+c]\,k\,r_s) - \mathrm{Ci}(k\,r_s)]. \tag{C.14}$$

In the above equations, $r_s = r_{\rm vir}/c(z, M)$.

C.1.3 Halo bias

The mass function is associated with the fraction of matter inside halos. If we want to obtain the correct spatial statistics for the halos, we must account for their position in space. Hence, it is important to introduce the halo bias, which is defined by the ratio of the halo power spectrum, $P_h(k)$, to the linear dark matter power spectrum, $P_{\rm lin}(k)$ (Tinker et al., 2010), as in Equation (6.4):

$$b^2(k) = \frac{P_h(k)}{P_{\rm lin}(k)}. \tag{C.15}$$

It is best understood within the context of the Peak-Background Split (PBS), where the long-wavelength modes δ_l enhances the probability of forming halos by decreasing the threshold δ_c (Figure C.5, right) for overdensities which are located at the peak δ_s (short modes) of large-scale (background) fluctuations δ_l: $\delta = \delta_l + \delta_s$. This is illustrated in Figure C.5.

It can be either derived from analytical mass functions, giving the Press-Schechter and Sheth-Tormen halo biases, or from Equation (C.15) via numerical simulations. From Tinker et al. (2010), the bias is given by

$$b(\nu) = 1 - A\frac{\nu^a}{\nu^a + \delta_c^a} + B\nu^b + C\nu^c, \tag{C.16}$$

where $\nu = \delta_c/\sigma$ and A, B, C, a, b, and c are fitted from simulations, depending on the matter perturbations at virialisation Δ, and can be found through the relations:

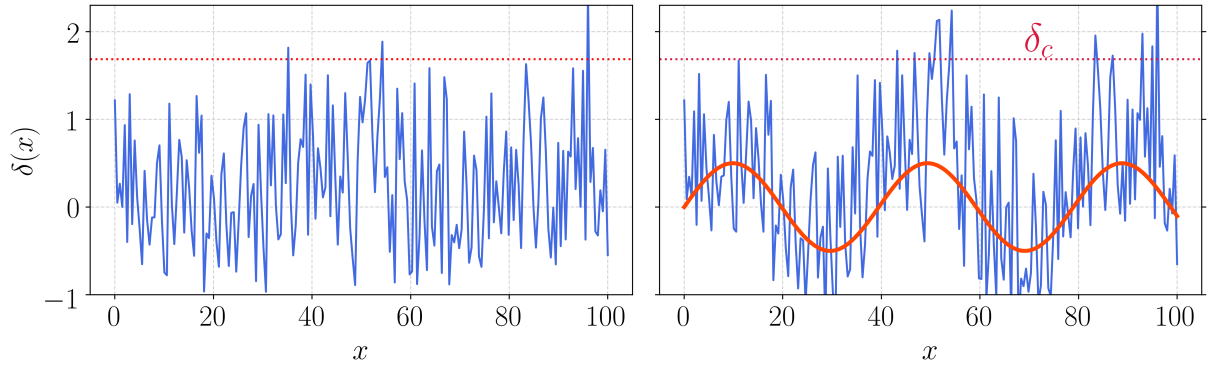


Figure C.5: The Peak-Background Split idea: the fluctuations of the peaks of the density field δ "sitting" over high densities of the background fluctuation have a lower threshold for the collapse, i.e. $\delta_s = \delta_c - \delta_l$. (*Left*) Random density field in the absence of long-wavelength (background) perturbations. (*Right*) Random density field depicted on the left, but in the presence of a background perturbation δ_l (*red curve*).

$$A = 1.0 + 0.24\,y\,\mathrm{e}^{-(4/y)^4} \quad B = 0.183 \quad C = 0.019 + 0.107\,y + 0.19\mathrm{e}^{-(4/y)^4}$$
$$a = 0.44y - 0.88 \quad b = 1.5 \quad c = 2.4$$

where $y = \log(\Delta)$, and $\nu = \delta_c/\sigma$.

Tinker et al. (2010) obtained C.16 with the power spectrum of 15 N-body simulations. Because their focus was on the large-scale bias, with Gaussian initial conditions, $b(\nu)$ is scale-independent: it was found by analysing the mean of the 10 (large boxes) to the 5 (smaller boxes) largest wavelengths of the simulations set.

We show the bias behaviour as a function of halo mass in Figure C.6. For comparison, we compare the Tinker bias with the Press-Schechter and Sheth-Tormen biases, found in da Silva (2017, p. 30).

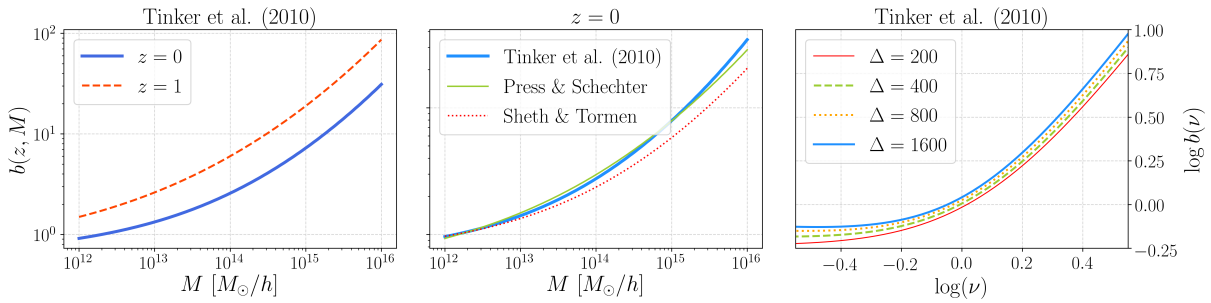


Figure C.6: (*Left*) Tinker bias for $\Delta = 200$, at redshifts $z = 0$ and $z = 1$. (*Middle*) Biases obtained from different fits, at $z = 0$. The Press-Schechter and Sheth-Tormen biases can be found in da Silva (2017, p. 30). (*Right*) Large-scale bias for different values of Δ, as a function of ν.

C.2 Power spectrum

With all these ingredients at hands, we can compute the 1- and 2-halo power-spectrum terms of the halo model. When summed up, they describe the clustering statistics of the dark matter distribution in the universe from the linear regime (dominated by the 2-halo term) up to the non-linear scales (1-halo term). The names come from the statistical description: the 1-halo term computes the spectrum of dark matter within the dark matter halos themselves, whereas the 2-halo term computes the spectrum for the statistics of distinct dark matter halos. With the halo model and numerical errors that exist, it agrees within 40% with the `halofit` spectrum given by the `CLASS` code for $\Delta = 400$, and within 20% for $\Delta = 200$. We show these results in Figure C.7.

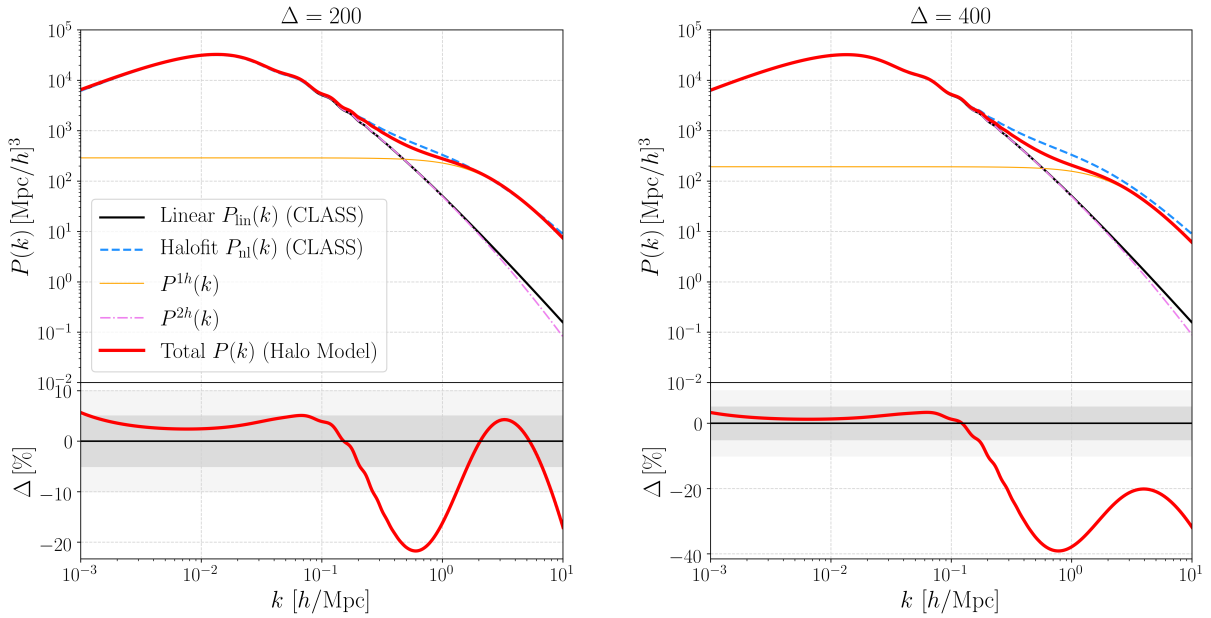


Figure C.7: Comparison between the 1- and 2-halo contributions to the matter power spectrum with the linear and non-linear (`halofit`) spectra from `CLASS`. The full non-linear halo model spectrum, obtained with the ingredients described in the previous sections, is also shown for the case of the Tinker mass function and large-scale halo bias. In this case, the accuracy found between our power spectrum and the `halofit` was found to be around 20% for $\Delta = 200$ (*left*) and 40% for $\Delta = 400$ (*right*). Light and darker-grey shaded regions represent the 10% and 5% regions, respectively.

Appendix D

Fisher Matrix Formalism

> "*Physics is, after all, the science of measurement. That is, physics is a quantification of* observed *phenomena. And observations contain noise, or fluctuations. The physical paradigm equations define the fluctuations or errors from ideal values that occur in such observations. That is,* the physics lies in the fluctuations."
>
> — Bernard Roy Frieden, 1998.

D.1 Definitions

Suppose we have a set of N observations, $\boldsymbol{x} = \{x_1, \ldots, x_N\}$, which is drawn from the probability density function (PDF) $P(X)$ of the random variable X. Consider also a model parametrized by the set of M parameters $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_M\}$. How much can we learn about the model from this data set?

One way to tackle this question is by asking how probable it is to get $\boldsymbol{x}$ assuming that it follows some PDF which is parametrized by the unknown parameters via the model assumed. As an example, we could assume the data to be drawn from a Gaussian distribution $P(\boldsymbol{x}; \boldsymbol{\theta})$ (Kay, 1993):

$$P(\boldsymbol{x}; \boldsymbol{\theta}) = (2\pi\sigma^2)^{-N/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{n=1}^{N}[x_n - \bar{x}_n(\boldsymbol{\theta})]\right), \tag{D.1}$$

where $\bar{x}_n(\boldsymbol{\theta})$ is the hypothetical model. If the data is irrelevant to test the theoretical assumption, then no information can be extracted from $\boldsymbol{x}$ for the model $\bar{x}(\boldsymbol{\theta})$ and, therefore, the parameters cannot be constrained. If $\boldsymbol{\theta}$ is considered to be a deterministic variable, we are in the realm of classical estimation theory. If, on the other hand, $\boldsymbol{\theta}$ is also a random variable, incorporating previous knowledge or assumptions on the possible values $\boldsymbol{\theta}$ can take, we enter the field of Bayesian estimation theory.

Regardless of being within the classical or Bayesian theory, the goal is to determine $\boldsymbol{\theta}$

based on $\boldsymbol{x}$. In order to do that, one can build an estimator $\hat{\boldsymbol{\theta}}$ through some function f:

$$\hat{\boldsymbol{\theta}} = f(\boldsymbol{x}). \tag{D.2}$$

If the information contained in the experiment is poor, then the error on $\hat{\boldsymbol{\theta}}$ is large.

Finally, by increasing the of observations we expect $\boldsymbol{\theta}$ to be better constrained. The Fisher matrix is a mathematical way to assess the information content of a set of observations, satisfying the requirements of being conditional on the relevance of the data set for the hypothesis testing, increasing with the number of observations, and being related to the experiment's precision (James, 2006, Chapter 5). It is defined by

$$F_{ij} \equiv -\left\langle \frac{\partial^2 \mathcal{L}(\boldsymbol{x}, \boldsymbol{\theta})}{\partial\theta_i \partial\theta_j} \right\rangle, \tag{D.3}$$

where $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\theta}) \equiv -\ln P(\boldsymbol{x}, \boldsymbol{\theta})$ is called the log-likelihood function. It is not a PDF, because in this case $\boldsymbol{x}$ is a fixed set of observations. If we have a broad likelihood, then $\boldsymbol{\theta}$ is poorly constrained, that is, has a large variance. Therefore, the curvature at the peak of the likelihood function is directly associated with the estimator's performance, and Equation (D.3) should be evaluated at the peak value (that is, for the "true" parameters, at the likelihood maximum).

We follow Tegmark et al. (1997) and consider

$$P(\boldsymbol{x}, \boldsymbol{\theta}) = \frac{(2\pi)^{-N/2}}{\sqrt{\det C}} \exp\left[-\frac{1}{2} (\boldsymbol{x} - \bar{\boldsymbol{x}})^\mathsf{T} C_{ij}^{-1} (\boldsymbol{x} - \bar{\boldsymbol{x}}) \right], \tag{D.4}$$

where $C_{ij} \equiv \mathrm{Cov}(X_i, X_j)$ is the covariance matrix of the data set, then

$$-2\mathcal{L} \propto \ln(\det C) + (\boldsymbol{x} - \bar{\boldsymbol{x}})^\mathsf{T} C_{ij}^{-1} (\boldsymbol{x} - \bar{\boldsymbol{x}}). \tag{D.5}$$

Defining the data matrix

$$D \equiv (\boldsymbol{x} - \bar{\boldsymbol{x}}) (\boldsymbol{x} - \bar{\boldsymbol{x}})^\mathsf{T} \equiv \boldsymbol{d}\, \boldsymbol{d}^\mathsf{T}, \tag{D.6}$$

the covariance can be written as $C = \langle D \rangle$. Using the known result $\ln\det C = \mathrm{Tr}[\ln C]$ and $\boldsymbol{d}^\mathsf{T} C^{-1} \boldsymbol{d}^\mathsf{T} = \mathrm{Tr}[\boldsymbol{d}\, C^{-1}\, \boldsymbol{d}^\mathsf{T}] = \mathrm{Tr}[\boldsymbol{d}^\mathsf{T}\, \boldsymbol{d}\, C^{-1}]$,

$$-2\mathcal{L} \propto \mathrm{Tr}\left[\ln C + C^{-1} D \right]. \tag{D.7}$$

Equation (D.3) becomes

$$F_{ij} = \frac{1}{2} \left\langle \frac{\partial^2}{\partial\theta_i \partial\theta_j} \mathrm{Tr}\left[\ln C + C^{-1} D \right] \right\rangle = \frac{1}{2} \left\langle \mathrm{Tr}\left[\frac{\partial}{\partial\theta_i} \left(C^{-1} C_{,j} - C^{-1} C_{,j} C^{-1} D + C^{-1} D_{,j} \right) \right] \right\rangle. \tag{D.8}$$

Therefore

$$F_{ij} = \frac{1}{2}\mathrm{Tr}\left[C^{-1}C_{,i}C^{-1}C_{,j} + 2C^{-1}\frac{\partial\bar{\boldsymbol{x}}}{\partial\theta_i}\frac{\partial\bar{\boldsymbol{x}}}{\partial\theta_j}\right], \tag{D.9}$$

after evaluating the derivatives at the likelihood maximum (i.e. $D_{,i} = 0$).

D.2 Power-spectrum Fisher matrix

In this section we derive Equation (3.26). Recall that our data $\boldsymbol{x}$ is the density contrast δ_α of some tracer α, and the parameters we would like to estimate are the band powers i of the power spectrum $P_\alpha(k_i)$ of tracer α. In this case, $\langle\boldsymbol{x}\rangle = \langle\delta_\alpha\rangle = 0$. The multi-tracer Fisher matrix can be written explicitly as the trace over the integer part (i.e. tracers) and the continuous (redshift space) positions, that is, a generalisation of Equation (D.9) (Abramo et al., 2015, ASL):

$$F_{\mu,i;\nu,j} = \frac{1}{2}\sum_{\alpha\beta\gamma\sigma}\int \mathrm{d}^3x_1\,\mathrm{d}^3x_2\,\mathrm{d}^3x_3\,\mathrm{d}^3x_4\,\left[C^{-1}_{\alpha\beta}(\boldsymbol{x}_1,\boldsymbol{x}_2)\frac{\partial C_{\beta\gamma}(\boldsymbol{x}_2,\boldsymbol{x}_3)}{\partial P_{\mu,i}}\times\right.$$
$$\left.\times C^{-1}_{\gamma\sigma}(\boldsymbol{x}_3,\boldsymbol{x}_4)\frac{\partial C_{\sigma\alpha}(\boldsymbol{x}_4,\boldsymbol{x}_1)}{\partial P_{\nu,j}}\right], \tag{D.10}$$

with the covariances $C_{\alpha\beta}(\boldsymbol{x}_1,\boldsymbol{x}_2)$ given in terms of theoretical quantities

$$C_{\alpha\beta}(\boldsymbol{x}_1,\boldsymbol{x}_2) = \int\frac{\mathrm{d}^3k}{(2\pi)^3}\,\mathrm{e}^{i\boldsymbol{k}\cdot(\boldsymbol{x}_1-\boldsymbol{x}_2)}\left[P(\bar{x},\boldsymbol{k})\,b_\alpha(\boldsymbol{x}_1,\boldsymbol{k})\,b_\beta(\boldsymbol{x}_2,\boldsymbol{k}) + \frac{\delta_{\alpha\beta}}{\sqrt{\bar{n}_\alpha(\boldsymbol{x}_1)\,\bar{n}_\beta(\boldsymbol{x}_2)}}\right]. \tag{D.11}$$

Above, b_α is the generalised bias of the tracers, such that its theoretical power spectrum is $P_\alpha = b_\alpha^2\,P$, as defined in Equation (3.7), with P the matter power spectrum.

Using the stationary phase approximation, which assumes that both the bias and power spectrum of the tracers vary slowly with $\boldsymbol{k}$, the covariance in Equation (D.11) becomes

$$C_{\alpha\beta}(\boldsymbol{x}_1,\boldsymbol{x}_2) \approx \delta^D(\boldsymbol{x}_1-\boldsymbol{x}_2)\left[P(\bar{x},\boldsymbol{k})b_\alpha(\boldsymbol{x}_1,\boldsymbol{k})b_\beta(\boldsymbol{x}_2,\boldsymbol{k}) + \frac{\delta_{\alpha\beta}}{\sqrt{\bar{n}_\alpha(\boldsymbol{x}_1)\bar{n}_\beta(\boldsymbol{x}_2)}}\right]. \tag{D.12}$$

Inverting The first object we tackle in order to obtain an expression for Equation (D.10) is the inverse of the covariance matrix of Equation (D.12). To do that, we follow ASL. The inverse of the covariance must obey the integral constraint

$$\sum_\beta\int \mathrm{d}^3x_2\,C_{\alpha\beta}(\boldsymbol{x}_1,\boldsymbol{x}_2)\,C^{-1}_{\beta\gamma}(\boldsymbol{x}_2,\boldsymbol{x}_3) = \delta_{\alpha\gamma}\delta^D(\boldsymbol{x}_1-\boldsymbol{x}_3). \tag{D.13}$$

We can write Equation (D.12) as

$$C_{\alpha\beta}(\boldsymbol{x}_1,\boldsymbol{x}_2) \approx \delta^D(\boldsymbol{x}_1-\boldsymbol{x}_2)M_{\alpha\beta}, \tag{D.14}$$

where we defined

$$M_{\alpha\beta} \equiv b_\alpha(\boldsymbol{x}_1, \boldsymbol{k})\, P(\bar{x}, \boldsymbol{k})\, b_\beta(\boldsymbol{x}_2, \boldsymbol{k}) + \frac{\delta_{\alpha\beta}}{\sqrt{\bar{n}_\alpha(\boldsymbol{x}_1)\, \bar{n}_\beta(\boldsymbol{x}_2)}}. \tag{D.15}$$

Since the inverse of $\delta^D(\boldsymbol{x}_1 - \boldsymbol{x}_2)$ is the Dirac delta itself,

$$\int \mathrm{d}^3 x_2 \delta^D(\boldsymbol{x}_1 - \boldsymbol{x}_2)\delta^D(\boldsymbol{x}_2 - \boldsymbol{x}_3) = \delta^D(\boldsymbol{x}_1 - \boldsymbol{x}_3), \tag{D.16}$$

$$C^{-1}_{\alpha\beta}(\boldsymbol{x}_1, \boldsymbol{x}_2) \approx \delta^D(\boldsymbol{x}_1 - \boldsymbol{x}_2)\, M^{-1}_{\alpha\beta}. \tag{D.17}$$

Above, $M^{-1}_{\alpha\beta}$ can be directly obtained if we make use of the so-called Sherman-Morrison-Woodbury identity, which tells us that, for $M_{\alpha\beta} = \delta_{\alpha\beta} + v_\alpha v_\beta$, $M^{-1}_{\alpha\beta} = \delta_{\alpha\beta} - v_\alpha v_\beta/(1 + v^2)$, where $v^2 = \sum_\mu v_\mu^2$. Thus:

$$\bar{n}_\alpha^{1/2}\bar{n}_\beta^{1/2} M_{\alpha\beta} = \delta_{\alpha\beta} + (\bar{n}_\alpha^{1/2}\, P_\alpha^{1/2})(\bar{n}_\beta^{1/2}\, P_\beta^{1/2}), \tag{D.18}$$

where we made use of the fact that $b_\alpha b_\beta P \equiv P_{\alpha\beta} = (P_\alpha P_\beta)^{1/2}$. By identifying $\bar{n}_\alpha P_\alpha^{1/2}$ with v_α, we arrive at

$$\bar{n}_\alpha^{-1/2}\, \bar{n}_\beta^{-1/2} M^{-1}_{\alpha\beta} = \delta_{\alpha\beta} - \frac{(\bar{n}_\alpha^{1/2}\, P_\alpha^{1/2})(\bar{n}_\beta^{1/2}\, P_\beta^{1/2})}{1 + \sum_\mu \bar{n}_\mu\, P_\mu}. \tag{D.19}$$

Therefore

$$C^{-1}_{\alpha\beta}(\boldsymbol{x}_1, \boldsymbol{x}_2) \approx \delta^D(\boldsymbol{x}_1 - \boldsymbol{x}_2)\left[\bar{n}_\alpha(\boldsymbol{x}_1)\, \delta_{\alpha\beta} - \bar{n}_\alpha(\boldsymbol{x}_1)\frac{b_\alpha(\boldsymbol{x}_1, \boldsymbol{k})\, P(\bar{x}, \boldsymbol{k})\, b_\beta(\boldsymbol{x}_2, \boldsymbol{k})}{1 + \sum_\mu \mathcal{P}_\mu}\, \bar{n}_\beta(\boldsymbol{x}_2)\right], \tag{D.20}$$

with $\mathcal{P}_\mu = \bar{n}_\mu(\boldsymbol{x}_1)\, b_\mu^2(\boldsymbol{x}, \boldsymbol{k})\, P(\boldsymbol{x}, \boldsymbol{k})$ being the clustering strength of each tracer.

Deriving It remains for us to find an expression for the derivatives appearing in Equation (D.10):

$$\frac{\partial C_{\alpha\beta}(\boldsymbol{x}_1, \boldsymbol{x}_2)}{\partial P_{\eta,i}} = \int \frac{\mathrm{d}^3 k}{(2\pi)^3}\, \mathrm{e}^{i\boldsymbol{k}\cdot(\boldsymbol{x}_1 - \boldsymbol{x}_2)}\, \frac{\partial P_{\alpha\beta}}{\partial P_{\eta,i}}, \tag{D.21}$$

where the power-spectrum appearing in the derivatives make reference to some tracer η, at real and Fourier space bins $\boldsymbol{x}_i$ and $\boldsymbol{k}_i$, with the former coming from a possible position dependence on the tracer bias. The above expression came from the fact that the shot noise term, which does not depend at all on cosmology, vanishes once we take the derivative with respect to the power spectrum.

Once again, we write $P_{\alpha\beta} = (P_\alpha P_\beta)^{1/2}$ to get

$$\frac{\partial P_{\alpha\beta}}{\partial P_{\eta,i}} = \frac{1}{2}\frac{1}{(P_\alpha P_\beta)^{1/2}}\left[\frac{\partial P_\alpha(\boldsymbol{x}_1,\boldsymbol{k})}{\partial P_\eta(\boldsymbol{x}_i,\boldsymbol{k}_i)}P_\beta(\boldsymbol{x}_2,\boldsymbol{k}) + P_\alpha(\boldsymbol{x}_1,\boldsymbol{k})\frac{\partial P_\beta(\boldsymbol{x}_2,\boldsymbol{k})}{\partial P_\eta(\boldsymbol{x}_i,\boldsymbol{k}_i)}\right], \tag{D.22}$$

$$= \frac{1}{2}\frac{1}{b_\alpha b_\beta P}\left[\frac{b_\alpha^2}{b_\eta^2}\delta_{\alpha\eta}\,\delta^i_{\boldsymbol{x}_1,\boldsymbol{k}}P_\beta + P_\alpha\frac{b_\beta^2}{b_\eta^2}\delta_{\beta\eta}\,\delta^i_{\boldsymbol{x}_2,\boldsymbol{k}}\right], \tag{D.23}$$

$$= \frac{b_\alpha b_\beta}{2b_\eta^2}\left(\delta_{\alpha\eta}\,\delta^i_{\boldsymbol{x}_1,\boldsymbol{k}} + \delta_{\beta\eta}\delta^i_{\boldsymbol{x}_2,\boldsymbol{k}}\right). \tag{D.24}$$

Above, we employed the notation introduced in ASL to denote

$$\frac{\partial P(\boldsymbol{x},\boldsymbol{k})}{\partial P(\boldsymbol{x}_i,\boldsymbol{k}_i)} = \delta^i_{\boldsymbol{x},\boldsymbol{k}} \tag{D.25}$$

as the restriction on integrals over the position and Fourier space to a certain volume, that is:

$$\int \frac{\mathrm{d}^3x\mathrm{d}^3k}{(2\pi)^3}(\ldots)\;\delta^i_{\boldsymbol{x},\boldsymbol{k}} \equiv \int_{V_i}\frac{\mathrm{d}^3x\mathrm{d}^3k}{(2\pi)^3}(\ldots), \tag{D.26}$$

with $V_i = V_{\boldsymbol{x}_i}V_{\boldsymbol{k}_i}$ the phase-space volume of some bin i. Therefore,

$$\frac{\partial C_{\alpha\beta}(\boldsymbol{x}_1,\boldsymbol{x}_2)}{\partial P_{\eta,i}} = \int\frac{\mathrm{d}^3k}{(2\pi)^3}\left[\mathrm{e}^{i\boldsymbol{k}\cdot(\boldsymbol{x}_1-\boldsymbol{x}_2)}\,\frac{b_\alpha(\boldsymbol{x}_1,\boldsymbol{k})b_\beta(\boldsymbol{x}_2,\boldsymbol{k})}{2b_\eta^2(\boldsymbol{x}_i,\boldsymbol{k}_i)}\left(\delta_{\alpha\eta}\,\delta^i_{\boldsymbol{x}_1,\boldsymbol{k}} + \delta_{\beta\eta}\,\delta^i_{\boldsymbol{x}_2,\boldsymbol{k}}\right)\right]. \tag{D.27}$$

Multi-tracer Fisher matrix Finally, with the expressions derived above, Equation (D.10) becomes

$$\begin{aligned}F_{\mu,i;\nu,j} \approx \frac{1}{2}\sum_{\alpha\beta\gamma\sigma}\int \mathrm{d}^3x_1\,\mathrm{d}^3x_2\,\mathrm{d}^3x_3\,\mathrm{d}^3x_4\;\delta^D(\boldsymbol{x}_1-\boldsymbol{x}_2)\left(\bar{n}_\alpha\,\delta_{\alpha\beta} - \bar{n}_\alpha\,\frac{b_\alpha\,P\,b_\beta}{1+\mathcal{P}}\bar{n}_\beta\right)\\ \times\int\frac{\mathrm{d}^3k}{(2\pi)^3}\,\mathrm{e}^{i\boldsymbol{k}\cdot(\boldsymbol{x}_2-\boldsymbol{x}_3)}\frac{b_\beta(\boldsymbol{x}_2,\boldsymbol{k})b_\beta(\boldsymbol{x}_3,\boldsymbol{k})}{2b_\mu^2(\boldsymbol{x}_i,\boldsymbol{k}_i)}\left(\delta_{\beta\mu}\,\delta^i_{\boldsymbol{x}_2,\boldsymbol{k}} + \delta_{\gamma\mu}\,\delta^i_{\boldsymbol{x}_3,\boldsymbol{k}}\right)\\ \times\delta^D(\boldsymbol{x}_3-\boldsymbol{x}_4)\left(\bar{n}_\gamma\,\delta_{\gamma\sigma} - \bar{n}_\gamma\frac{b_\gamma\,P\,b_\sigma}{1+\mathcal{P}}\bar{n}_\sigma\right)\\ \times\int\frac{\mathrm{d}^3k'}{(2\pi)^3}\,\mathrm{e}^{i\boldsymbol{k}'\cdot(\boldsymbol{x}_4-\boldsymbol{x}_1)}\frac{b_\sigma(\boldsymbol{x}_4,\boldsymbol{k}')b_\alpha(\boldsymbol{x}_1,\boldsymbol{k}')}{2b_\nu^2(\boldsymbol{x}_j,\boldsymbol{k}_j)}\left(\delta_{\sigma\nu}\,\delta^j_{\boldsymbol{x}_4,\boldsymbol{k}} + \delta_{\alpha\nu}\,\delta^j_{\boldsymbol{x}_1,\boldsymbol{k}}\right).\end{aligned} \tag{D.28}$$

If we integrate over $\boldsymbol{x}_4$, $\boldsymbol{x}_4 \to \boldsymbol{x}_3$, and over $\boldsymbol{x}_2$, $\boldsymbol{x}_2 \to \boldsymbol{x}_1$, we obtain

$$\begin{aligned}F_{\mu\nu} \approx \frac{1}{2}\sum_{\alpha\beta\gamma\sigma}\int\frac{\mathrm{d}^3k}{(2\pi)^3}\,\mathrm{d}^3x_1\left(\bar{n}_\alpha\delta_{\alpha\beta} - \bar{n}_\alpha\frac{b_\alpha\,P\,b_\beta}{1+\mathcal{P}}\bar{n}_\beta\right)\times\\ \times\int\frac{\mathrm{d}^3k}{(2\pi)^3}\,\mathrm{d}^3x_3\,\mathrm{e}^{i\boldsymbol{k}\cdot(\boldsymbol{x}_1-\boldsymbol{x}_3)}\mathrm{e}^{i\boldsymbol{k}'\cdot(\boldsymbol{x}_3-\boldsymbol{x}_1)}\left(\delta_{\beta\mu}\,\delta^i_{\boldsymbol{x}_1,\boldsymbol{k}} + \delta_{\gamma\mu}\,\delta^i_{\boldsymbol{x}_3,\boldsymbol{k}}\right)\frac{b_\alpha\,b_\beta}{2\beta_\mu^2}\end{aligned}$$

$$\times\left(\bar{n}_\gamma\delta_{\gamma\sigma}-\bar{n}_\gamma\frac{b_\gamma\,P\,b_\sigma}{1+\mathcal{P}}\bar{n}_\sigma\right)\times\left(\delta_{\sigma\nu}\,\delta^i_{\boldsymbol{x}_3,\boldsymbol{k}}+\delta_{\alpha\nu}\,\delta^i_{\boldsymbol{x}_1,\boldsymbol{k}}\right)\frac{b_\sigma b_\alpha}{2b_\nu^2}. \tag{D.29}$$

Once again, using the stationary phase approximation, $\mathrm{e}^{i\boldsymbol{k}\cdot(\boldsymbol{x}_1-\boldsymbol{x}_3)}\mathrm{e}^{i\boldsymbol{k}'\cdot(\boldsymbol{x}_3-\boldsymbol{x}_1)}=\mathrm{e}^{i(\boldsymbol{k}'-\boldsymbol{k})(\boldsymbol{x}_3-\boldsymbol{x}_1)}\approx(2\pi)^3\delta^D(\boldsymbol{k}'-\boldsymbol{k})\delta^D(\boldsymbol{x}_3-\boldsymbol{x}_1)$ leads to

$$\begin{aligned}F_{\mu\nu}\approx\frac{1}{8\,b_\mu^2b_\nu^2}\int\frac{\mathrm{d}^3x\,\mathrm{d}^3k}{(2\pi)^3}\sum_{\alpha\beta\gamma\sigma}&\left(\bar{n}_\alpha\delta_{\alpha\beta}-\bar{n}_\alpha\frac{b_\alpha\,P\,b_\beta}{1+\mathcal{P}}\bar{n}_\beta\right)\left(\delta_{\beta\mu}\,\delta^i_{\boldsymbol{x},\boldsymbol{k}}+\delta_{\gamma\mu}\,\delta^i_{\boldsymbol{x},\boldsymbol{k}}\right)\\&b_\alpha b_\beta\left(\bar{n}_\gamma\delta_{\gamma\sigma}-\bar{n}_\gamma\frac{b_\gamma\,P\,b_\sigma}{1+\mathcal{P}}\bar{n}_\sigma\right)\left(\delta_{\sigma\nu}\,\delta^i_{\boldsymbol{x},\boldsymbol{k}}+\delta_{\alpha\nu}\,\delta^i_{\boldsymbol{x},\boldsymbol{k}}\right)b_\sigma b_\alpha,\end{aligned} \tag{D.30}$$

which, by considering Equation (D.26), gives

$$F_{\mu\nu}\approx\frac{1}{8\,b_\mu^2b_\nu^2}\int_{V_i}\frac{\mathrm{d}^3x\,\mathrm{d}^3k}{(2\pi)^3}\sum_{\alpha\beta\gamma\sigma}(\ldots). \tag{D.31}$$

Expanding everything inside $(\ldots)$ yields 16 terms. For example, the first, second, fifth and sixth terms are, respectively:

$$\sum_{\alpha\beta\gamma\sigma}\delta_{\alpha\beta}\bar{n}_\alpha\delta_{\beta\mu}b_\beta b_\gamma\delta_{\gamma\sigma}\bar{n}_\gamma\delta_{\sigma\nu}b_\sigma b_\alpha=\bar{n}_\mu b_\mu^2\bar{n}_\nu b_\nu^2, \tag{D.32}$$

$$\sum_{\alpha\beta\gamma\sigma}\delta_{\alpha\beta}\bar{n}_\alpha\delta_{\beta\mu}b_\beta b_\gamma\delta_{\gamma\sigma}\bar{n}_\gamma\delta_{\alpha\nu}b_\sigma b_\alpha=\bar{n}_\mu b_\mu^2\sum_\sigma\bar{n}_\sigma b_\sigma^2, \tag{D.33}$$

$$\sum_{\alpha\beta\gamma\sigma}\delta_{\alpha\beta}\bar{n}_\alpha\delta_{\gamma\mu}b_\beta b_\gamma\delta_{\gamma\sigma}\bar{n}_\gamma\delta_{\sigma\nu}b_\sigma b_\alpha=\bar{n}_\mu b_\mu^2\sum_\beta\bar{n}_\beta b_\beta^2, \tag{D.34}$$

$$\sum_{\alpha\beta\gamma\sigma}\delta_{\alpha\beta}\bar{n}_\alpha\delta_{\gamma\mu}b_\beta b_\gamma\delta_{\gamma\sigma}\bar{n}_\gamma\delta_{\alpha\nu}b_\sigma b_\alpha=\bar{n}_\mu b_\mu^2\bar{n}_\nu b_\nu^2. \tag{D.35}$$

By summing Equation (D.32) with (D.35), and (D.33) with (D.34), we obtain $2\bar{n}_\mu b_\mu^2\bar{n}_\nu b_\nu^2$ and $2\bar{n}_\mu b_\mu^2\sum_\sigma\bar{n}_\sigma b_\sigma^2$, respectively. Other terms involving the total clustering strength appear analogously. By considering the sum of all these 16 terms, we are left with

$$F_{\mu\nu}\approx\frac{1}{8}\left[2\bar{n}_\mu b_\mu^2\bar{n}_\nu b_\nu^2+2\bar{n}_\mu b_\mu^2\sum_\sigma\bar{n}_\sigma b_\sigma^2-2\frac{\bar{n}_\mu b_\mu^2}{1+\mathcal{P}}\delta_{\mu\nu}\sum_{\alpha\beta}\bar{n}_\alpha b_\alpha^2P\bar{n}_\beta b_\beta^2-6\frac{\bar{n}_\mu b_\mu^2\bar{n}_\nu b_\nu^2}{1+\mathcal{P}}\mathcal{P}+\frac{4\bar{n}_\mu b_\mu^2\bar{n}_\nu b_\nu^2}{(1+\mathcal{P})^2}\right]. \tag{D.36}$$

Noticing that, by multiplying up and down by P^2, the P^2 in the denominator goes with $b_\mu^2\,P\,b_\nu^2\,P$ giving $P_\mu P_\nu$, and by simplifying the expression we arrive at the final result of ASL:

$$F_{\mu\nu}=\frac{1}{4}\frac{\delta_{\mu\nu}\mathcal{P}_\mu\mathcal{P}(1-\mathcal{P})+\mathcal{P}_\mu\mathcal{P}_\nu(1-\mathcal{P})}{(1+\mathcal{P})^2}. \tag{D.37}$$

This gives everything we need to build an optimal estimator for the power spectrum of LSS tracers.

D.3 Bispectrum Fisher matrix

In this section we present the Fisher matrix for the bispectrum utilised in Chapter 7. We propose the construction of the unbiased estimator $\hat{\phi}$ for the coefficients of the redshift distribution expansion, Equation (7.5):

$$\hat{\phi}_3^a = \sum_d \left(\mathcal{F}^{-1}\right)^a{}_d \sum_{\boldsymbol{k}_\perp,\boldsymbol{q}_\perp} \sum_{b,c} \sum_{b',c'} \mathcal{B}^{dbc}_{ghh}(\boldsymbol{k}_\perp,\boldsymbol{q}_\perp)\,(\mathsf{C}_3^{-1})^{cb,c'b'}(\boldsymbol{k}_\perp,\boldsymbol{q}_\perp)\,\hat{\mathcal{B}}^{Tb'c'}_{ghh}(\boldsymbol{k}_\perp,\boldsymbol{q}_\perp), \tag{D.38}$$

where

$$\mathsf{C}_3^{ab,cd}(\boldsymbol{k}_\perp,\boldsymbol{q}_\perp) \equiv \mathrm{Cov}[\hat{B}^{Tab}_{ghh}(\boldsymbol{k}_\perp,\boldsymbol{q}_\perp), \hat{\mathcal{B}}^{Tcd}_{ghh}(\boldsymbol{k}_\perp,\boldsymbol{q}_\perp)] \tag{D.39}$$

is the bispectrum covariance, $\hat{\mathcal{B}}^{Tbc}_{ghh}(\boldsymbol{k}_\perp,\boldsymbol{q}_\perp) = \langle\Delta_g(\boldsymbol{k}_\perp)\,\delta_h(\boldsymbol{q}_\perp)\delta_h(-\boldsymbol{k}_\perp-\boldsymbol{q}_\perp)\rangle$, and we called $\mathcal{F}$ the transverse Fisher matrix

$$\mathrm{Cov}^{-1}(\hat{\phi}_3^a,\hat{\phi}_3^d) \equiv \mathcal{F}^a{}_d = \sum_{\boldsymbol{k}_\perp,\boldsymbol{q}_\perp} \sum_{b,c} \sum_{b',c'} \mathcal{B}^{abc}_{ghh}(\boldsymbol{k}_\perp,\boldsymbol{q}_\perp)\,(\mathsf{C}_3^{-1})^{cb,c'b'}\,\mathcal{B}^{db'c'}_{ghh}(\boldsymbol{k}_\perp,\boldsymbol{q}_\perp). \tag{D.40}$$

To write this as proposed in Yadav et al. (2007), namely

$$F^B_{\alpha\beta} = \sum_{XX'} \frac{A}{4\pi}\frac{1}{6}\int_0^\infty \mathrm{d}k_\perp\,\mathrm{d}p_\perp\,\mathrm{d}q_\perp\,\frac{k_\perp\,p_\perp\,q_\perp}{\pi^2 A_T}\,\partial_\alpha\mathcal{B}^{XYZ}(k_\perp,p_\perp,q_\perp)\,\partial_\beta\mathcal{B}^{X'Y'Z'}(k_\perp,p_\perp,q_\perp)\times \\ \mathcal{I}^{XX'}(k_\perp)\,\mathcal{I}^{YY'}(p_\perp)\,\mathcal{I}^{ZZ'}(q_\perp), \tag{D.41}$$

we change the sum over the transverse vectors, $\boldsymbol{k}_\perp$ and $\boldsymbol{q}_\perp$, to an integral over the 2D moduli $k_\perp, q_\perp$ and $p_\perp = |\boldsymbol{k}_\perp + \boldsymbol{q}_\perp|$. By doing so, we consider the bispectrum estimator (Scoccimarro et al., 1998)

$$\hat{B}(k_\perp,q_\perp,p_\perp) = \mathcal{N}\int_{k_\perp}\mathrm{d}^2k_\perp\int_{q_\perp}\mathrm{d}^2q_\perp\int_{p_\perp}\mathrm{d}^2p_\perp\;\delta^D(k_\perp+q_\perp+p_\perp)\,\delta(\boldsymbol{k}_\perp)\,\delta(\boldsymbol{q}_\perp)\,\delta(\boldsymbol{p}_\perp), \tag{D.42}$$

where $\int_{k_\perp}\mathrm{d}^2k_\perp$ represents an integral in the ring (instead of a 3D shell) $|\boldsymbol{k}_\perp| \in [k_\perp, k_\perp + \Delta k]$ and

$$\mathcal{N} = \frac{A_T}{4\pi\,A\,k_\perp\,q_\perp\,p_\perp\,(\Delta k)^3} \tag{D.43}$$

is a normalisation factor. For this specific estimator, it is possible to show that the variance is given by

$$\mathrm{Var}[\hat{B}(k_\perp,q_\perp,p_\perp)] = (2\pi)^4\,\mathcal{N}\,s_{kqp}\,P(k_\perp)\,P(q_\perp)\,P(p_\perp), \tag{D.44}$$

where s_{kqp} is a counting factor that appears from the Dirac delta functions confining the wavemodes to rings. For equilateral, isosceles and scalene triangles $s_{kqp} = 6, 2$ and 1, respectively.

The final Fisher matrix, Equation (D.41) can be worked out with its symmetry proper-

ties: we expand all possible combinations of tracers, $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{X}', \boldsymbol{Y}'$ and $\boldsymbol{Z}'$, paying particular attention to the derivative operator, and grouping all the equivalent terms. For example, collecting the terms with 3 galaxy entries (TTT) in one of the bispectra, $\mathcal{B}^{XYZ}$ or $\mathcal{B}^{X'Y'Z'}$, and 1 hydrogen (TTa, TaT) in the remaining one, the following term $F^{3T2T}_{\alpha\beta} \sim \partial_\alpha \mathcal{B}^{TTT} \partial_\beta \mathcal{B}^{Ta'T}$, which appears twice in the Fisher matrix expansion, is equivalent to $\partial_\alpha \mathcal{B}^{TTT} \partial_\beta \mathcal{B}^{TTa'}$. Therefore, we have $3F^{3T2T}_{\alpha\beta}$. By symmetry, there will be another three terms in the form $F^{3T2T}_{\beta\alpha}$. We group these two contributions, $3F^{3T2T}_{\alpha\beta}$ and $3F^{3T2T}_{\beta\alpha}$ into a single term $6F^{3T2T}_{(\beta\alpha)}$ for simplicity.

The complete Fisher matrix can be expressed as the sum of 9 non-vanishing contributions (neglecting the entries that do not have one galaxy overdensity field, as the derivatives vanish):

$$F^B_{k_\parallel q_\parallel} = F^{3T3T}_{k_\parallel q_\parallel} + 6F^{3T2T}_{(k_\parallel, q_\parallel)} + 6F^{3T1T}_{(k_\parallel, q_\parallel)} + 3F^{2T2T+}_{k_\parallel q_\parallel} + 6F^{2T2T\times}_{k_\parallel q_\parallel} + 12F^{2T1T+}_{(k_\parallel, q_\parallel)} + 6F^{2T1T\times}_{(k_\parallel, q_\parallel)} + 3F^{1T1T+}_{k_\parallel q_\parallel} + 6F^{1T1T\times}_{k_\parallel q_\parallel}. \tag{D.45}$$

For the sake of simplicity, we do not show all the 9 non-vanishing components, but we focus on one more term ($1T1T$), that is, the ones containing 1 galaxy and 2 hydrogen slices. There are two terms like $F^{1T1T(+)}_{\alpha\beta} \sim \partial_\alpha \mathcal{B}^{Tab} \partial_\beta \mathcal{B}^{Ta'b'}$, which is equivalent to the other possible combination of tracers: $\partial_\alpha \mathcal{B}^{abT} \partial_\beta \mathcal{B}^{a'b'T}$. Hence, they add up, contributing as $3F^{1T1T(+)}_{\alpha\beta}$. Finally, we have twice each one of the following terms: $\partial_\alpha \mathcal{B}^{Tab} \partial_\beta \mathcal{B}^{a'Tb'}$, $\partial_\alpha \mathcal{B}^{Tab} \partial_\beta \mathcal{B}^{a'b'T}$ and $\partial_\alpha \mathcal{B}^{abT} \partial_\beta \mathcal{B}^{Ta'b'}$, which all contribute to yield $6F^{1T1T(\times)}_{\alpha\beta}$.

To lowest order in P_{gh}/P_{hh} and ignoring any information from B^{TTT} or $B^{TTk_\parallel}$, only $3F^{1T1T(+)}_{\alpha\beta}$ will contribute to the Fisher matrix:

$$F^B_{k_\parallel q_\parallel} \simeq \delta^K_{k_\parallel + q_\parallel} \frac{AL}{4\pi} \frac{1}{6} \int \mathrm{d}k_\perp \mathrm{d}q_\perp \mathrm{d}p_\perp \, \frac{k_\perp q_\perp p_\perp}{\pi^2 A_T} \mathcal{I}^{TT}(k_\perp) \int \frac{\mathrm{d}p_\parallel}{2\pi} \frac{B^2_{ghh}(k_\parallel, p_\parallel, -k_\parallel - p_\parallel, \Delta_\perp)}{P_{hh}(p_\parallel, q_\perp) P_{hh}(-k_\parallel - p_\parallel, p_\perp)}. \tag{D.46}$$