Alma Mater Studiorum - Università di Bologna

School of Science
Department of Physics and Astronomy
Master Degree in Astrophysics and Cosmology

Testing General Relativity with Relativistic Redshift-Space Distortions in Clustering Statistics

Graduation Thesis

Submitted by:
Francesca Giusti

Supervisor:
Prof. Federico Marulli

Academic Year [2024/2025]
Graduation Date [I]

Sail!

Contents

Abstract

The large-scale structure (LSS) of the Universe offers a powerful probe of fundamental cosmological information on structure formation, dark matter and dark energy content. Clustering statistics, in particular the 2-point auto-correlation function (2PCF) and the 2-point cross-correlation function, are among the most widely used tools in this field. The observed number density of tracers of the matter field is affected by several perturbative effects, which induce distortions in the redshift. While the dominant contribution arises from cosmological expansion, further corrections, such as peculiar velocities, must be included to avert misinterpretations. Recent studies point towards modelling first-order relativistic effects, including gravitational potential terms, relativistic Doppler and transverse Doppler shifts, gravitational lensing, the integrated Sachs–Wolfe (ISW) effect, as well as light-cone, wide-angle, and evolution terms. These corrections can induce a dipole signal in the cross-correlation between differently biased populations, a signal that is otherwise absent in the auto-correlation. This thesis aims to validate state-of-the art models of the various contributions to the large-scale dipole signal in the cross-correlation function over separations between 20 and 140 h^{-1}Mpc. Disentangling all the different contributions in real survey is non-trivial, requiring both high-precision redshift measurements and very large data sets. This work relies on numerical simulations. We used, in particular, the full-sky light-cone realisation of the RayGalGroupSims numerical simulation suite, which spans a redshift range of $[0, 0.5]$ and covers a volume of $8.34\,(h^{-1}\mathrm{Gpc})^3$. Adopting a fiducial Λ-cold dark matter (ΛCDM) model based on general relativity (GR), we find that dipole is largely dominated by the relativistic Doppler term over all the scales considered. After modelling different contributions, the resulting reduced chi-squared values falled within the range $[0.3, 2.9]$ for all tested configurations, but for potential terms. We finally perform Markov Chain Monte Carlo (MCMC) fits for two cross-correlation cases involving the two most massive populations, constraining the linear growth rate f and bias parameters, and compare the results with those from the auto-correlation. Agreement with model predictions is discussed in light of the obtained constraints.

Sommario

La struttura a grande scala (LSS) dell'Universo fornisce informazioni cosmologiche di fondamentale importanza riguardanti la formazione delle strutture cosmiche e il contenuto di materia ed energia oscure. La statistica di clustering rappresenta uno strumento largamente utilizzato in questo campo: ne sono un esempio le funzioni di auto-correlazione di cross-correlazione a due punti. Il numero dei traccianti del campo di materia (come le galassie e gli aloni di materia oscura) che osserviamo è influenzato da diversi effetti perturbativi, che inducono distorsioni nei redshift. Sebbene il contributo dominante al redshift derivi dall'espansione cosmologica, sono presenti ulteriori effetti distorsivi, come quelli causati dalle velocità peculiari degli oggetti. Studi recenti si sono inoltre focalizzati sugli effetti relativistici al primo ordine, inclusi i termini del potenziale gravitazionale, l'effetto Doppler relativistico, il lensing gravitazionale, l'effetto Sachs–Wolfe integrato (ISW), nonché le distorsioni dovute al cono luce, gli effetti evolutivi e quelli che derivano dal rilassare l'ipotesi di osservatore distante. Tali perturbazioni possono indurre un segnale di dipolo nella cross-correlazione tra popolazioni con bias differenti, un segnale altrimenti assente nell'auto-correlazione. Questa tesi si propone di validare i più recenti modelli per descrivere i diversi contributi al segnale di dipolo su larga scala, fra 20 e 140 h^{-1}Mpc, nella funzione di cross-correlazione. Isolare tutti i contributi coinvolti in survey di galassie reali non è banale e richiede sia misure di redshift ad alta precisione, sia data set molto estesi. Questo lavoro si è basato sull'analisi di simulazioni numeriche. Abbiamo utilizzato, in particolare, la versione full-sky del set di simulazioni numeriche RayGalGroupSims, che copre un intervallo di redshift $[0, 0{,}5]$ e un volume di $8.34\,(h^{-1}\mathrm{Gpc})^3$. Adottando un modello ΛCDM (costante cosmologica e materia oscura fredda), basato su relatività generale (GR), abbiamo riscontrato che il dipolo è largamente dominato dal termine Doppler relativistico su tutte le scale considerate. Dopo aver modellato i differenti termini, i valori ridotti di χ^2 ottenuti rientrano nell'intervallo $[0.3, 2.9]$ per tutte le configurazioni testate. Infine, abbiamo effettuato fit basati su catene di Markov Monte Carlo (MCMC), per due casi di cross-correlazione tra le due popolazioni più massicce, con l'obiettivo di vincolare il tasso di crescita lineare f e i bias delle due popolazioni, confrontando i risultati con quelli ottenuti dall'auto-correlazione. L'accordo con le predizioni dei modelli è stato discusso

alla luce dei risultati ottenuti.

Introduction

The LSS of the Universe is a central focus of contemporary cosmology. The spatial distribution of haloes, galaxies, clusters, and cosmic voids is not random but encodes the imprint of the initial conditions under which cosmic structures formed, namely the primordial scalar fluctuations (i.e. gravitational potential fluctuations in linear theory) that seeded cosmic structure. Analysing this structure offers a powerful tool for probing the nature of dark matter and dark energy, and for testing GR on cosmological scales. A detailed relativistic framework for understanding the LSS and its observational signatures is presented by Tsagas et al. (2008).

A key statistical feature of LSS is the clustering of cosmic structures, reflecting their tendency to aggregate on characteristic scales. This phenomenon is typically quantified through the 2PCF, which measures the excess probability, compared to a random distribution, of finding a pair of objects at a given separation. The clustering signal arises from the interplay between the initial density fluctuations and the subsequent gravitational evolution of matter. As such, the 2PCF encodes valuable information about both the composition and the geometry of the Universe, as well as the laws governing structure formation. For example, clustering properties, particularly the angular 2PCF, have been recently used in analysis of galaxies in the Sloan Digital Sky Survey (SDSS), treated as tracers of the dark matter field, to investigate the evolution of cosmic structure (Franco et al., 2025).

A succession of major observational campaigns has progressively enhanced our understanding of the cosmic web. The Planck mission provided precise measurements of the cosmic microwave background (CMB), delivering tight constraints on key ΛCDM (Λ-Cold Dark Matter) parameters such as the matter density parameter Ω_m, the Hubble constant H_0, and the scalar spectral index n_s (Aghanim et al., 2020). These parameter constraints serve as a benchmark for LSS analyses at lower redshifts.

Among spectroscopic galaxy surveys, WiggleZ Dark Energy Survey was one of the earliest to probe the baryon acoustic oscillation (BAO) signal and redshift-space distortions (RSD) at intermediate redshifts ($z \sim 0.2$–1), providing key evidence for the accelerating expansion of the Universe (Blake et al., 2011). The SDSS – Baryon Oscillation Spectroscopic Survey (BOSS) later improved on these constraints by measuring

BAO and RSD more precisely in the range $z \sim 0.3$–0.7, enabling accurate reconstructions of the expansion history and the growth rate of cosmic structures (Beutler et al., 2014; Alam et al., 2017b). Building on this, eBOSS (extended BOSS) extended spectroscopic observations to higher redshifts ($z \sim 0.6$–2.1), using a combination of emission-line galaxies, luminous red galaxies, and quasars (Bautista et al., 2020). Its final cosmological analysis placed stringent constraints on both the distance-redshift relation and the growth of structure, enabling joint tests of cosmic expansion and gravitational dynamics.

In parallel, the VIMOS Public Extragalactic Redshift Survey targeted galaxies in the redshift range $z \sim 0.5$–1.2, providing high-fidelity measurements of galaxy clustering and RSD over a wide cosmic volume (Mohammad et al., 2018). This helped bridge the observational gap between BOSS and high-redshift probes like eBOSS.

Complementary photometric surveys, such as the Dark Energy Survey (DES), have further contributed by measuring galaxy clustering and weak gravitational lensing over large sky areas (Abbott et al., 2022). These data have yielded independent constraints on cosmological parameters, revealing mild but persistent tensions with Planck's CMB-based predictions, particularly regarding the amplitude of matter clustering on intermediate scales.

Looking ahead, the *Euclid* mission, led by the European Space Agency (ESA), aims to map over 15000 deg^2 of the sky and to obtain spectroscopic redshifts for tens of millions of galaxies up to $z \sim 2$ (Amendola et al., 2018). By combining galaxy clustering and weak lensing, *Euclid* is expected to achieve percent-level precision on both the expansion rate and the growth of structures, thereby providing a stringent testbed for GR and alternative theories.

Among the various objectives of modern cosmology, testing GR on large scales has become increasingly crucial. Despite the remarkable success of the theory in local (Dyson et al., 1920; Pound and Rebka, 1960) and strong-field regimes (Antoniadis et al., 2013), its validity on cosmological scales remains an open question. At such scales, new physical effects could come into play, possibly related to dark energy or to modifications of the metric tensor itself. Moreover, the validity of the equivalence principle, a foundational tenet of GR, has yet to be confirmed beyond the Solar System. As shown in Castello et al. (2024), galaxy clustering statistics offer a novel way to test the equivalence principle via measurable quantities that could reveal additional forces acting on dark matter. However, the clearest signatures of modified gravity theories are often found in the perturbative sector. Growth functions, for instance, may diverge by several percent from those predicted by standard dark energy models, even when the Hubble expansion history is the same. This implies that joint measurements of the Hubble and growth functions are essential for constraining the underlying gravitational physics (Lue et al., 2004). The growth function can be inferred from both weak lensing and galaxy clustering, particularly via RSD.

In the classical Newtonian framework, RSD arise from neglecting peculiar velocities of galaxies, which are superimposed on the Hubble flow velocities. Neglecting these contributions can lead to systematic biases in the inferred LSS. The pioneering work by Kaiser (1987), followed by Hamilton (1992), demonstrated how multipole moments of the 2PCF, notably the quadrupole and hexadecapole, can constrain cosmological parameters such as the matter density parameter. Since then, RSD have become a cornerstone of cosmological analyses.

More recently, it has become evident that a fully relativistic treatment of RSD is necessary (Elkhashab et al., 2025), especially for present and next-generation surveys such as the Euclid Wide Spectroscopic Survey. This relativistic approach opens new opportunities to test GR on cosmological scales. Relativistic corrections include terms such as the Sachs-Wolfe effect, its integrated counterpart, and the Shapiro time delay. Additional terms stem from the relativistic Doppler effect, gravitational redshift, light-cone projection, lensing, source evolution, and wide-angle effects. These contributions, together with their observational signatures in galaxy surveys, have been studied, for instance, by Yoo (2010).

To accurately model and detect these relativistic contributions, it is essential to move from the standard auto-correlation function to the cross-correlation function between distinct galaxy populations. In auto-correlation, only even multipoles appear due to the symmetry of the system. However, when cross-correlating different populations selected, for instance, by luminosity or mass, the symmetry is broken, and odd multipoles emerge. Among these, the dipole is particularly compelling, as it exhibits a marked sensitivity to relativistic effects. This has led to a growing body of work aimed at modelling and measuring the dipole (McDonald, 2009; Bonvin and Durrer, 2011; Gaztanaga et al., 2017; Tansella et al., 2018; Breton et al., 2019), with recent studies investigating its potential as a probe of fundamental physics and its detectability in surveys such as DESI (Dark Energy Spectroscopic Instrument), see Bonvin et al. (2023). In preparation for *Euclid*, the effect of gravitational redshift on the dipole has recently been studied using the Flagship simulation, which reproduces realistic galaxy distributions expected from the mission (Lepori et al., 2025).

Looking ahead, the newly acquired data from *Euclid*, together with forthcoming observations, are expected to mark a significant turning point in this field. As shown by Castorina and di Dio (2022), relativistic effects typically amount to corrections at the sub-10% level compared to the standard RSD term. For this reason, Euclid, with its high-precision redshift measurements and extensive sky coverage, will enable, for the first time, a statistically robust detection of these effects, thereby opening a direct observational window onto the relativistic regime of cosmic structure formation. The coming years may thus prove pivotal in either consolidating or challenging GR as a pillar of our cosmological model and of physics as a whole.

This is precisely why the present work focuses on measuring and modelling the dipole as a testbed for GR. The approach involves both the development of numerical estimators capable of extracting the dipole from halo and galaxy catalogues (which may be real or simulated), and the implementation of theoretical models describing its expected behaviour. The structure unfolds as follows.

- Chapter 1 outlines the theoretical background necessary to acquire a general overview of the fundamental concepts which modern cosmology is based on;
- Chapter 2 serves to the theoretical development of cosmic clustering and its strengths as a cosmological probe, with particular emphasis on the multipole expansion of the cross-correlation function and the treatment of standard RSD;
- Chapter 3 introduces the relativistic contributions to RSD. The theoretical foundations are presented, together with a discussion of their implications for the dipole moment of the cross-correlation function;
- Chapter 4 is divided into two main parts. The first part provides an overview of the RayGalGroupSims simulations, which form the basis for the numerical data used in this thesis. The second part is dedicated to the CosmoBolognaLib, which have been extensively used and further extended to meet the specific requirements of this thesis. This section includes code and data validation, as well as detailed explanations of the new implementations for data analysis and dipole modelling. A brief description is also given of the Bayesian statistical analysis and MCMC methods, used in the post-processing stage;
- Chapter 4 presents the main results along with broader considerations and interpretations;
- Chapter 5 provides the concluding remarks and discusses future prospects and possible applications of the work.

Chapter 1

Theoretical framework

This chapter provides a brief overview of the theoretical framework of modern cosmology, largely following Coles and Lucchin (2002) and Dodelson and Shmidt (2021) and focusing on the principles and concepts most relevant to this thesis. In particular, it aims to establish the necessary background to understand the cosmological context in which LSS formation occurs, the statistical description of clustering, and the origin and interpretation of redshift distortions, in particular for what concerned relativistic effects on the dipole of the 2PCF. While the presentation is not exhaustive, it includes the key elements required to follow the discussions in the main body of the thesis. Readers already familiar with standard cosmology may choose to skip the first two chapters and proceed directly to chapter 3.

1.1 Cosmological principle and FLRW metric

Observations on cosmological scales, far exceeding those of galaxies and galaxy clusters, reveal that the Universe is homogeneous and isotropic. Homogeneity refers to invariance under translation, while isotropy denotes invariance under rotation. Thus, the observer's perspective depends neither on the direction in which they look nor on their location. We elevate these properties to a fundamental principle, known as the cosmological principle, which is assumed to hold not only within the observed universe but throughout the Universe as a whole. To proceed, we need to introduce a metric, beginning with the framework of GR and the 4-dimensional spacetime continuum. It is crucial to highlight that time is not absolute either in special or in GR as in Newtonian mechanics: it depends on the relative motion between different reference frames and on the observer's location within the curvature of spacetime. The infinitesimal interval between two points is expressed as:

$$\mathrm{d}s^2 = g_{\alpha\beta}\mathrm{d}x^\alpha \mathrm{d}x^\beta, \tag{1.1}$$

where the indices $\alpha, \beta = 0, 1, 2, 3$ correspond to time and spatial coordinates, with the Einstein summation convention for repeated indices. If we decompose the terms into time, spatial, and mixed components, we obtain the following:

$$\mathrm{d}s^2 = g_{00}\mathrm{d}t^2 + 2g_{0i}\mathrm{d}t\mathrm{d}x^i + g_{ij}\mathrm{d}x^i\mathrm{d}x^j. \tag{1.2}$$

In order to derive an expression for the 3-dimensional metric, we first consider the 2-dimensional case. The spatial separation between two points in polar coordinates is given by:

$$\mathrm{d}l^2 = \mathrm{d}\rho^2 + \rho^2\mathrm{d}\phi^2, \tag{1.3}$$

where ρ is the radial coordinate, with $0 \leq \rho < \infty$, and ϕ is the angular coordinate, with $0 \leq \phi < 2\pi$. It is convenient to express the radial coordinate as a function of the scale factor $a(t)$, so that $\rho = a(t)\, r$, where r is a dimensionless coordinate. Substituting the angular coordinate ϕ with the solid angle element $\mathrm{d}\Omega = \mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2$, we adapt this approach for the 3-dimensional case, resulting in the spatial distance for the flat geometry to be:

$$\mathrm{d}l^2 = a^2(t)(\mathrm{d}r^2 + r^2\mathrm{d}\Omega^2). \tag{1.4}$$

Applying analogous reasoning to spherical and hyperbolic geometry, we derive the following general expression:

$$\mathrm{d}s^2 = c^2\mathrm{d}t^2 - a^2(t)\left[\frac{\mathrm{d}r^2}{1-kr^2} + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right)\right]. \tag{1.5}$$

Thus, spatial geometry is governed by the parameter k, which takes one of the following values:

- $k = -1$: Hyperbolic geometry, corresponding to an open universe.
- $k = 0$: Euclidean geometry, corresponding to a flat universe.
- $k = 1$: Spherical geometry, corresponding to a closed universe.

The Eq. (1.5) is known as the Friedmann-Lemaître-Robertson-Walker (FLRW) metric.

1.2 Friedmann equations

With the metric formulation in place, we now turn to GR, which is intended to provide a broad overview of the fundamental concepts underpinning the analysis conducted in this thesis. The primary aim now is to derive the crucial equations of cosmology, namely

the Friedmann equations. The cornerstone equation of GR is the Einstein field equation, expressed as

$$R_{\mu\nu} - \frac{1}{2} R\, g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}, \tag{1.6}$$

where $R_{\mu\nu}$ denotes the Ricci tensor, R the Ricci scalar, and $T_{\mu\nu}$ the energy-momentum tensor, with $\mu, \nu = 0, 1, 2, 3$. In this analysis, we adopt the assumption that the energy-momentum tensor can be effectively represented by that of an ideal fluid. Consequently, we neglect viscosity and thermal conduction, as the energy-momentum tensor is then reduced to a function dependence solely on pressure and density. In particular, we express the energy-momentum tensor as

$$T_{\mu\nu} = -p\, g_{\mu\nu} + \left(p + \rho\, c^2\right) \mathbf{u}_\mu \mathbf{u}_\nu, \tag{1.7}$$

$\mathbf{u}_\mu$, $\mathbf{u}_\nu$ being the 4-velocities. This formulation represents a system of 10 equations, since the metric tensor has to be symmetric. Due to the cosmological principle, however, only two independent equations remain, famously referred to as the Friedmann equations (Friedmann, 1922):

$$\dot{a}^2 + K c^2 = \frac{8\pi G}{3} \rho\, a^2 \tag{1.8}$$

$$\ddot{a} = -\frac{4\pi G}{3} \left(\rho + \frac{3\, p}{c^2}\right) a \tag{1.9}$$

where the dots represent derivatives with respect to cosmological time. We can manipulate Eq. (1.8) to obtain the following expression:

$$\frac{K\, c^2}{a^2} = \left(\frac{\dot{a}}{a}\right)^2 \left(\frac{\rho}{\rho_c} - 1\right), \tag{1.10}$$

where the critical density ρ_c, is defined as

$$\rho_c = \frac{3}{8\pi G} \left(\frac{\dot{a}}{a}\right)^2. \tag{1.11}$$

Next, we introduce the density parameter, defined as

$$\Omega = \frac{\rho}{\rho_c}, \tag{1.12}$$

so that the geometry of space is open, flat, or closed depending on whether Ω is less than, equal to, or greater than unity. Figure 1.1 illustrates the evolution of the cosmic scale factor for the three possible spatial curvatures, under different combinations of cosmological fluid components. Together with the Friedmann equations, it is essential

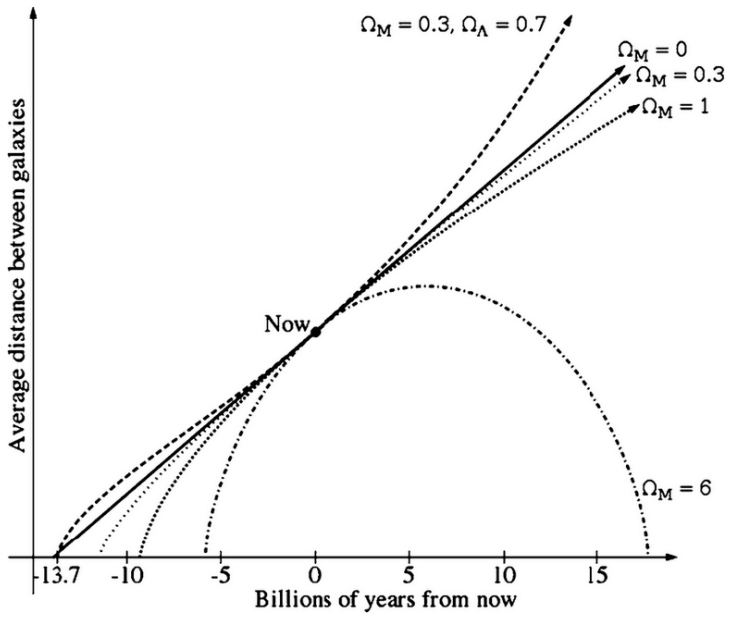

Figure 1.1: Time evolution of the cosmic scale factor for different density parameter values (Casado, 2020).

to account for the adiabatic condition, as the universe constitutes a closed system. This condition can be expressed as follows:

$$\mathrm{d}(\rho c^2\, a^3) = -p\, \mathrm{d}a^3, \tag{1.13}$$

where ρc^2 represents the energy density per unit volume.

It is often convenient to introduce the conformal time τ, defined as

$$\eta = \int \frac{\mathrm{d}t}{a(t)}, \tag{1.14}$$

which enables the FLRW metric to be written as

$$\mathrm{d}s^2 = a(\eta)^2 \left[(c\, \mathrm{d}\eta)^2 - \left(\frac{\mathrm{d}r^2}{1 - k\, r^2} + r^2 \mathrm{d}\Omega^2 \right) \right]. \tag{1.15}$$

We now examine the behaviour of the Friedmann equations in order to find a static solution, that is, one where $\dot{a} = \ddot{a} = 0$. As we shall show, this necessitates the introduction of an additional term. From Eq. (1.9), we observe that imposing $\ddot{a} = 0$ implies a relationship between pressure and density that causes the term in parentheses to vanish. Thus:

$$\rho = -\frac{3p}{c^2}. \tag{1.16}$$

Given the requirement for pressure and density to be positive quantities, Einstein introduced a new constant Λ in the field equation to address this concern. This constant was later shown to be incorrect in the context in which Einstein conceived it, yet it is still employed today, albeit with a different interpretation. Defining a new term for the energy-momentum tensor such as

$$\tilde{T}_{ij} := T_{ij} + \frac{\Lambda c^4}{8\pi G}\, g_{ij}, \tag{1.17}$$

allows us to preserve the form of the Einstein field equation by replacing T_{ij} with $\tilde{T}_{ij}$, while maintaining the structure of the energy-momentum tensor. This is achieved by substituting pressure and density with the modified expressions:

$$\tilde{p} = p + p_\Lambda = p - \frac{\Lambda c^4}{8\pi G}, \tag{1.18}$$

and

$$\tilde{\rho} = \rho + \rho_\Lambda = \rho + \frac{\Lambda c^2}{8\pi G}. \tag{1.19}$$

With these modifications, we can also rewrite the two Friedmann equations, thus preserving their formal structure.

1.3 The Hubble-Lemaître law

Starting with the FLRW metric and setting $\mathrm{d}t = 0$, we define the proper distance d_P between two points as:

$$d_P := \int_0^r \frac{a\,\mathrm{d}r'}{(1 - kr'^2)^{1/2}} = a(t) f(r). \tag{1.20}$$

The function $f(r)$ depends on the value of k, and takes the following form in each case:

- $f(r) = r$, for $k = 0$;
- $f(r) = \sin^{-1}(r)$, for $k = 1$;
- $f(r) = \sinh^{-1}(r)$, for $k = -1$.

Next, we define the comoving distance as the proper distance at the present time t_0:

$$d_C := d_P(t_0) = a_0\, f(r) = \frac{a_0}{a(t)} d_P(t), \tag{1.21}$$

where $a_0 \equiv a(t_0)$.

Taking the time derivative of the proper distance, we obtain the Hubble-Lemaître law (Lemaître, 1927; Hubble, 1929)

$$\frac{\mathrm{d}}{\mathrm{d}t} d_P = \dot{a}(t) f(r) = \frac{\dot{a}(t)}{a(t)} d_P, \tag{1.22}$$

where, in the final step, we multiply and divide by the scale factor $a(t)$. We define the Hubble parameter $H(t)$ such as

$$H(t) := \frac{\dot{a}(t)}{a(t)}, \tag{1.23}$$

and the Hubble parameter computed at the present time t_0 is

$$H_0 = \left.\frac{\dot{a}(t)}{a(t)}\right|_{t=t_0}. \tag{1.24}$$

The Hubble parameter at the present time indicates the isotropic expansion rate of the Universe, and it is a constant throughout space for a fixed time. Current estimates in the literature hover near $67\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ (Kozmanyan et al., 2019; Sneppen et al., 2023). Nevertheless, it is essential to highlight the so-called "H_0 tension", which arises from the two distinct methods used to estimate the Hubble constant. Specifically, measurements from the local universe (using a distance ladder method) yield a higher value of H_0 compared to those based on the Cosmic Microwave Background (CMB). In fact, there is a 4σ to 6σ deviation between these two estimated values of H_0 and this may suggest the need for new physics beyond the ΛCDM model (Hu and Wang, 2023).

We can also parameterise H_0 in unity of $100\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, as follows:

$$H_0 = h\,100\,\mathrm{km\,s^{-1}\,Mpc^{-1}}. \tag{1.25}$$

We call h the reduced Hubble parameter: it is useful to take into account the uncertainty in H_0.

1.4 Redshift

Sources (such as galaxies or clusters) in our Universe follow the expansion of the universe itself. Consider a source located at a comoving coordinate r. The redshift z is a directly observable variable, related to the scale factor a, and can be defined as

$$z := \frac{\lambda_o - \lambda_e}{\lambda_e}, \tag{1.26}$$

where λ_o is the wavelength observed at point O (the location of the observer) at a time t_0, and λ_e is the wavelength emitted by the source at an earlier time t_e. Consider a source emitting a first pulse at a time t_e, with a wavelength λ_e that reaches the observer at a time t_0, with a wavelength λ_0, and a second pulse at a time $t = t_e + \delta t_e$ that reaches the observer at a time $t' = t_0 + \delta t_0$. If δt_e and δt_0 are small, since light rays travel across null geodesic (i.e. $ds^2 = 0$), we can integrate the FLRW metric over a light-like interval to obtain

$$\frac{a_0}{a} = \frac{\lambda_o}{\lambda_e}, \tag{1.27}$$

where we choose a coordinate system in which light rays travel along a path with $d\theta = d\phi = 0$. Thus, we can directly link the redshift to the scale factor $a(t)$:

$$z + 1 = \frac{a_0}{a(t)}. \tag{1.28}$$

A varying scale factor between the present time and the time at which the light ray was emitted causes a shift in the wavelength. If $a_0 > a(t)$, then $z > 0$, which results in a redshift of the wavelength. In contrast, if $a_0 < a(t)$, we observe a blue-shift.

1.5 Cosmological distances

Depending on the specific requirements, different definitions of distance may be adopted in cosmology. The proper distance, while formally defined, has limited practical utility, owing to intrinsic challenges in directly measuring the separations between objects. Another distance, already introduced, is the comoving distance (i.e. the proper distance computed at the present time), which we redefine here in a more general form as:

$$\chi(z) = \int_0^z \frac{c\,\mathrm{d}z'}{H(z')} \tag{1.29}$$

We now proceed to define the luminosity distance, derived from the conservation of flux:

$$l = \frac{L}{4\pi d_L^2}. \tag{1.30}$$

We find

$$d_L = a_0\, r(1 + z), \tag{1.31}$$

where r is a dimensionless polar coordinate. This distance became particularly useful with the discovery of the standard candles (objects whose intrinsic luminosity is known,

such as Type Ia supernovae) and, in principle, can be measured directly by using this method. The final distance we define is the angular distance, denoted as:

$$d_A = \frac{\mathrm{d}s}{\mathrm{d}\theta} = a(t)r, \tag{1.32}$$

where we employ the FLRW metric, setting $\mathrm{d}\phi = 0$, and $\mathrm{d}t = 0$ (i.e. we know the proper length of the object). This leads to the following result:

$$\mathrm{d}s^2 = a(t)^2 r^2 \mathrm{d}\theta^2. \tag{1.33}$$

The relation between these two distances is governed by the redshift and is given by:

$$\frac{d_A}{d_L} = \frac{a(t)\, r}{a_0\, r\, (1+z)} = \frac{1}{(1+z)^2}. \tag{1.34}$$

This is referred to as the duality relation: any deviation from this relation would imply an inconsistency of the metric, entailing a reassessment of the assumptions of homogeneity and isotropy.

1.6 The deceleration parameter

Let us expand the scale factor in a Taylor series around t_0, up to the second order:

$$a(t) = a_0 + \left.\frac{\mathrm{d}a}{\mathrm{d}t}\right|_{t_0} (t - t_0) + \frac{1}{2} \left.\frac{\mathrm{d}^2 a}{\mathrm{d}t^2}\right|_{t_0} (t - t_0)^2 + O(t - t_0)^3, \tag{1.35}$$

with $a_0 \equiv a(t_0)$. This can be rewritten as

$$a(t) = a_0 \left[1 + H_0(t - t_0) - \frac{1}{2}\, q_0\, H_0^2 (t - t_0)^2\right], \tag{1.36}$$

by introducing the deceleration parameter

$$q := -\frac{\ddot{a}(t)\, a(t)}{\dot{a}^2(t)}. \tag{1.37}$$

Consequently, the deceleration parameter at the present time is

$$q_0 := -\frac{\ddot{a}_0\, a_0}{\dot{a}_0^2}. \tag{1.38}$$

The deceleration parameter is a dimensionless quantity by construction, and serves as a measure of the variation in the expansion rate of our Universe. Since the first term of the expansion dominates on small scales, the measurement of q_0 is best achieved at cosmological distances, where the second term becomes predominant. The deceleration parameter at the present time has been found to be negative, indicating that the expansion of the Universe is accelerating.

1.7 Friedmann models

In this section, we shall examine a class of homogeneous and isotropic cosmological models, governed by the Friedmann equations, which provide a relatively simple description of an evolving universe. Using the value of the critical density at the present time, Eq. (1.8) can be reformulated as follows:

$$H_0^2 \left(1 - \frac{\rho_0}{\rho_{0,c}}\right) = -\frac{kc^2}{a_0^2}. \tag{1.39}$$

We introduce the general equation of state for a cosmological fluid:

$$p = w\rho c^2, \tag{1.40}$$

where $w = 0$ and $w = \dfrac{1}{3}$ correspond to the equation of state for the non-relativistic matter and the radiative components, respectively.

In conventional physics w is constrained within the range $0 \leq w < 1$, a domain referred to as "the Zel'dovich interval". Notably, the value of w associated with the cosmological constant lies beyond this range. Substituting the pressure, as derived from the general equation of state, into the adiabatic condition, Eq. (1.13), subsequently expanding the differential, and separating variables, we obtain:

$$(1+w)\frac{\mathrm{d}a^3}{a^3} = -\frac{\mathrm{d}\rho}{\rho} \longrightarrow \rho \propto a^{-3(1+w)}. \tag{1.41}$$

As evident from this equation, the matter and radiative components evolve differently with respect to the scale factor $a(t)$. Regarding the cosmological constant, characterised by an equation of state parameter $w = -1$, it is noteworthy that its energy density remains unchanged over time. In the present epoch, the cosmological constant constitutes the dominant component, although the contribution from matter cannot be disregarded. The significant role played by Λ drives the universe to accelerate expansion: this can be easily inferred from the second Friedmann equation, if incorporating all the contributions from both matter and Λ. Including both terms also into the Eq. (1.8) we obtain:

$$\dot{a}^2 + kc^2 = \frac{8\pi}{3} G \left(\rho_M + \rho_\Lambda\right) a^2. \tag{1.42}$$

Dividing by a_0^2, assuming $a_0 = 1$ and explicitly expressing ρ_Λ, we derive

$$\dot{a}^2 - \frac{8\pi}{3} G\rho_M a^2 - \frac{\Lambda c^2}{3} a^2 = -kc^2, \tag{1.43}$$

which, with the appropriate substitutions, takes the form

$$H_0^2 \left(1 - \Omega_\mathrm{M} - \Omega_\Lambda\right) = -kc^2. \tag{1.44}$$

Throughout this analysis, we have omitted the radiative component, as it is negligible in the present epoch, given that $\Omega_{0,R} \sim 10^{-5}$. By defining $\Omega_{0,tot} = \Omega_{0,M} + \Omega_{0,\Lambda}$, it becomes evident that the necessary condition for the Universe to be flat is $\Omega_{0,M} + \Omega_{0,\Lambda} = 1$. Starting again from the first Friedmann equation, we now derive a general expression for the Hubble parameter. Analyzing the evolution of the density, normalized to its present-day value, taking into account the definition of the critical density, and then generalising the equation to include all relevant cosmological fluid components, we can express H as follows:

$$H^2(z) = H_0^2(1+z)^2 \left\{ \left(1 - \sum_i \Omega_{0,i} \right) + \sum_i \left[\Omega_{0,i} \left(1+z\right)^{1+3w_i} \right] \right\}, \tag{1.45}$$

where the subscript i denotes the i-th component.

1.8 Single-component Universe

The simplest model of the Universe is that of a Universe filled with a single-component cosmological fluid. As shown in Eq. (1.45) the evolution of several key quantities can be determined by substituting different values of w, corresponding to different components. Moreover, we can distinguish between two possible spatial geometries of the Universe: flat and curved.

1.8.1 Flat Universe

Since we are assuming a flat universe, we set $\Omega_{0,w} = 1$, causing the term inside the first parenthesis to vanish, and thus leading to the following relation:

$$\dot{a} = \frac{\mathrm{d}a}{\mathrm{d}t} \propto a^{-\frac{1+3w}{2}} \tag{1.46}$$

Separating the variables and solving for the scale factor, we obtain

$$a(t) = a_0 \left(\frac{t}{t_0} \right)^{\frac{2}{3(1+w)}} = \begin{cases} a \propto t^{1/2} & \text{for } w = 1/3 \\ a \propto t^{2/3} & \text{for } w = 0 \end{cases} \tag{1.47}$$

It is interesting to notice that, with w constrained within the Zel'dovich interval, the condition $t \to 0$ implies $a \to 0$, which mandates the existence of a singularity. Conversely, as $t \to \infty$, the scale factor shows the same asymptotic behaviour, leading to an infinite expansion of the universe. Another fundamental quantity to derive is the Hubble parameter, given by

$$H = \frac{\dot{a}}{a} = \frac{2}{3(1+w)t}, \tag{1.48}$$

alongside the deceleration parameter

$$q = \frac{1+3w}{2}, \tag{1.49}$$

and the density evolution

$$\rho = \frac{1}{6\pi G(1+w)^2\, t^2}. \tag{1.50}$$

In particular, the model with $w = 0$ is the so-called "Einstein-de Sitter Universe" (Einstein and de Sitter, 1932).

At first glance, defining these quantities might seem unnecessary, given that observational evidence confirms our Universe does not conform to a single-component model. However, the significance of such a model lies in the fact that, for most of its life, every universe, whether curved or flat, exhibits behaviour similar to that of such a kind of universe.

1.8.2 Curved Universe

In the case the universe is either closed or open, we again start our analysis with the first Friedmann equation, Eq. (1.45), divided by $\left(\frac{a_0}{a}\right)^2$. We recall that we are considering Friedmann models, which describe universes with a single-component fluid. The equation then takes the form:

$$\left(\frac{\dot{a}}{a_0}\right)^2 = H_0^2 \left[(1-\Omega_0) + \Omega_0 \left(\frac{a_0}{a}\right)^{1+3w}\right]. \tag{1.51}$$

Our focus is on comparing the two terms inside the square brackets, and we take the absolute value of the first term to account for both open and closed universes. By writing

$$|1-\Omega_0| \ll \Omega_0 \left(\frac{a_0}{a}\right)^{1+3w}, \tag{1.52}$$

we obtain

$$\frac{a_0}{a} = 1+z \gg \left|\Omega_0^{-1} - 1\right|^{\frac{1}{1+3w}} \equiv \frac{a_0}{a^*} = 1+z^*. \tag{1.53}$$

This relation indicates that, for the first term in the square brackets to be negligible, the following condition must hold: $a \ll a^*$, or equivalently $z \gg z^*$. In other words, if we are sufficiently close to the Big Bang (the precise proximity to be determined) we

can safely neglect the curvature term $(1 - \Omega_0)$. As expected, this leads to an equation analogous to the one previously derived for a single-component universe:

$$\frac{\dot{a}}{a_0} = H_0 \Omega_0^{\frac{1}{2}} \left(\frac{a_0}{a}\right)^{\frac{1+3w}{2}}, \tag{1.54}$$

with the only distinction being the additional factor of $\Omega_0^{\frac{1}{2}}$. This illustrates why, as mentioned earlier, even curved universes exhibit behaviour similar to flat ones, at least during the primordial phases. Having now reached a strong conviction that $\Omega_0 \sim 1$, substituting even extreme values (such as 10^{-1} or 10) for it in the equation for a^*, and assuming $w = 0$, would lead to the fact that the curvature term only becomes non-negligible in the relatively recent past (for $z \sim 10$ or less). Conversely when considering the case in which $a > a^*$, or equivalently $z < z^*$, the first term within the square brackets becomes negligible. This allows us to investigate the asymptotic behaviour in regimes approaching to the present epoch or even extending into the future. Under these conditions, the first Friedmann equation becomes:

$$\left(\frac{\dot{a}}{a_0}\right)^2 = H_0^2(1 - \Omega_0). \tag{1.55}$$

At this stage, we distinguish between two different cases: one in which $\Omega_0 > 1$, corresponding to a closed universe, and another in which $\Omega_0 < 1$, describing an open universe.

- $\Omega_0 < 1$. By separating variables in Eq. (1.55) and solving for the scale factor $a(t)$, we obtain $a \propto t$. This implies that the ultimate fate of open universes is an infinite linear expansion over time.

- $\Omega_0 > 1$. On the other hand, for curvature parameter values exceeding unity, there exists a critical moment at which the expression within the square brackets of Eq. (1.51) vanishes. At this point, cosmic expansion ceases, and the scale factor begins to decrease. In this scenario, the universe ultimately undergoes a re-collapse into a singularity, an event referred to as the "Big Crunch".

The final aspect we aim to examine is the evolution of the density parameter. Starting from the first Friedmann equation, expressed in terms of H^2 as a function of the redshift,

$$H^2(z) = H_0^2(1 + z)^2 \left[(1 - \Omega_0) + \Omega_0(1 + z)^{1+3w}\right], \tag{1.56}$$

and recalling the definition of Ω as the ratio between the density and a critical threshold, alongside the evolution of the density,

$$\rho = \rho_0 \left(\frac{a_0}{a}\right)^{3(1+3w)}, \tag{1.57}$$

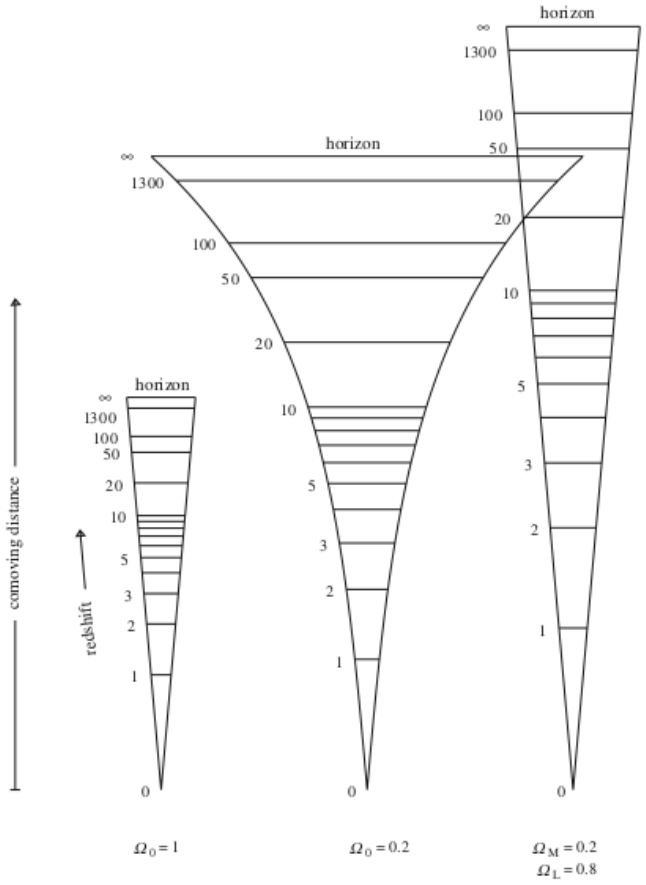

Figure 1.2: Relation between distances and redshifts for three different cosmological models: flat (left) and open (middle) universe with a single matter component, and flat universe with a cosmological constant (right). The cones originate from the observer's location. (Hamilton, 1997).

we derive the evolution of the density parameter $\Omega\,(z)$, given by

$$\Omega^{-1}(z) - 1 = \frac{\Omega_0^{-1} - 1}{(1+z)^{1+3w}}. \tag{1.58}$$

Several implications follow from this expression. Firstly, in the asymptotic limit $z \to \infty$, we recover $\Omega \to 1$, as anticipated. This confirms that, near the Big Bang, all universes exhibit behaviour analogous to the Einstein-de Sitter universe. Moreover, since the denominator of the second term remains strictly positive for any redshift and any physical admissible value of w (which is constrained within the Zel'dovich interval), we observe the following: for $\Omega_0 > 1$, or equivalently, $\Omega_0^{-1} < 1$, it follows that $\Omega^{-1}(z) < 1$ and thus $\Omega(z) > 1$. In other words, if the universe was initially closed, it will remain closed throughout its entire evolution. The same reasoning applies to the case of an

open universe, reinforcing the conclusion that the curvature sign is preserved over cosmic time.

1.9 ΛCDM model

The ΛCDM model (where Λ is the cosmological constant and CDM stands for Cold Dark Matter) currently represents the standard and most widely accepted framework in cosmology, as it accounts for a wide range of observational data. As these findings reveal, our Universe, at the present epoch, cannot be characterised by a single-component fluid, since there are two major constituents we have identified. The first is matter, with $\Omega_M \sim 0.3$, and the second is the cosmological constant Λ, which can be interpreted as dark energy, for which $\Omega_\Lambda \sim 0.7$. Matter is generally understood to consist of both baryonic and dark components, with the latter predominantly composed of CDM, although its precise nature remains a subject of ongoing debate and is yet to be definitively determined (Feng, 2010). CDM is estimated to constitute the majority of the matter component, with $\Omega_{DM} \sim 0.26$. It is crucial to consider also the influence of the cosmological constant, which is supported by several observational data, such as the study of supernovae (Riess et al., 1998; Adil et al., 2024) and the LSS of the universe (Barrow and P., 1993; Hameeda et al., 2022).

The relationship between distances and the cosmic scale factor is illustrated in Figure 1.2. On the right-hand side, the flat case, including a cosmological constant, is shown. This model reflects our current understanding of the Universe, and comparing it with the other two cases in the figure clearly reveals the significant impact of the cosmological constant on cosmic evolution.

1.10 Thermal history of the Universe

In this study the ΛCDM cosmological model is adopted: this is embedded in the framework of the "hot Big Bang" model, wherein the universe originates from a singularity with infinite density. The concept of an infinite density at the very beginning follows from a mathematical extrapolation of the equations of the model, and should not be regarded as a physical state, as the current understanding of physics breaks down at this point. Although this model does not describe the physical laws governing the Big Bang itself, it remains the most widely accepted paradigm because of its consistency with contemporary observations.

Our knowledge begins at 10^{-43} s after the Big Bang. At this epoch, temperatures are extremely high, on the order of 10^{32} K, and the entropy is equal to unity. On timescales shorter than or comparable to the Planck time, both gravitational and quantum effects

are present. From this moment until approximately 10^{-33} seconds after the Big Bang the strong, weak and electromagnetic interactions might be unified, as described by Grand Unified Theories (GUTs).

At around $t \sim 10^{-33}$ s, the energy falls below 10^{15} GeV and the inflationary era occurs. Without delving too deeply into the details of inflationary models, as this lies beyond the scope of the present thesis, we highlight only the key concepts and the reasons why they are essential to complement the standard cosmological model.

Inflation is a phase of accelerated expansion in the early Universe, characterised by a well-defined onset and end. Here, we consider slow-roll models involving a matter scalar field. These models typically rely on the potential energy of a scalar field, ϕ, which offers an easy and natural explanation for such an expansion. In slow-roll inflation, the scalar field evolves slowly over time, resulting in a nearly constant Hubble parameter throughout the inflationary period. It is worth noting that no known scalar field is, in fact, a good candidate for such models. However, the necessity of an inflationary phase is widely acknowledged, as it addresses several shortcomings of the standard Big Bang model:

- Horizon problem: inflation allows for the justification of the extreme isotropy (such as the temperature of the CMB, section 1.11) between regions that, without inflation, would never have entered into causal connection in the past.

- Flatness problem: if the Universe had been born with a value of $\Omega_{\rm tot}$ even slightly different from unity, we would now observe values significantly deviating from one. The inflationary epoch solves this issue by driving the total density parameter extremely close to unity, to within one part in 10^{-60}.

- Magnetic monopoles problem: although we have not explored GUTs in detail, these theories predict the existence of magnetic monopoles, which, however, have never been observed. Inflationary models, and, thus, the era of accelerated expansion, predict such a dramatic dilution of magnetic monopoles, that they are rendered virtually absent in the Universe today.

- Primordial fluctuations: the standard Big Bang model does not account for the origin of the primordial fluctuation seeds. Inflation, on the other hand, naturally predicts their generation. Moreover, during this phase, it becomes possible for particles to survive, since a baryon-antibaryon asymmetry emerges.

Eventually, after the inflationary era, also the electroweak and the strong forces decouple.

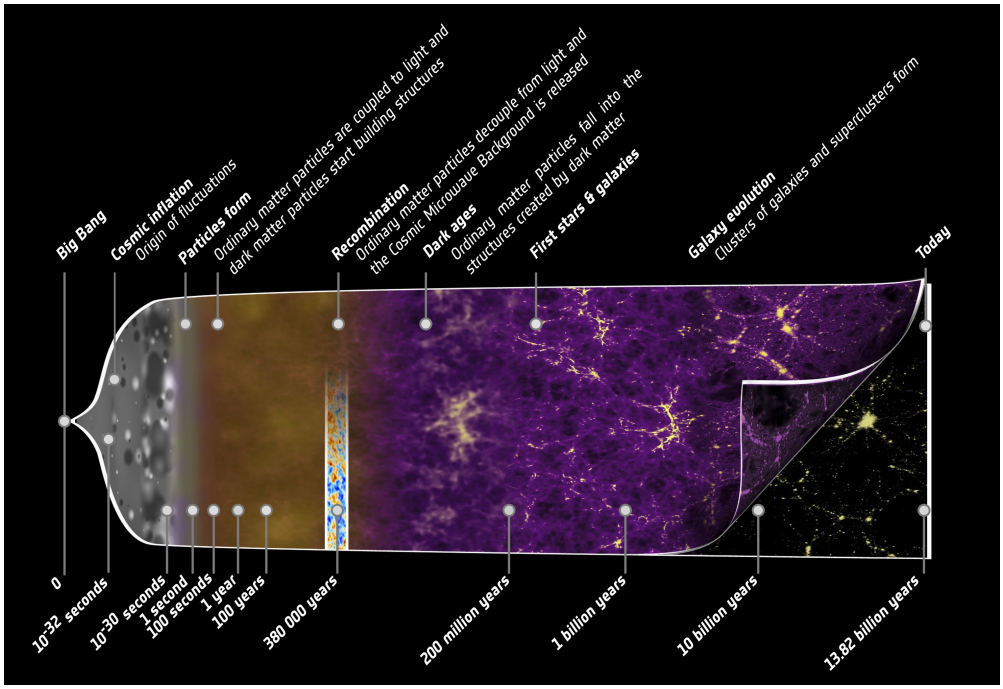

Figure 1.3: The large-scale history of the Universe, from the Big Bang to the present day. This illustration highlights several key epochs in cosmic evolution, including the inflationary period and the release of the CMB, tracing the transition from the dark ages to the emergence of cosmic structures that eventually reionised the Universe. https://www.esa.int (Planck Collaboration).

1.10.1 Cosmological eras

We now provide a brief summary of the different cosmological eras, each characterised by the contributions of various particle species. As the temperature decreases, particles decouple from radiation or undergo annihilation, depending on their mass. This process leads to the loss of their contribution to the energy budget of the cosmic fluid. In fact, in the early Universe, the energy contributions from decoupled particles, or from those still coupled but no longer relativistic, become negligible, and can be safely disregarded in the overall energetics of the system.

- Quark epoch: At approximately 10^{-12} seconds after the Big Bang, the weak and electromagnetic interactions separate. However, the temperature remains too high for quarks to form hadrons, and the cosmic fluid exists as a quark-gluon plasma.

- Hadron epoch: This epoch begins when the energy drops to $\sim$ 150 MeV (corresponding to $t \sim 10^{-5}$ s after the Big Bang) and lasts until around 1 second post-Big Bang. During this period, quarks begin to bind into hadrons, marking

the onset of nucleon formation.

- Leptonic epoch: this era begins with the annihilation of pions and continues until the final particles annihilation (i.e. electrons-positrons annihilation), which occurs 10 s after the Big Bang. During this epoch, muons annihilate, neutrinos decouple from radiation, and electron – positron pairs undergo annihilation, shaping the subsequent evolution of the Universe.

- Radiative epoch: following the annihilation of electron-positron pairs, the radiative era begins. This period is characterised by the initial formation of deuterium, after which other light nuclei start to form. For instance, helium forms approximately $200 - 300$ s after the Big Bang. The formation of light nuclei during this epoch is well described by the theory of primordial nucleosynthesis, which hinges on the initial ratio of neutrons to protons, with its evolution driven by temperature. This represents one of the key successes of the Big Bang theory.

- Recombination era: this is the moment in which hydrogen recombines, meaning electrons start to bind to nuclei, allowing photons to propagate freely, without being continuosly absorbed and re-emitted. The plasma transitions from being optically thick to transparent. The cosmic microwave background (CMB, section 1.11) photons detected today were last scattered during this epoch.

- Equivalence: at a redshift $z \sim 10^3$ the density of matter equals the density of radiation. From this moment on the matter component prevails, and first cosmic structures emerge.

A summary of the cosmic history of the Universe over the past 13.8 billion years is shown in Figure 1.3.

1.11 Cosmic Microwave Background

As mentioned in the previous section, the CMB originates at the time of last scattering (at $z \sim 1100$), after wich photons can propagate freely throughout the universe. We now turn our attention to a statistical analysis of the CMB, starting from the temperature map of photons.

As depicted in Figure 1.4, the CMB temperature is not completely homogeneous but exhibits subtle fluctuations around a mean value of 2.72548 ± 0.00057 K (Fixsen, 2009). These variations are the lingering imprints of primordial density perturbations, the tiny seeds that gave rise to the cosmic structures we observe today. As previously outlined, their origin can be traced back to quantum fluctuations arose during the inflationary epoch.

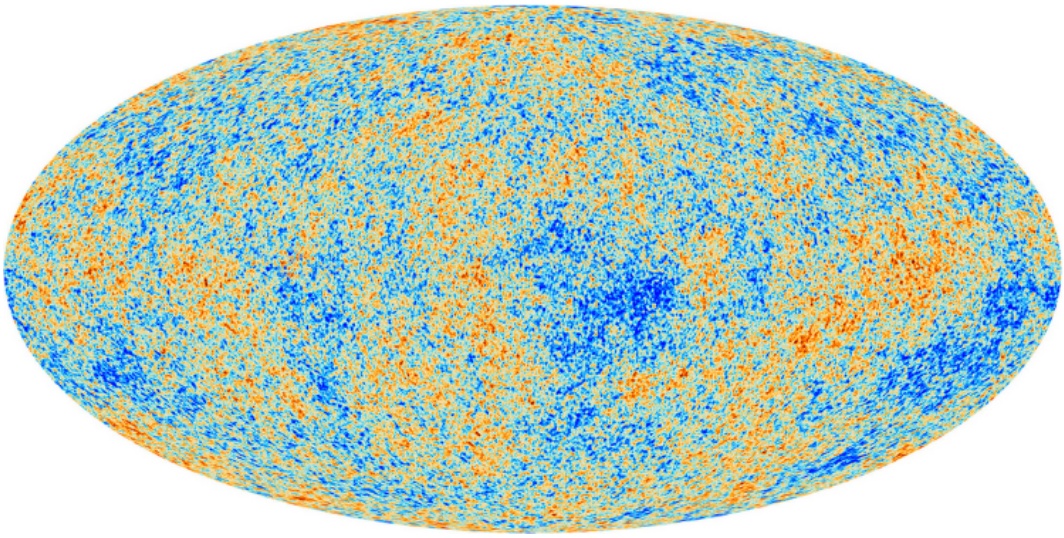

Figure 1.4: Temperature map of the CMB around its mean value of ~ 2.726 *K. Regions shaded in light blue and deep blue indicate temperatures below the average, while areas rendered in orange and red correspond to regions with temperatures above the mean (Ade et al., 2014).*

Given a mean temperature $\overline{T}$, one can characterise these angular temperature deviations using the expression:

$$\frac{\delta T}{\overline{T}}(\theta,\phi) = \frac{T(\theta,\phi)-\overline{T}}{\overline{T}} = \sum_{l=0}^{\infty}\sum_{m=-l}^{+l} a_{l,m}Y_{l,m}(\theta,\phi) \tag{1.59}$$

where, in the second step, we have expanded the temperature field in terms of spherical harmonics. Here, θ and ϕ denote angular coordinates on the celestial sphere, $a_{l,m}$ are the coefficients encoding the amplitude of fluctuations at given values of l, m, and $Y_{l,m}$ are the spherical harmonics, which are themselves defined in terms of Legendre polynomials:

$$Y_{l,m}(\theta,\phi)\sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}\,\mathcal{P}_l^m(\cos\theta)e^{i\phi}. \tag{1.60}$$

We define the angular power spectrum $\mathcal{C}_l$ as

$$\mathcal{C}_l = \langle |a_{l,m}|^2 \rangle = \frac{1}{2l+1}\sum_{m=-l}^{+l} |a_{l,m}|^2, \tag{1.61}$$

which quantifies the contribution of each multipole l to the reconstruction of the temperature fluctuation field. As mentioned above, the CMB exhibits temperature anisotropies which are not only intrinsic (i.e. reflecting the physical conditions at the

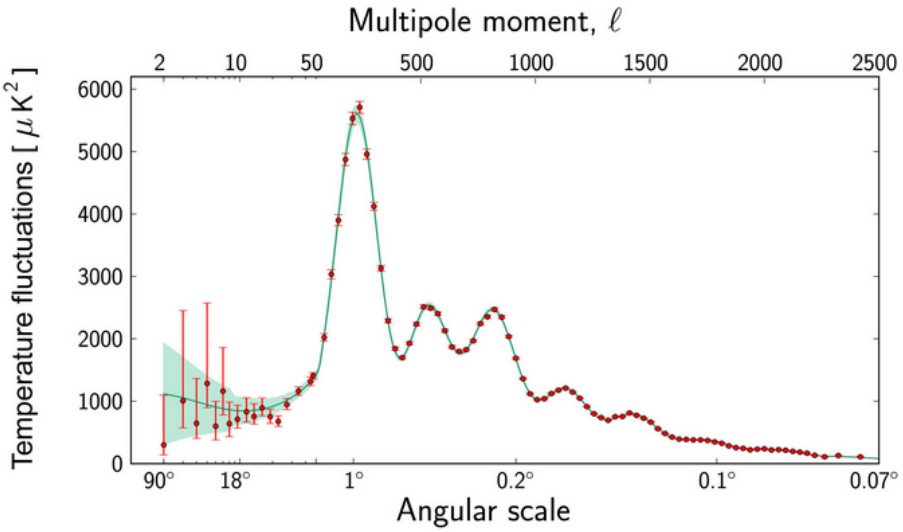

Figure 1.5: Angular power spectrum of the CMB starting from $l = 2$. The x-axis displays the angular scales, with the corresponding representations of multipoles shown at the top. The red dots represent Planck satellite measurements, together with their error bars. The green curve illustrates the best-fitting model within the framework of the ΛCDM model. The surrounding pale green band encompasses the spread of predictions from model variants that remain consistent with observational data (Ade et al., 2014).

time of last scattering) but also the result of all the interactions experienced by photons from redshift $z \sim 1100$ to the present day. These secondary anisotropies are of great cosmological significance, as they encode valuable information about the growth of density perturbations and the subsequent formation of cosmic structures.

In Figure 1.5 we present the angular power spectrum of the CMB, which encapsulates a wealth of physical information. A brief overview of the most striking features is presented below.

At large angular scales, the spectrum retains the imprint of primordial fluctuations, those generated in the very early Universe. Moving to smaller scales, the regular series of peaks and troughs correspond to acoustic oscillations of the tightly coupled photon–baryon fluid, oscillating within the gravitational potential wells shaped by dark matter. These oscillations are progressively damped on the smallest scales due to Silk damping, a diffusive mechanism that erases anisotropies as photons gradually escape from over-dense regions through repeated scattering. Of particular interest is the position of the first acoustic peak, which is linked to the size of the sound horizon at recombination. Encoded in its multipole position is crucial information about the

geometry of the universe, allowing to infer the value of the total energy density parameter Ω_{tot}. Although numerous other physical processes contribute to the detailed shape of the spectrum, a full treatment would extend beyond the scope of this thesis. We therefore limit ourselves to a brief mention of some secondary anisotropies, which arise from photon interactions occurring after the last scattering. Even in this domain, a rich variety of effects emerge, of which we highlight only a few below.

- The early ISW effect occurs because photons traveling from last scattering to today experience time-dependent gravitational potential variations due to the non-negligible radiation energy density at recombination. It appears on angular scales just larger than the first acoustic peak, at slightly lower multipole moments, l.

- The late Sachs–Wolfe effect results from the decay of gravitational potentials when cosmological perturbations evolve differently than in an EdS universe, evolving more rapidly or more slowly, as in the case of the ΛCDM model. It becomes significant at redshifts $z \lesssim 1$, affecting large angular scales near the horizon size, when dark energy dominates.

- The Rees-Sciama effect is the nonlinear counterpart of the ISW effect, linked to potential changes in the nonlinear regime.

- The Sunyaev-Zeldovich effect arises as photons, much cooler than hot cluster plasmas, gain energy through Inverse Compton scattering, showing prominently at large l values, corresponding to small angular cluster sizes.

While the correspondence between theory and observation is remarkably tight across small and intermediate angular scales, a mild suppression is evident at large scales, particularly around 6 degrees. These large-scale anomalies in the CMB pattern may point to features not yet fully accounted for by the standard model, hinting at the need to revisit or refine certain aspects.

1.12 Formation of cosmic structures

Linear theory

To study the evolution of the Universe, it is essential to comprehend the mechanisms through which primordial matter fluctuations amplify over time. In this context, the introduction of the Jeans theory (Jeans, 1902) is mandatory. The fundamental approach starts from the core equations of linear Jeans theory, which are the conservation of mass, the Euler and the Poisson equation:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho v) = 0, \tag{1.62}$$

$$\frac{\partial v}{\partial t} + (v \cdot \nabla)v = -\frac{\nabla p}{\rho} - \nabla \phi, \tag{1.63}$$

$$\nabla^2 \phi = 4\pi G \rho. \tag{1.64}$$

Assuming an adiabatic behaviour of the perturbations, the entropy evolution is governed by:

$$\frac{\partial S}{\partial t} = 0. \tag{1.65}$$

Thus, the equation of state $p = p(\rho, S)$ simplifies to $p = p(\rho)$, with pressure p now depending solely on the density field. These equation chracterise the evolution of large-scale matter distribution.

The next step in our treatment involves incorporating the effects of the background expansion of the Universe. To this end, we introduce two quantities that will prove essential in the forthcoming analyses: the conformal time η and the conformal Hubble parameter $\mathcal{H}$, defined respectively as:

$$\eta(z) = \int_z^\infty \frac{dz'}{H(z')}, \tag{1.66}$$

$$\mathcal{H} = \frac{a'}{a} = \frac{\mathrm{d}a}{\mathrm{d}\eta}\frac{1}{a} = aH. \tag{1.67}$$

where a prime denotes a derivative with respect to the conformal time.

The velocity can be decomposed into two components: the background Hubble flow and the peculiar velocity $\boldsymbol{u}$, resulting in

$$\boldsymbol{v} = \mathcal{H}\boldsymbol{x} + a\boldsymbol{u}. \tag{1.68}$$

The density contrast δ, a key quantity in structure formation, is given by:

$$\delta(\boldsymbol{x}, \eta) = \frac{\rho(\boldsymbol{x}, \eta) - \overline{\rho}}{\overline{\rho}} \tag{1.69}$$

where $\overline{\rho}$ denotes the background (mean) density. By substituting this definition into the continuity and Euler equations, using the conformal quantities introduced before, we obtain:

$$\delta' + \nabla \left[(1 + \delta)\boldsymbol{u}\right] = 0, \tag{1.70}$$

$$\boldsymbol{u}' + \mathcal{H}\boldsymbol{u} + (\boldsymbol{u}\cdot\nabla)\boldsymbol{u} = -\nabla\Phi - \frac{\nabla p}{\rho} \tag{1.71}$$

where Φ is the gravitational potential.

In the linear regime, assuming small perturbations, these equations reduce to

$$\delta' + \nabla\boldsymbol{u} = 0, \tag{1.72}$$

$$\boldsymbol{u}' + \mathcal{H}\boldsymbol{u} = -\nabla\Phi \tag{1.73}$$

Taking the divergence of the linearised Euler equation, and combining it with the continuity equation, leads to a second-order differential equation governing the evolution of the density contrast:

$$\delta'' + \mathcal{H}\delta' - \frac{3}{2}\Omega_{\mathrm{M},0}\mathcal{H}^2\delta = 0 \tag{1.74}$$

where $\Omega_{\mathrm{M},0}$ is the present-day matter density parameter

The general solution to this equation can be written as a linear combination of two independent solutions:

$$\delta(\boldsymbol{x},\eta) = D_+(\eta)A(\boldsymbol{x}) + D_-(\eta)B(\boldsymbol{x}), \tag{1.75}$$

being A, B functions which describe the initial density field. $D_+(\eta)$ and $D_-(\eta)$ are the amplitude of the growing and decaying modes, respectively. The decaying mode rapidly becomes negligible, and the growth of structures is governed by the growing mode $D_+(\eta)$, commonly referred to as the linear growth factor. As a result, the density contrast evolves proportionally to this growth factor:

$$\delta(\boldsymbol{x},\eta) = D_+(\eta)\delta(\boldsymbol{x},0). \tag{1.76}$$

In a FLRW Universe containing only matter and vacuum energy, with densities Ω_{M} and Ω_Λ, the linear growth factor is given by

$$D_+(a) = \frac{5}{2}H(a)\Omega_{\mathrm{M},0}\int_0^a \frac{\mathrm{d}a'}{a'^3H(a')} \tag{1.77}$$

with $H(a) = (\Omega_{\mathrm{M}}a^{-3} + \Omega_{\mathrm{k}}a^{-2} + \Omega_\Lambda a^{-1})^{1/2}$ and the curvature density Ω_{k} is defined to be the density deficit $\Omega_{\mathrm{k}} \equiv 1 - \Omega_{\mathrm{M}} - \Omega_\Lambda$, which is respectively positive, zero, and negative in open, flat, and closed Universes (Hamilton, 2001). Its logarithmic derivative is the dimensionless linear growth rate, f:

$$f = \frac{\mathrm{d}\ln D_+(a)}{\mathrm{d}\ln a} \tag{1.78}$$

is related to the amplitude of the peculiar velocities and thus to redshift space distortions (RSD, section 2.4). The growth of structure provides a particularly sensitive

probe for testing modified gravity scenarios. Within the ΛCDM model, based on GR, a well-established relationship exists between the linear growth rate and the matter density parameter:

$$f \simeq \Omega_{\mathrm{M}}^{\gamma}(z) \tag{1.79}$$

where the growth index parameter $\gamma \simeq 0.545$ is a precise prediction of GR. This relation therefore offers a powerful means to test the validity of GR on cosmological scales: any significant deviation between the observed growth rate and the prediction under GR assumptions could point towards new physics governing the LSS of the Universe.

Chapter 2

Clustering

2.1 The Newtonian gauge

Before delving into the analysis of clustering statistics, it is crucial to clarify the gauge in which our physical quantities are defined. The observed distribution of galaxies, both on the sky and along the line of sight, is shaped not only by the underlying matter density, but also by a range of relativistic effects, all of which must be framed within a consistent theoretical setup.

It is therefore appropriate at this stage to make our choice of gauge explicit. While any gauge is, in principle, equally legitimate, some gauges lend themselves more naturally to physical interpretation. In the context of galaxy clustering, the conformal Newtonian gauge stands out as particularly convenient: it is well suited for dealing with scalar perturbations of the metric, although it can be generalised to include vector and tensor modes (Bertschinger, 1993). This gauge will underpin our relativistic description of density fluctuations, peculiar velocities, RSD, and gravitational effects, all of which leave a distinct imprint on the observed LSS. This choice of gauge is also central to modern treatments of relativistic galaxy number counts, as it provides a clear link between metric perturbations and observational effects such as RSD, relativistic Doppler terms, and gravitational lensing (Yoo, 2010; Bonvin and Durrer, 2011; Breton et al., 2019).

In the conformal Newtonian gauge, the perturbed FLRW metric with scalar perturbations takes the form:

$$\mathrm{d}s^2 = a^2(\eta)\left[-(1+2\Psi)\mathrm{d}\eta^2 + (1-2\Phi)\delta_{ij}\mathrm{d}x^i\mathrm{d}x^j\right], \tag{2.1}$$

where Ψ and Φ are the Bardeen potentials (Bardeen, 1980) and we have set $c = 1$. In the absence of anisotropic stress, the two potential coincide, $\Psi = \Phi$, and the metric simplifies to:

$$\mathrm{d}s^2 = a^2(\eta)\left[-(1+2\Phi)\mathrm{d}\eta^2 + (1-2\Phi)\delta_{ij}\mathrm{d}x^i\mathrm{d}x^j\right]. \tag{2.2}$$

In this case, the potential Φ can be interpreted as a gravitational potential in the Newtonian limit. The assumption of negligible anisotropic stress is not particularly far-fetched for most of cosmic history, as in a matter- and dark energy-dominated universe such contributions are indeed negligible.

Within this gauge, the Poisson equation reads:

$$\Delta\Phi = \frac{3\Omega_0}{2}\frac{\delta(\boldsymbol{x},\eta)}{a}, \tag{2.3}$$

where all quantities are expressed in comoving coordinates and adopting a conformal time parametrisation, given by Eqs. (1.66) and (1.67), assuming a flat background and a linear perturbation theory. At linear order, the comoving peculiar velocity field is related to the potential by:

$$v^i(\boldsymbol{x},\eta) = \frac{2}{3\Omega_0}\mathcal{H}f\,\nabla^i\Phi. \tag{2.4}$$

This result can be obtained by solving the continuity and Euler equations in linear theory under the assumption $\delta \propto D_+(\eta)$ with D_+ the linear growth factor, and using the Eq. (1.78), thereby neglecting decaying modes.

As mentioned in the beginning of this section, the following treatment adopts this specific gauge: it is therefore not gauge-invariant, and hence not fully general. Nevertheless, it is the most commonly employed approach in the literature, because, being quasi-Newtonian, it allows, under certain well-justified assumptions (such as the absence of anisotropic stress), to express the relevant quantities in a familiar Newtonian form.

2.2 The power spectrum

As we have seen, cosmological perturbations originate from quantum fluctuations at the end of the inflationary era and evolve under the effect of gravitational interactions encoded in metric perturbations. In relativistic treatment, the perturbations $\delta(\boldsymbol{x},\eta)$, defined by Eq. (1.69), are functions of both conformal time η and spatial coordinates $\boldsymbol{x}$, and evolve under the action of the metric potentials expressed in a given gauge, such as the Newtonian gauge.

Standard single-field slow-roll inflationary models predict that the fluctuations follow a Gaussian distribution, so that the probability of observing a perturbation with a density contrast $\delta = \delta(\boldsymbol{x},\eta)$ is given by

$$p(\delta) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{\delta^2}{2\sigma^2}}, \tag{2.5}$$

where σ denotes the standard deviation. It is important here to consider the ergodic hypothesis, which states that, for a sufficiently large and statistically homogeneous system, ensemble averages can be replaced by spatial averages within a single realisation. Since we have access to only one observable Universe, we cannot perform an average over multiple realisations of the cosmic density field. Instead, under the assumption of statistical homogeneity and isotropy on large scales, we rely on spatial averages within our single realisation to estimate ensemble properties. This substitution allows us to infer statistical quantities, such as the 2PCF, from observations of the LSS.

We have previously stated that our primary interest is in the statistical analysis of perturbations, δ. However, it is also aknowledged that not all matter components are observable. Let us define the numerical fluctuation of galaxies as the ratio between the number of galaxies within a given volume and the mean number of galaxies, expressed as

$$\delta_g = \frac{\delta N_g(V)}{\overline{N}_g}. \tag{2.6}$$

It is reasonable to hypothesise that galaxies, like other tracers, could serve as a fiducial sample of the total matter distribution, and thus follow the same density fluctuations. Consequently, at linear order, we can relate these quantities as

$$\delta_g = b\,\delta_M, \tag{2.7}$$

where b denotes the linear bias factor, which is dependent on redshift and mass, and δ_M represents the matter density field averaged over a certain volume. This quantity can be defined at each position $\boldsymbol{x}$ as

$$\delta_M(\boldsymbol{x}) = \frac{\delta M(\boldsymbol{x})}{\overline{M}} = \delta(\boldsymbol{x}) * W(\boldsymbol{x}, R), \tag{2.8}$$

where the symbol $*$ denotes a convolution operation, $\overline{M}$ is the mass averaged over a given volume and W is a convolution filter, commonly known as the window function. Several filters can be employed, such as the top-hat filter or the Gaussian filter. In the latter case, the advantage lies in the fact that convolving with a Gaussian function does not alter the properties of the density field itself, which remains of Gaussian type, though its variance evolves as

$$\sigma_M^2(R) = \langle \delta_M^2(\boldsymbol{x}) \rangle. \tag{2.9}$$

where $\langle\rangle$ is the ensemble average. The variance is a decreasing monotonic function with increasing radius. In the limiting case where the filtering radius tends to ∞, the result is the averaged field. Since σ_M^2 is proportional to the square density field fluctuations, and the growth of these fluctuations is scale-independent, it follows that, starting from an initial variance distribution, one can expect it to exhibit self-similarity after a certain amount of time. Given that $\sigma_M^2 = 1$ defines the threshold below which linear theory is applicable, it is noteworthy that smaller scales are the first to transition into the nonlinear regime. The variance is linked to the power spectrum $P(k)$, and we aim to investigate this relationship. Standard inflation models predict the absence of preferred scales, allowing the primordial power spectrum to be characterised by a power-law form, such as:

$$P_{\rm in}(k) = A \left(\frac{k}{k_*}\right)^{n_s}, \tag{2.10}$$

where A represents the amplitude, n_s denotes the spectral index and k_* is a pivot scale. Notably, the spectral index can take values greater than or less than zero, signifying a dominance of power on either small or large scales, respectively. However, since the variance depends on k through an integral, it is necessary for it to converge that $n > -3$ for $k \to 0$ and $n < -3$ for $k \to \infty$. Moreover, given the moment at which a certain mass becomes nonlinear, a relationship between time and mass is established, we must further constrain the values of the spectral index, if we wish for smaller masses to form prior to larger ones, and energy to be an increasing function of mass. This is called hierarchical formation, which requires n to be in the range $-3 < n < 1$. Among the several values that n can assume, there exists a particular one, predicted by inflation, which derives from the fact that potential fluctuations are mass independent, exhibiting a white-noise-like (i.e. scale-invariant) behaviour. This is realised for $n = 1$ and corresponds to the so-called Zel'dovich spectrum (Zeldovich, 1972). As perturbations evolve from inflation to the present epoch, different physical processes modify their amplitudes in a scale-dependent way. This evolution is encoded in the transfer function $T(k)$, which relates the present-day potential fluctuations to the primordial ones. Accordingly, the linear matter power spectrum becomes:

$$P_{\rm L}(k, a) \propto D_+^2(a)\, T^2(k)\, P_{\rm in}(k), \tag{2.11}$$

where $D + (a)$ is the growth factor. For biased tracers, such as galaxies, the power spectrum reads:

$$P_g(k, a) = b^2(a) P_{\rm L}(k, a). \tag{2.12}$$

In modern cosmological analyses, $T(k)$ is computed using Boltzmann codes such as

CLASS (Lesgourgues, 2011) or CAMB (Lewis and Challinor, 2011), and include the effects of dark matter, baryons, neutrinos, radiation and the cosmological constant.

2.3 The 2-point correlation function

We now focus on the counterpart of the power spectrum in configuration space, to characterise structural clustering, that is, the correlation function, as originally formulated in the 1960s (Totsuji and Kihara, 1969), and further developed during the 1970s and 1980s (Peebles and Hauser, 1974).

Considering the probability of finding a first galaxy (or any other object) in a small comoving volume $\mathrm{d}V_1$ at a position $\boldsymbol{x}_1$, and a second galaxy in a volume $\mathrm{d}V_2$ at a position $\boldsymbol{x}_2$, we have, respectively:

$$\mathrm{d}P_1 = \bar{n}\mathrm{d}V_1, \tag{2.13}$$

$$\mathrm{d}P_2 = \bar{n}\mathrm{d}V_2. \tag{2.14}$$

where $\bar{n}$ denotes the mean number density of objects per unit volume.

The uncorrelated probability of simultaneously finding one galaxy in $\mathrm{d}V_1$ and another in $\mathrm{d}V_2$ is then

$$\mathrm{d}^2P = \bar{n}^2\,\mathrm{d}V_1\,\mathrm{d}V_2, \tag{2.15}$$

with the mean number of objects in a volume V given by $\langle N\rangle = \bar{n}\,V$.

However, when spatial clustering is present, the joint probability of locating a second galaxy within $\mathrm{d}V_2$, given that a first one has already been observed in $\mathrm{d}V_1$, is modified to:

$$\mathrm{d}^2P = \bar{n}^2\mathrm{d}V_1\mathrm{d}V_2\left[1+\xi(\boldsymbol{x}_1,\boldsymbol{x}_2)\right], \tag{2.16}$$

where $\xi(\boldsymbol{x}_1,\boldsymbol{x}_2)$ is the 2PCF. Statistical homogeneity and isotropy imply that ξ depends only on the comoving separation $\boldsymbol{r} = \boldsymbol{x}_2 - \boldsymbol{x}_1$, and hence $\xi(\boldsymbol{x}_1,\boldsymbol{x}_2) = \xi(\boldsymbol{r}) = \xi(r)$. Moreover, $\xi(r) \geq -1$ ensures that $\mathrm{d}^2P \geq 0$.

As evident in Eq. (2.16), the 2PCF, $\xi(r)$, serves as a measure of the excess (or deficit) in the joint probability relative to the expectation under a random spatial distribution. The 2PCF can also be directly related to the density field, as follows. Given that $\rho_b = \bar{n}\,m$, and $\rho(\boldsymbol{x}) = n(\boldsymbol{x})m$, we can express the number density as

$$n(\boldsymbol{x}) = \frac{\rho(\boldsymbol{x})}{\rho_b}\,\bar{n}. \tag{2.17}$$

The probability of locating an object within a volume $\mathrm{d}V$, at the position $\boldsymbol{x}$, is

$$\mathrm{d}P(\boldsymbol{x}) = n(\boldsymbol{x})\,\mathrm{d}V = \frac{\rho(\boldsymbol{x})}{\rho_b}\,\bar{n}\mathrm{d}V. \tag{2.18}$$

Hence, the probability of finding two objects, one situated at position $\boldsymbol{x}$ and the other at position $\boldsymbol{x}+\boldsymbol{r}$, is given by

$$\mathrm{d}^2P = \frac{\bar{n}^2}{\rho_b^2}\,\langle\rho(\boldsymbol{x})\,\rho(\boldsymbol{x}+\boldsymbol{r})\rangle\,\mathrm{d}V_1\mathrm{d}V_2. \tag{2.19}$$

Since

$$\delta(\boldsymbol{x}) = \frac{\rho(\boldsymbol{x})}{\rho_b} - 1, \tag{2.20}$$

then $\xi(\boldsymbol{r}) = \langle\delta(\boldsymbol{x})\delta(\boldsymbol{x}+\boldsymbol{r})\rangle$, which is the 2PCF in the continuum.

The 2PCF is positive at small scales and becomes negative at large scales. This means that dark matter haloes and galaxies cluster on small scales with a probability larger than that of the homogeneous case. In particular, if we want to estabilish the 2PCF for biased objects, we have to multiply the dark matter 2PCF by the square of the linear bias factor b:

$$\xi_{\mathrm{obj}}(\boldsymbol{r}) = b^2\xi_{\mathrm{DM}}(\boldsymbol{r}). \tag{2.21}$$

Finally, the relation between the 2PCF and the power spectrum is the following:

$$\xi(\boldsymbol{r}) = \frac{1}{(2\pi)^3}\int \mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{r}}P(\boldsymbol{k})\,\mathrm{d}^3\boldsymbol{k}. \tag{2.22}$$

2.3.1 The 2-point cross-correlation function

We now aim to introduce the cross-correlation function that this thesis focuses on. The formalism follows that previously introduced for the auto-correlation, with the difference that we now consider two sets of tracers, labeled 1 and 2. The cross-correlation function ξ_{12}, at a separation $\vec{r}$, is thus defined through the joint probability of locating an object from set 2 in a volume $\mathrm{d}V_2$, given that an object from set 1 has already been found in a volume $\mathrm{d}V_1$:

$$\mathrm{d}^2P = \bar{n}_1\bar{n}_2\mathrm{d}V_1\mathrm{d}V_2\left[1+\xi_{12}(\boldsymbol{r})\right], \tag{2.23}$$

where $\bar{n}_1, \bar{n}_2$ are the mean number densities per unit volume of objects of sets 1 and 2, respectively. The cross-correlation function between the two populations is:

$$\xi_{12}(\boldsymbol{r}) = \langle\delta_1(\boldsymbol{r}_1)\delta_2(\boldsymbol{r}_2)\rangle, \tag{2.24}$$

where $\delta_1(\boldsymbol{r}_1)$ and $\delta_2(\boldsymbol{r}_2)$ are the overdensisites of populations 1 and 2.

The cross-correlation between two populations of galaxies or haloes is anti-symmetric under exchange of the line-of-sight coordinates. Suppose that we have a brighter population, B, and a fainter population, F, that we want to cross-correlate: the correlation changes if the bright galaxy is in front of or behind the faint one. The cross-correlation function $\xi_{\rm BF}$ is made of a symmetric and an anti-symmetric part:

$$\xi_{\rm BF} = \xi^s_{\rm BF} + \xi^a_{\rm BF}, \tag{2.25}$$

with

$$\xi^s_{\rm BF} \equiv \frac{1}{2}\left[\xi_{\rm BF}(\Delta x_z, \Delta\boldsymbol{x}_\perp) + \xi_{\rm BF}(-\Delta x_z, -\Delta\boldsymbol{x}_\perp)\right], \tag{2.26}$$

$$\xi^a_{\rm BF} \equiv \frac{1}{2}\left[\xi_{\rm BF}(\Delta x_z, \Delta\boldsymbol{x}_\perp) - \xi_{\rm BF}(-\Delta x_z, -\Delta\boldsymbol{x}_\perp)\right]. \tag{2.27}$$

The terms Δx_z and $\Delta_{\boldsymbol{x}_\perp}$ represent the line-of-sight (along z-axis) and transverse separation:

$$\Delta x_z \equiv x_{z,\rm B} - x_{z,\rm F}, \tag{2.28}$$

$$\Delta_{\boldsymbol{x}_\perp} \equiv \boldsymbol{x}_{\perp,\rm B} - \boldsymbol{x}_{\perp,\rm F}. \tag{2.29}$$

2.3.2 Estimators of the 2-point correlation function

From an observational point of view, the 2PCF is inferred from discrete tracers of the underlying matter distribution. This requires comparing the number of pairs of objects separated by a specific distance with the expected number of pairs that would occur if the objects were distributed randomly. The most straightforward estimator for the 2PCF (Peebles and Hauser, 1974) can be expressed as

$$\xi(r) = \frac{DD(r)}{RR(r)} - 1, \tag{2.30}$$

where $DD(r)$ and $RR(r)$ denote the normalised number of data-data and random-random object pairs, respectively, each separated by a distance r, and normalised with respect to the total number of random-random pairs.

More accurate estimators for both the auto-correlation and cross-correlation functions can be found in the literature. For this thesis project, the Landy-Szalay estimator was used to determine both the auto-correlation and the cross-correlation functions (Landy and Szalay, 1993):

$$\xi(r) = \frac{DD(r) - 2DR(r) - RR(r)}{RR(r)}, \tag{2.31}$$

$$\xi(r) = \frac{D_1D_2(r) - D_1R_2(r) - R_1D_2(r) + R_1R_2(r)}{R_1R_2(r)}, \tag{2.32}$$

where the subscripts 1 and 2 refer to the two different catalogues used, D_1D_2 and R_1R_2 represent the pair counts for data-data and random-random objects, respectively, while D_1R_2, R_1D_2, DR correspond to mixed data-random terms.

2.3.3 Multipole moments

The 2PCF has already been defined as the probability of finding two objects separated by a distance r. Consequently, if the system is isotropic, the 2PCF depends only on the distance. However, this assumption ceases to hold once redshifts are used to infer distances without accounting for peculiar velocities and relativistic effects, thereby breaking the isotropy of the system. In such circumstances, the 2PCF must account for two dependencies: one on the norm of the distance r, and the other on the angle along the line of sight. As a result, 2PCF is generally expressed as $\xi = \xi(r, \mu)$, where μ is the cosine of the angle between the line of sight and the comoving separation r.

Multipole moments are important summary statistics for analysing the anisotropic 2PCF, $\xi(r, \mu)$. When considering an isotropic system, the only non-vanishing multipole is ξ_0, that is, the monopole. When introducing RSD (section 2.4), the isotropy breaks, and if we further consider a non-symmetric distribution under inversion, even odd multipoles arise. The l^{th} order multipole moment of the 2PCF is given by:

$$\xi_l(r) = \frac{2l+1}{2} \int_{-1}^{1} \xi(r, \mu)\mathcal{L}_l(\mu)\, \mathrm{d}\mu, \tag{2.33}$$

where $\mathcal{L}_l(\mu)$ is the Legendre polynomial of l^{th} order. In practice, due to the finite number of object pairs, the integral is approximated by a discrete sum as:

$$\xi_l(r) \simeq \frac{2l+1}{2} \sum_{\mu=-1}^{1} \xi(r, \mu)\mathcal{L}_l(\mu)\, \delta\mu. \tag{2.34}$$

Here we consider the first five multipoles:

$$\xi_0(r) = \sum_{\mu=0}^{1} \xi(r, \mu)\, \delta\mu, \tag{2.35}$$

$$\xi_1(r) = \frac{3}{2} \sum_{\mu=-1}^{1} \xi(r, \mu)\mu\, \delta\mu, \tag{2.36}$$

$$\xi_2(r) = \frac{5}{2} \sum_{\mu=0}^{1} \xi(r, \mu)(3\mu^2 - 1)\, \delta\mu, \tag{2.37}$$

$$\xi_3(r) = \frac{7}{4}\sum_{\mu=-1}^{1} \xi(r,\mu)(5\mu^3 - 3\mu)\,\delta\mu, \tag{2.38}$$

$$\xi_4(r) = \frac{9}{8}\sum_{\mu=0}^{1} \xi(r,\mu)(35\mu^4 - 30\mu^2 + 3)\,\delta\mu. \tag{2.39}$$

2.4 Mapping from real to redshift space

Our primary objective, from now on, is to derive all the contributions that lead to an asymmetry in the cross-correlation function, and consequently to the emergence of odd multipoles, with particular focus on extracting relativistic effects, as will be discussed later. We begin by describing all sources of distortion, starting with the standard RSD. We emphasise that its derivation relies on the Newtonian linear approximation and the distant-observer limit. We subsequently address geometric distortions, light-cone and evolutionary effects, and finally lensing contributions. A dedicated chapter will be reserved for relativistic distortions, followed by a comprehensive summary of all the aforementioned contributions.

In redshift space, the observed positions of cosmological objects are affected by RSD, which arise when the observed redshifts are interpreted as purely cosmological, neglecting the contribution from peculiar velocities. This effect induces an anisotropy in the observed matter field, enhancing the observed clustering in redshift space relative to that in real space.

On large scales, peculiar motions are predominantly governed by the gravitational infall of structures, such as galaxies, during cosmic formation, a phenomenon known as the Kaiser effect: this process induces a characteristic anisotropy in the correlation function, manifesting as a flattening along the line of sight. Conversely, on smaller scales, RSD arise primarily from randomised velocities within virialised structures, but contributions may also stem from coherent or chaotic motions in their surroundings, especially in the quasi-linear infall regions, giving rise to the so-called Fingers of God (FoG). This effect causes elongation of the structures along the line of sight. All these effects are shown in Figure 2.1.

It is possible to establish the relationship between the power spectrum in real space and in redshift space by drawing on key theoretical considerations (Kaiser, 1987). The first one states that the number density of objects must be preserved when mapping between these two representations, which is expressed as

$$n_s(\boldsymbol{x}_s)d^3\boldsymbol{x}_s = n_r(\boldsymbol{x}_r)d^3\boldsymbol{x}_r, \tag{2.40}$$

where the subscripts s and r denote quantities in redshift and real space, respectively.

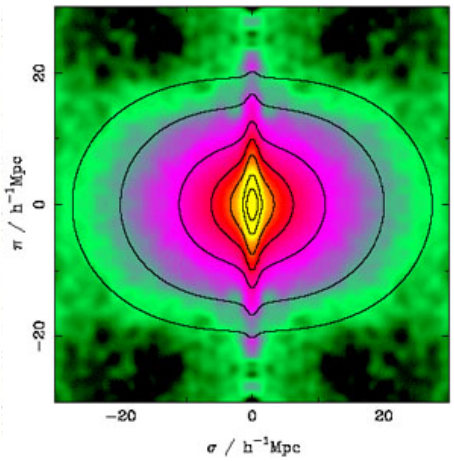

Figure 2.1: The 2-dimensional redshift-space 2PCF $\xi(\sigma, \pi)$ of the 2dF Galaxy Redshift Survey is shown as a function of transverse (σ) and radial (π) pair separations. The contour lines correspond to constant values of the correlation function at $\xi = 10, 5, 2, 1, 0.5, 0.2, 0.1$. Notable features include the radial elongation at small separations (the so-called "Fingers of God") and the large-scale flattening due to coherent infall (Kaiser effect). For visual clarity and to enhance the perception of anisotropy, the data from the first quadrant (upper right) have been symmetrically reflected about both coordinate axes (Peacock et al., 2001).

The second aspect is that the observed redshift, at first order, is given by:

$$z_{\text{obs}} \simeq H_0 x + \boldsymbol{v} \cdot \hat{\boldsymbol{x}}, \tag{2.41}$$

where the first term corresponds to the Hubble-Lemaître flow, while the second arises from the peculiar velocity component along the line of sight. The observed position of an object is given as:

$$x_s = x + \frac{\boldsymbol{v} \cdot \hat{\boldsymbol{x}}}{aH}. \tag{2.42}$$

Introducing the Jacobian J as

$$J := \left| \frac{d^3\boldsymbol{x}}{d^3\boldsymbol{x}_s} \right|, \tag{2.43}$$

and solving for Eq. (2.42), while taking into account the Eq. (2.40), leads to

$$1 + \delta_s = (1 + \delta) \left[1 - \frac{\partial}{\partial x} \left(\frac{\boldsymbol{v} \cdot \hat{\boldsymbol{x}}}{aH} \right) \right]. \tag{2.44}$$

A first-order expansion reveals that, in redshift space, the overdensity field consists of the sum of its real counterpart and a correction term. We adopt the distant observer

approximation, which corresponds to the small-angle or flat-sky limit: the core assumption is that the line-of-sight direction, $\hat{\boldsymbol{n}} = \boldsymbol{x}/x$, can be regarded as constant across the observed region. In other words, provided that galaxies lie close to one another within the (x, y) plane, we may replace $\hat{\boldsymbol{x}}$ in Eq. (2.44) with a fixed radial unit vector directed towards the centre of the survey area, that is, for example, $\hat{\boldsymbol{e}}_z$. For these reasons, taking the Fourier transform of the overdensity yields the following expression for the perturbation in the density field of haloes and galaxies:

$$\delta_s(\boldsymbol{k}) = \left[b + f\mu_k^2\right] \delta(\boldsymbol{k}), \tag{2.45}$$

where $\mu_k = \boldsymbol{k} \cdot \hat{\boldsymbol{e}}_z$ denotes the cosine of the angle between the line of sight and the wavevector $\boldsymbol{k}$. As previously introduced, the factor f is the growth rate, defined as the logarithmic derivative of the growth factor D_+, Eq. (1.78), which describes the growth of the matter perturbations. There exists an empirical fitting formula for the growth rate f as follows:

$$f(a) \sim \Omega_m^\gamma(a), \tag{2.46}$$

where the exponent is $\gamma \simeq 0.545$ in GR, and we will explain its implications in what follows.

Eq. (2.45) formalises the behaviour regarding large-scale RSD. Firstly, given that $f\mu_k^2 \geq 0$, the observed overdensity in redshift space always exceeds that in real space, where it would simply be $b\,\delta(\boldsymbol{k})$. A second key aspect of Eq. (2.45) is that the amplification specifically affects perturbations whose wavevectors are aligned with the line of sight. In contrast, perturbations with $\boldsymbol{k}$ perpendicular to the line of sight are unaffected by RSD. Therefore the galaxy power spectrum in redshift space also depends on the orientation of $\boldsymbol{k}$, not only on its value: this dependence is accounted for by μ_k.

Thus, the power spectrum in redshift space can be expressed in terms of its real-space counterpart as

$$P_s(\boldsymbol{k}, \mu_k) = P_{\mathrm{L}}(\boldsymbol{k}) \left(1 + f\mu_k^2\right)^2 + P_{\mathrm{N}}, \tag{2.47}$$

where $P_{\mathrm{L}}(\boldsymbol{k})$ is the linear matter power spectrum, P_{N} is a white-noise term in the galaxy power spectrum (i.e. a scale-independent constant).

Let us now recall Eq. (2.45) to see how this is related to the multipoles of the correlation function through the multiplying factor:

$$\begin{aligned}(b + f\mu_k^2)^2 &= b^2 + f^2\mu_k^4 + 2f\mu_k^2 \\ &= b^2 + \frac{1}{5}f^2 + \frac{2}{3}bf + \left(\frac{4}{7}f^2 + \frac{4}{3}bf\right)\mathcal{L}_2(\mu_k) + \frac{8}{35}\mathcal{L}_4(\mu_k).\end{aligned} \tag{2.48}$$

As we can see from this expression, this introduces two even multipoles in addition to the monopole of the 2PCF and power spectrum in redshift space.

It is important at this point to recall the expression for the growth rate f. In fact, f provides a direct observational window into the nature of dark energy: a more rapid expansion of the Universe, driven by dark energy, leads to a lower growth rate of cosmic structures. Conversely, if a modified gravity model is considered at fixed $H(z)$, an enhanced growth rate is generally expected, as most modifications to gravity tend to increase its strength. A particularly significant test, therefore, involves comparing the observed growth rate with the value predicted by GR. In this context, measurements of RSD can place constraints on f, potentially leading to challenges or further validations of GR.

2.5 Geometric distortions

Geometric distortions (GD) arise when an incorrect cosmological model is assumed to convert redshifts and angular positions into comoving coordinates. This effect overlaps with the dynamical distortions discussed in the previous section, but manifests itself even in the absence of peculiar velocities. Since radial and transverse coordinates depend differently on cosmological parameters, any mismatch between the true and the assumed cosmology leads to a deformation of the observed structures. As a result, intrinsically isotropic distributions may appear elongated or compressed along the line of sight, providing a powerful geometric test of the underlying cosmological model.

2.5.1 The Alcock-Paczynski test

In 1979, Alcock and Paczynski devised a test aimed at placing constraints on cosmological parameters (Alcock and Paczynski, 1979). In what follows, we reconstruct its fundamental steps.

Consider the observed galaxy position, given by the spherical coordinates (θ, ϕ) and the redshift z. These coordinates can be converted into a 3D position $\boldsymbol{x}_{\rm obs}$ and this is possible if we choose a fiducial cosmology and thus a fiducial distance-redshift relation $\chi(z)$, which differs from the one of the real Universe:

$$\chi_{\rm fid}(z) = \chi_{\rm true}(z) + \delta\chi(z). \tag{2.49}$$

To simplify the calculations, we can set $\boldsymbol{x}_{\rm obs} = 0$ and fix the redshift to a mean value, $z = \bar{z}$. For the transverse component, the assigned values are

$$(x_1, x_2)_{\rm obs} = \chi_{\rm fid}(z) \cdot (\theta_1, \theta_2), \tag{2.50}$$

where θ_1, θ_2 represent the two components of θ. Meanwhile, if the true cosmology was considered, the relation would be:

$$(x_1, x_2)_{\rm true} = \chi_{\rm true}(z) \cdot (\theta_1, \theta_2), \tag{2.51}$$

or, in terms of $(x_1, x_2)_{\rm obs}$, as

$$(x_1, x_2)_{\rm true} = \left(1 - \frac{\delta\chi(z)}{\chi_{\rm fid}(z)}\right)(x_1, x_2)_{\rm obs}. \tag{2.52}$$

Thus, as evident from this equation, if $\delta\chi > 0$ then the comoving distance χ that we have assigned to the galaxy in the transverse direction is larger than the real one, and the opposite happens for $\delta\chi < 0$.

The line-of-sight observed component $x_{3,\rm obs}$ is related to the redshift (recall that we have fixed it to a mean value $\bar{z}$) through the following expression:

$$x_{3,\rm obs}(z) = \chi_{\rm fid}(z) - \chi_{\rm fid}(\bar{z}) \simeq \frac{1}{H_{\rm fid}(\bar{z})}(z - \bar{z}), \tag{2.53}$$

where the second step follows from a first-order expansion in $z - \bar{z}$. This approximation is valid when considering narrow redshift slices. It also derives from the definition of the derivative of the comoving distance with respect to redshift, $\mathrm{d}\chi/\mathrm{d}z = 1/H$, assuming $c = 1$. As for the transverse coordinates, the true value of x_3 (corresponding to the correct cosmology) is given by:

$$x_{3,\rm true}(z) = \frac{1}{H_{\rm true}(\bar{z})}(z - \bar{z}) = \frac{H_{\rm fid}(\bar{z})}{H_{\rm true}(\bar{z})}\, x_{3,\rm obs}(z). \tag{2.54}$$

Using $\delta H(z) = H_{\rm true}(z) - H_{\rm fid}(z)$ we obtain:

$$x_{3,\rm true}(z) = \left(1 + \frac{\delta H(\bar{z})}{H_{\rm fid}(\bar{z})}\right) x_{3,\rm obs}. \tag{2.55}$$

Finally, the relationship between the true and observed coordinates, to account for due to a wrong distance-redshift relation is

$$\boldsymbol{x}_{\rm true} = ([1 - \alpha_\perp]x_{1,\rm obs},\ [1 - \alpha_\perp]x_{2,\rm obs},\ [1 - \alpha_\parallel]x_{3,\rm obs}), \tag{2.56}$$

where

$$\alpha_\parallel = -\frac{\delta H}{H_{\rm fid}(z)}, \tag{2.57}$$

$$\alpha_\perp = \frac{\delta\chi}{\chi_{\rm fid}(z)}, \tag{2.58}$$

which, as said, we can safely evaluate at $\bar{z}$. We can also express $\alpha_\parallel$ and $\alpha_\perp$ in terms of true quantities as:

$$\alpha_\parallel + 1 = \frac{H_{\rm true}(z)}{H_{\rm fid}(z)}, \tag{2.59}$$

$$\alpha_\perp + 1 = \frac{\chi_{\rm true}(z)}{\chi_{\rm fid}(z)}. \tag{2.60}$$

This effect is known as the "Alcock-Paczynski" (AP) effect (from which follows the AP test).

We now briefly show how this wrong cosmology can affect the power spectrum. We start from defining the true wavevector in terms of the observed one as follows:

$$\boldsymbol{k}_{\rm true} = ([1-\alpha_\perp]k_{1,\rm obs},\ [1-\alpha_\perp]k_{2,\rm obs},\ [1-\alpha_\parallel]k_{3,\rm obs}). \tag{2.61}$$

By considering the Fourier transform of the redshift-space overdensity, we can easily write

$$P_s(\boldsymbol{k}_{\rm obs}, \bar{z}) = P_{\rm L}(\boldsymbol{k}, \bar{z}) \left(1 + f\mu_k^2\right)^2 \Big|_{([1-\alpha_\perp]k_{1,\rm obs},\ [1-\alpha_\perp]k_{2,\rm obs},\ [1-\alpha_\parallel]k_{3,\rm obs})} + P_{\rm N}. \tag{2.62}$$

The difference between this equation and the Eq. (2.47) lies in the inclusion here of coordinate rescaling effects. The form of the expression is preserved, thanks to the fact that the number of galaxies in a given volume is independent of the coordinate system used to describe them.

The Eq. (2.62) includes two effects:

- The RSD, due to ignoring peculiar velocities, derived in Section 2.4;
- The AP effect, arising from the fact that the coordinates assigned to galaxies are based on an assumed distance–redshift relation, which may differ from the true one.

The constant noise term $P_{\rm N}$ is unaltered by either of these effects. While RSD in the linear regime affects only the amplitude of the power spectrum, the AP effect modifies its shape. This distinction allows us to disentangle the two contributions. In this sense, the AP test is a powerful cosmological probe, helping us to discriminate between different cosmological models.

Chapter 3

Redshift-space distortions in the dipole of cross-correlation

One of the key scientific goals of the *Euclid* mission is to detect, with high statistical significance, the anisotropies in the large-scale distribution of galaxies. Such measurements offer a stringent observational test of GR on cosmological scales and have the potential to unveil subtle signatures of modified gravity or relativistic effects imprinted on the LSS of the Universe.

In this thesis, we have taken a further step beyond classical dynamical and geometrical distortions, introducing the full range of distortions, including relativistic contribution, that affect redshift measurements and leave imprints in the dipole of the 2-point cross-correlation function. These distortions have been the central focus of the work presented here. Most studies focus on the even multipoles of the 2PCF, which are largely dominated by standard RSD. In contrast, wide-angle, evolution, and relativistic effects induce an anti-symmetry in the observed distribution of cosmic objects when cross-correlating two different galaxy populations, thereby giving rise to odd multipoles. In recent years, many authors have investigated the impact of relativistic distortions on object counts and on the 2PCF (Bonvin and Durrer, 2011; Challinor and Lewis, 2011). Calculations remain highly non-trivial and are therefore typically carried out in the linear regime and under the distant-observer approximation. A linear mapping between real and redshift space is also assumed. Within this framework, one can isolate and analyse the various contributions to the multipoles of the 2PCF on linear scales. Extending beyond standard RSD effects to include relativistic contributions allows for additional tests of GR and provides valuable insights into the nature of dark matter and dark energy. It is essential to account for these relativistic RSD effects when performing precise analyses, as standard RSD alone do not generate an antisymmetric component in the cross-correlation function. Although incorporating these extra terms is more challenging due to their subtlety, it enables more stringent tests of GR on large cosmological scales.

This is especially important because these relativistic contributions depend on the linear growth rate, which characterises the evolution of cosmic structures and is closely connected to GR via the growth index parameter γ, Eq. (1.79).

3.1 Wide-angle effect

It is essential to incorporate additional terms when seeking more accurate predictions than those provided by standard RSD.

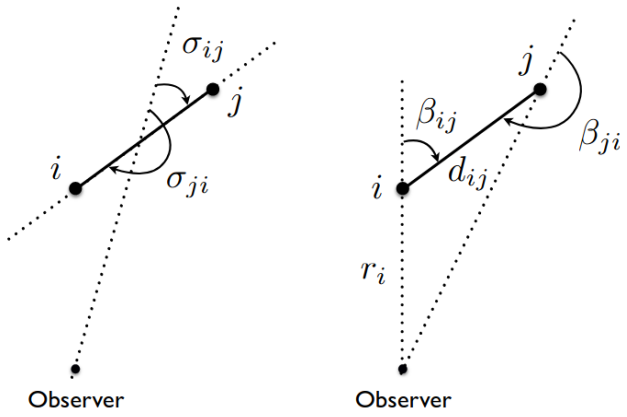

Figure 3.1: Two alternative definitions of the angle employed to extract the dipole are considered, both arising from the breakdown of the distant observer approximation (Gaztanaga et al., 2017).

One such correction arises upon relaxing the distant observer approximation, wherein the lines of sight to the two galaxies in a pair are no longer parallel. This deviation introduces modifications to the standard Kaiser expressions for the monopole, quadrupole, and hexadecapole, and, moreover, induces an additional antisymmetric contribution to the 2PCF. The precise form of this wide-angle effect depends crucially on the definition of the angle used to extract the multipoles.

There are various ways of defining the angle between a pair of galaxies and the observer's line of sight. In Figure 3.1 two of the most common choices are illustrated: σ_{ij} represents the angle between the median and the vector connecting the median to j, while β_{ij} is the angle between the direction of i and the vector connecting i to j. The former definition ensures that i and j are handled symmetrically. The latter, β_{ij}, appears more natural when capturing relativistic effects: it centres the coordinate system on one galaxy and characterises how the correlation signal varies around it.

This is particularly relevant for effects such as gravitational redshift, which are expected to generate a dipolar modulation about each galaxy. In galaxy groups and

clusters it is due to the differential gravitational potential between the central, massive galaxy, located deeper in the potential well, and the surrounding, less massive companions (Wojtak et al., 2011).

At leading order in d/r, the mathematical expression of the dipole does not depend on the choice of angular variable. Differences arising from angular parametrisation only appear at order $(d/r)^3$, as a consequence of geometric projection effects rather than any physical dependence on coordinates. For the scales considered in this thesis, such higher-order corrections are entirely negligible. Let us take a step back and reconsider the cross-power spectrum, which may be expressed as:

$$P_{\rm gal} = (b_1 + f\nu_1^2)(b_2 + f\nu_2^2)P(k), \tag{3.1}$$

where $\nu_{1,2}$ denote the cosine of the angles between the wavevector $\hat{\boldsymbol{k}}_2$ and $\hat{\boldsymbol{k}}_2$ and the line of sight, respectively, and b_1, b_2 are the biases of the two populations.

This expression can be expanded in Legendre polynomials and transformed into configuration space via an inverse Fourier transform. However, this yields a result that depends on two angular variables, $\mu_1 = \hat{\boldsymbol{x}}_1 \cdot \hat{\boldsymbol{r}}$ and $\mu_2 = \hat{\boldsymbol{x}}_2 \cdot \hat{\boldsymbol{r}}$, which proves cumbersome for interpretation. For the purposes of this thesis we choose the angle defined by the scalar product between the median position of the galaxy pair and their separation vector. Being $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ the positions of objects of the first and second populations, we thus have

$$\boldsymbol{x} = \frac{\boldsymbol{x}_1 + \boldsymbol{x}_2}{2}, \tag{3.2}$$

$$\boldsymbol{r} = \boldsymbol{x}_2 - \boldsymbol{x}_1. \tag{3.3}$$

Considering the angle defined by $\mu = \hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{r}}$, which is symmetric under exchange of pairs, we do not need any additional terms for the dipole due to the choice of angle. We also have $\nu = \hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{x}}$, $\nu_1 = \hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{x}}_1$, $\nu_2 = \hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{x}}_2$ and $\nu_r = \hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{r}}$. Expressing all quantities using these definitions and expanding to linear order in $\mathcal{O}(r/x)$, we obtain:

$$(b_1 + f\nu_1^2)(b_2 + f\nu_2^2) = A + \frac{r}{x}B, \tag{3.4}$$

where A is the standard symmetric contribution responsible for the even multipoles, and B is the antisymmetric term that gives rise to the dipole and octupole terms:

$$B = (b_1 - b_2)f\left[\mathcal{L}_1(\nu_r)\mathcal{L}_1(\nu) - \frac{1}{3}\mathcal{L}_1(\mu) - \frac{2}{3}\mathcal{L}_1(\mu)\mathcal{L}_2(\nu)\right], \tag{3.5}$$

where $\mathcal{L}_l$ are the l^{th}-order Legendre polynomials (Gaztanaga et al., 2017; Breton et al., 2019).

The wide-angle contribution to the odd multipoles is given by:

$$\xi^{\rm wa}(r,\mu) = (b_1 - b_2)\frac{2r}{5x} f\left[-\mathcal{L}_1(\mu) + \mathcal{L}_3(\mu)\right] \gamma_2^0(r), \tag{3.6}$$

where

$$\gamma_2^0 = \frac{1}{(2\pi)^2}\int k^2 j_2(kr) P(k,z)\,\mathrm{d}k, \tag{3.7}$$

with $P(k,z)$ the linear matter power spectrum, $j_2(kr)$ the 2^{nd}-order spherical Bessel function and $\mathcal{H}_0$ the conformal Hubble parameter at $z = 0$.

3.2 Light-cone and evolution effects

When analysing the clustering properties of high-redshift objects at a quantitative level, it is essential to properly account for light-cone effects (Kaiser, 2013). Cosmological observations are, in fact, only feasible on the light-cone hypersurface defined by the current observer. The time evolution of the sources, such as the bias, introduces contamination in the observational data.

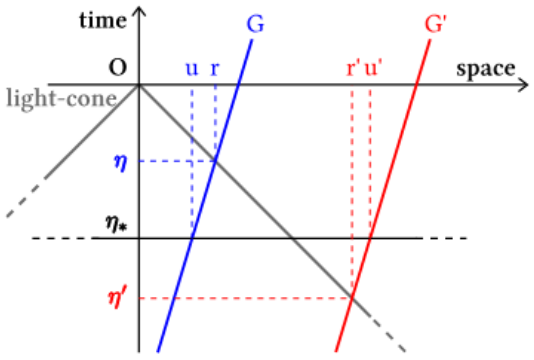

Figure 3.2: Representation of the light-cone effect. When considering observations on the light cone, the peculiar motions of galaxies G and G' cause their observed radial positions to differ from those they had at the reference conformal time η_*. As a result, their true separation along the line of sight at the time of observation, $r' - r$, does not coincide with the radial separation $u' - u$ that they had at η_* (Bonvin et al., 2014).

Given a metric $g_{\mu\nu}$, the past-light cone of a point x^μ is defined as the set of all points y^μ for which there exists a null geodesic connecting y^μ to x^μ. It therefore includes all events that can be observed at x^μ via light propagation[1] In practice, the light cone can

[1]To be more precise: let $(\mathcal{M}, g)$ be a 4D Lorentzian manifold, where $\mathcal{M}$ is a differentiable manifold representing spacetime, and g the metric tensor. At each point $p \in \mathcal{M}$, there exists a tangent space

be parameterised by a pair of angles, defining the direction of observation, and a redshift. This has a fundamental implication: we are unable to observe anything beyond our past light cone. As previously mentioned, cosmological observations do not correspond to a snapshot taken on a constant-time hypersurface, but rather to what is observed along the light cone. Consequently, the coordinate system used, i.e. angles and redshift, is affected by these perturbations and must be corrected accordingly.

These effects can be probed using a key observable: galaxy number counts. The observed galaxy counts are affected by perturbations in the underlying distribution, such that:

$$\Delta_{\rm obs} = \Delta + \delta\Delta, \tag{3.8}$$

where $\delta\Delta$ is the correction term and $\boldsymbol{x} = \boldsymbol{x}_{\rm obs} - \delta\boldsymbol{x}$ quantifies the difference between the observed and true position of the source. It should be noted that, in computing this quantity at a given point $x^\mu = (\eta, \boldsymbol{x})$, one typically neglects the fact that photons propagate through a perturbed spacetime, as including this would lead to higher-order contributions in the fluctuation expansion (Bonvin et al., 2014).

To illustrate the light-cone effect, consider a galaxy located at a position $\boldsymbol{q}$ at a conformal time η_*, and assume for simplicity that all galaxies share a common peculiar velocity $\boldsymbol{v}$. At a later conformal time η the position of the galaxy will be:

$$\boldsymbol{x} = \boldsymbol{q} + \boldsymbol{v}(\eta - \eta_*). \tag{3.9}$$

Since the observed galaxy position $(\eta, \boldsymbol{x})$ lies on the observer's light cone, the radial distance is given, to the lowest order, by $r = \eta_0 - \eta$, where η_0 is the conformal time today. If we denote by u the line-of-sight component of the initial position, $\boldsymbol{q}$, and by r the observed radial position, then we obtain:

$$r = u + \hat{n}^i v_i(\eta - \eta_*) = u + \hat{n}^i v_i(r_* - r). \tag{3.10}$$

This lies at the core of the light-cone effect, namely that we observe galaxies which are not stationary but have moved over time, and this displacement affects the coordinates by which we measure them. This is summarised in Figure 3.2: the true separation between galaxies along the line of sight is represented by the difference $r' - r$, which differs from their separation at time η_*. The light-cone effect thus introduces a density fluctuation in the observed number counts, given by

$$\Delta^{\rm LC} = \frac{\boldsymbol{v}\cdot\boldsymbol{n}}{\mathcal{C}}, \tag{3.11}$$

which gives rise to the following anti-symmetric term contributing to the dipole:

$T_p\mathcal{M}$: the tangent bundle is defined as the disjoint union of all tangent spaces over $\mathcal{M}$. The light cone at a point p is the set of tangent vectors $v \in T_p\mathcal{M}$ satisfying the condition $g(v, v) = 0$.

$$\xi^{\rm LC} = -\frac{1}{2\pi^2} f(b_1 - b_2)\frac{\mathcal{H}}{c}\int {\rm d}k\, k\, j_1(kr) P(k,z), \tag{3.12}$$

where j_l is the spherical Bessel function of l^{th}-order.

Let us now extract the evolution term, by connecting these fluctuations with that a fixed time η_* through a Taylor expansion as:

$$[\Delta + \delta\Delta]_\eta \simeq [\Delta + \delta\Delta]_{\eta_*} + \frac{\partial}{\partial\eta}[\Delta + \delta\Delta]\bigg|_{\eta_*}(\eta - \eta_*), \tag{3.13}$$

with $\eta - \eta_*$ small compared to the Hubble time. To the time difference $\eta - \eta_*$ corresponds a spatial difference $r - r_*$ with $r \equiv \eta_0 - \eta_*$, being η_0 the conformal time today. The second term on the right-hand side is referred to as the evolution term. This term also accounts for the spatial dependence of $\vec{v}$ since, if the number count of objects is conserved, at linear order we have

$$-\nabla_i v_i = \frac{\partial\Delta}{\partial\eta}. \tag{3.14}$$

These evolution contributions to the antisymmetry in the cross-correlation function are dominated by the evolution of the bias and of the growth rate. As we have already shown in the previous section, also in this case at the first order the choice of the angle does not affect these terms.

Such contributions to the dipole can be written as:

$$\langle\xi^{\rm evo1}\rangle = \frac{r}{6}\left[(b_1 - b_2)f' - f(b_1' - b_2')\right]\left[\gamma_0^0(r) - \frac{4}{5}\gamma_2^0(r)\right], \tag{3.15}$$

$$\langle\xi^{\rm evo2}\rangle = \frac{r}{2}(b_1 b_2' - b_2 b_1')\gamma_0^0(r), \tag{3.16}$$

where

$$\gamma_0^0(r) = \frac{1}{2\pi^2}\int {\rm d}k\, k^2 j_0(kr) P(k,z), \tag{3.17}$$

$$\gamma_2^0(r) = \frac{1}{2\pi^2}\int {\rm d}k\, k^2 j_2(kr) P(k,z). \tag{3.18}$$

A prime denotes a derivative with respect to the comoving distance.

We decomposed the evolution term into two contributions, $\xi^{\rm evo1}$ and $\xi^{\rm evo2}$, since $\xi^{\rm evo2}$ is also generated in real space and it does not contribute to the dipole when only the relativistic terms are considered by subtracting the contributions in real space.

3.3 Gravitational lensing term

When light propagates through regions of gravitational potential, produced by massive objects such as galaxy clusters, it is deflected, which can result in the source of that light appearing at a different angular position from its true location. The study of gravitational lensing is not as recent as one might assume, with its origins predating even the development of GR. In the early 1800s, Soldner computed the deflection of light by modelling it as a particle and applying Newtonian mechanics, deriving a deviation angle of $2GM/(bc^2)$.

With the discovery of GR by Einstein, researchers confirmed that gravity bends light and explored its implications. Eddington and others predicted that gravitational lensing could generate multiple images of distant objects. Einstein demonstrated that stellar lensing was too subtle to detect, but Zwicky proposed that galaxies could produce observable image splittings, approximately 1 arcsecond, thus making lensing a valuable tool in extragalactic astronomy, although it remains a relatively rare phenomenon.

Gravitational lensing was first observed in 1979 with the discovery of the "twin quasar" (Walsh et al., 1979). Since then, numerous lensed quasars, galaxies, arcs, and Einstein rings have been discovered, with current and upcoming surveys expected to identify thousands more.

Consider observing a source emitting photons, such as a galaxy, and let β denote the true angular position, which is the direction from which a photon would have arrived in the absence of any deflection. The angle θ corresponds to the observed direction of the photon, as seen by the observer. Under the distant observer approximation, where the deflection angle and the reduced deflection angle can be considered equivalent ($\hat{\alpha} \simeq \alpha$), the lens equation reads:

$$\beta = \theta - \alpha. \tag{3.19}$$

A schematic representation of this configuration is provided in Figure 3.3.

The deflection angle of a photon (whether reduced or not, as already noted) is given by:

$$\hat{\boldsymbol{a}} = \frac{2}{c^2} \int \boldsymbol{\nabla}_{\perp} \Phi \, \mathrm{d}\lambda. \tag{3.20}$$

This expression represents the gradient of the dimensionless Newtonian potential (Φ/c^2) perpendicular to the light ray, integrated along the photon's path. The factor of two arises because the perturbed Minkowski metric features identical perturbations in both its temporal and spatial components.

We now consider a simple application by examining the case of a point mass potential, $\Phi = -GM/r$. Let the photon travel along the z-axis, and let $\vec{x} = (x, y)$ represent coordinates in the plane perpendicular to the direction of motion. Under the

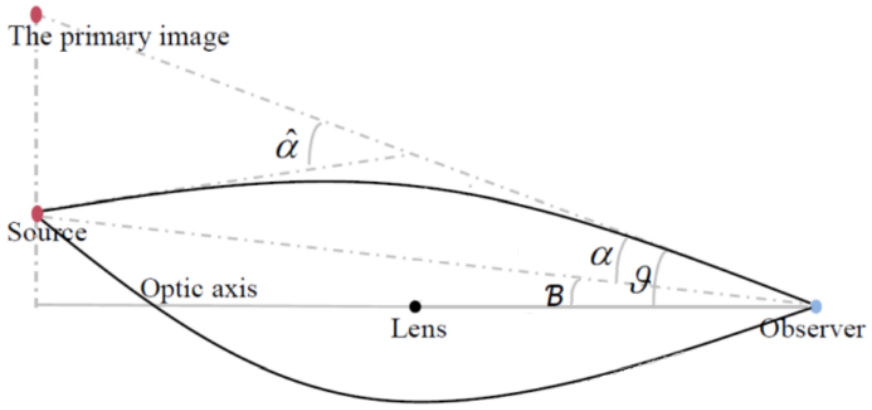

Figure 3.3: Lensing effect: the apparent source is seen from the observer under an angle θ, this is the result of the photon deflection due to the potential of the lens, encountered on its path (Gao, 2024).

assumption of distant observer, that is, the light ray is emitted at $z = -\infty$ and received at $z = +\infty$, we obtain:

$$\hat{\boldsymbol{a}} = \frac{2}{c^2}\int_{-\infty}^{+\infty} \boldsymbol{\nabla}_\perp \Phi \, \mathrm{d}z \tag{3.21}$$

$$= \frac{2}{c^2}\int_{-\infty}^{+\infty} \mathrm{d}z \begin{pmatrix} \partial\Phi/\partial x \\ \partial\Phi/\partial y \end{pmatrix} \tag{3.22}$$

$$= \frac{2}{c^2}\int_{-\infty}^{+\infty} \mathrm{d}z \frac{GM}{r^3} \begin{pmatrix} x \\ y \end{pmatrix} \tag{3.23}$$

$$= \frac{2GM}{c^2}\boldsymbol{x}\int_{-\infty}^{+\infty} \mathrm{d}z \frac{1}{(b^2+z^2)^{3/2}} \tag{3.24}$$

$$= \frac{4GM}{bc^2}\hat{\boldsymbol{x}}, \tag{3.25}$$

where $b = (x^2 + y^2)^{1/2}$ and $\hat{\boldsymbol{x}} = \boldsymbol{x}/b$ is the unit vector in the transverse plane. The final result is precisely twice what would be obtained from a purely Newtonian treatment: to a hypothetical massless particle moving at a speed c, a heuristic model inconsistent with Newtonian mechanics but useful as a point of comparison with GR. Imagine to observe two populations of galaxies, the first one brighter and a fainter second one located behind the first one. The lensing produced by the brighter population will affect the second population leading to fluctuations in the number galaxies counts, which induces an anti-symmetry. Thus, the lensing contribution to the dipole is the following, as given by Matsubara (2000):

$$\begin{aligned}\langle \xi^{\rm lens} \rangle &= \langle \Delta^{\rm std} \Delta^{\rm lens} \rangle \\ &= -\frac{9\Omega_{\rm M}(z) r \mathcal{H}^2}{8\pi}(b_1 - b_2) \int_{-1}^{1} \mu^2 {\rm d}\mu \int {\rm d}k\, k J_0 \left(kr\sqrt{1-\mu^2} \right) P(k,z),\end{aligned} \tag{3.26}$$

where J_0 is the first kind Bessel spherical function of 0^{th}-order.

3.4 Relativistic Doppler term

As has been extensively emphasised throughout, our primary observable is the galaxy number counts. These counts are not solely influenced by density fluctuations and velocity gradients, as considered by the standard RSD model. Instead, a richer phenomenology emerges, incorporating contributions from the gravitational potential and relativistic corrections arising from peculiar velocities. The relativistic Doppler effect provides, in fact, an additional contribution that must be taken into account, and its contribution to the number density fluctuation is given by:

$$\Delta^{\rm Rel,Doppler} = -\left(\frac{\dot{\mathcal{H}}}{c\mathcal{H}^2} + \frac{2}{\mathcal{H}\chi} \right) \boldsymbol{v} \cdot \boldsymbol{n}, \tag{3.27}$$

where dots denote derivatives with respect to conformal time. It should be noted that the most significant part of this contribution stems from the second term in the round brackets:

$$\Delta^{\rm div} = -\frac{2}{\mathcal{H}\chi} \boldsymbol{v} \cdot \boldsymbol{n}. \tag{3.28}$$

Let us rewrite Eq. (2.42) in an equivalent form as[2]:

$$\boldsymbol{s} = \boldsymbol{r} + \frac{\boldsymbol{v} \cdot \boldsymbol{n}}{\mathcal{H}}. \tag{3.29}$$

The volume elements in real and redshift space are, respectively:

$${\rm d}^3 r = r^2 {\rm d}r\, {\rm d}\Omega, \tag{3.30}$$

$${\rm d}^3 s = s^2 {\rm d}s\, {\rm d}\Omega, \tag{3.31}$$

with

$${\rm d}s = {\rm d}r + \frac{1}{\mathcal{H}} \nabla_r (\boldsymbol{v} \cdot \boldsymbol{n}) {\rm d}r. \tag{3.32}$$

[2]Here the time dependence is not considered.

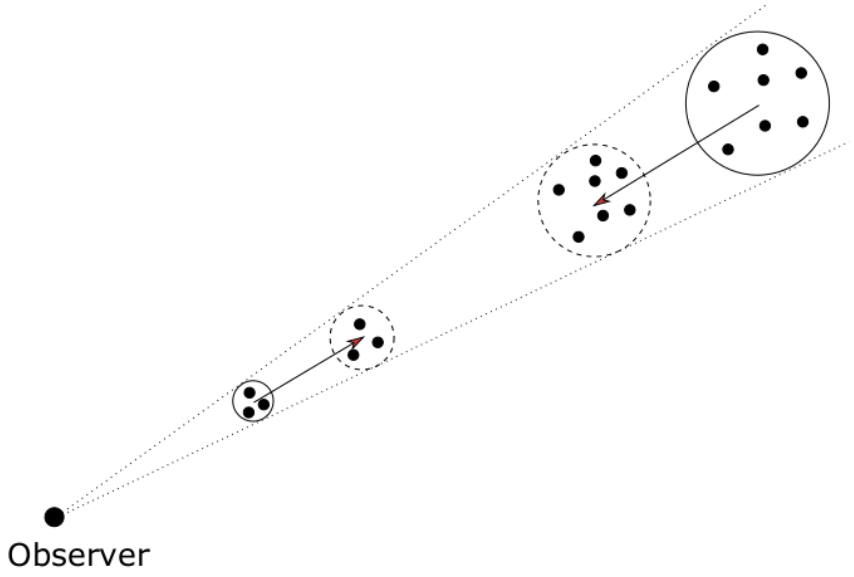

Figure 3.4: Line of sight divergence effect: the solid circles represent the true distribution of objects, whereas the dashed circles correspond to the observed distribution. Arrows denote the overall peculiar velocity of the system (Breton, 2018).

At linear order, the Jacobian reads:

$$\left|\frac{\mathrm{d}\boldsymbol{s}}{\mathrm{d}\boldsymbol{r}}\right| = 1 + \frac{1}{\mathcal{H}}\nabla_r(\boldsymbol{v}\cdot\boldsymbol{n}) + \frac{2}{r\mathcal{H}}(\boldsymbol{v}\cdot\boldsymbol{n}). \tag{3.33}$$

The third term of this equation leads to the effect commonly referred to in the literature as the line-of-sight divergence effect, and it is clearly illustrated in Figure 3.4: a volume subtending a given solid angle appears denser if it is moving towards the observer, and less dense if receding. The effect is proportional to the velocity field along the line of sight and becomes the dominant contribution to the dipole on large scales. Although this term originates from a purely geometric consideration, it naturally enters the relativistic framework when adopting a fully relativistic treatment of redshift distortions in a perturbed FLRW spacetime. It is therefore conventionally grouped under the label relativistic Doppler contribution, as it forms part of the complete set of relativistic corrections to galaxy number counts. By contrast, it is typically absent from standard treatments of RSD, as it becomes observationally relevant in the context of cross-correlating two distinct galaxy populations. This is because the projection of peculiar velocities along the line of sight differs depending on whether a galaxy lies in front of or behind its companion in the pair. As a result, an anti-symmetric contribution emerges in the correlation function, manifesting itself as a non-vanishing dipole.

The contribution from the accelerating expansion of the universe is given by

$$\Delta^{\mathrm{q}} = -\frac{\dot{\mathcal{H}}}{c\mathcal{H}^2}\boldsymbol{v}\cdot\boldsymbol{n}. \tag{3.34}$$

The whole contribution to the dipole, arising from the two terms considered here, is the following:

$$\langle\xi^{\mathrm{Rel,Doppler}}\rangle = \frac{1}{2c\pi^2}\left(\frac{\dot{\mathcal{H}}}{\mathcal{H}} + \frac{2c}{\chi}\right) f\,(b_1 - b_2)\int \mathrm{d}k\, kj_1(kr)P(k,z). \tag{3.35}$$

The first term in the round brackets is of the order of 10^{-3}, and it overwhelmingly dominates over the contribution arising from the acceleration of the cosmic expansion, which is of the order of 10^{-5}.

3.5 Contribution from acceleration

We also introduce an additional term contributing to the generation of a dipole that arises from a variation in the velocity of the observed objects along the line of sight and enters the fluctuation in number counts as:

$$\Delta^{\mathrm{acc}} = \frac{1}{\mathcal{H}c}\dot{\boldsymbol{v}}\cdot\boldsymbol{n}. \tag{3.36}$$

The contribution to the dipole reads:

$$\langle\xi^{\mathrm{acc}}\rangle = -\frac{1}{2c\pi^2}\left(f^2 + \frac{\dot{f}}{\mathcal{H}} + \frac{\dot{\mathcal{H}}}{\mathcal{H}^2}f\right)\mathcal{H}(b_1 - b_2)\int \mathrm{d}k\, kj_1(kr)P(k,z) \tag{3.37}$$

3.6 Gravitational redshift

The second effect we need to introduce is that arising from the gravitational potential (McDonald, 2009). As a photon climbs out of a gravitational potential well, it undergoes an energy loss, manifesting itself as a redshift. When considering two distinct galaxy populations that differ in mass-luminosity, the associated photons are redshifted to varying degrees. This differential gives rise to a breakdown of the isotropy in the observed galaxy distribution.

In Figure 3.5 a schematic difference between classic RSD (left panel) and gravitational redshift (right panel) is shown. We consider galaxies residing within the symmetric gravitational potential well of a cluster, observed from a distant point, denoted as O. Let B represent the central brighter galaxy of the cluster, located at the bottom of the gravitational potential well. The other cluster members are labeled F_1 and F_2, denoting

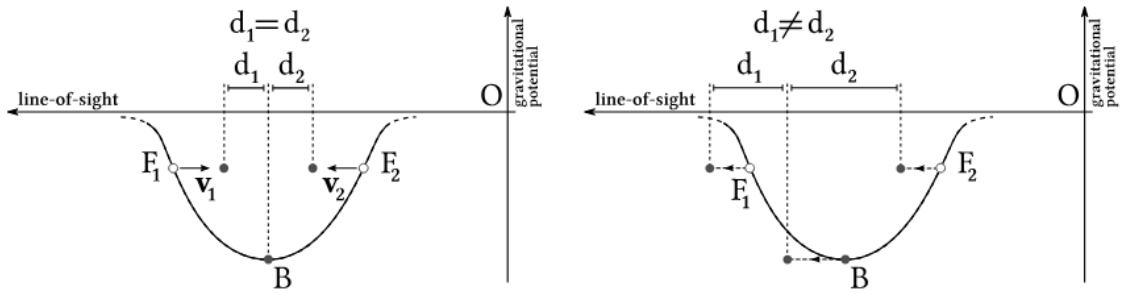

Figure 3.5: Schematic illustration of the difference between Doppler RSD, on the left-hand side and gravitational redshift distortions, on the right-hand side (Bonvin et al., 2014).

two such galaxies (fainter than B) positioned symmetrically with respect to B in real space. In redshift space, the relative positions of these galaxies are altered: galaxies appear compressed along the line of sight. However, this effect is symmetric, thus F_1 and F_2 remain equally displaced with respect to B.

The situation changes when gravitational redshift is considered. In this case, all three galaxies are shifted in the same direction, but the central galaxy B undergoes the largest gravitational redshift. This leads to an asymmetry: F_1 now appears closer to B than F_2, breaking the front-back symmetry.

The result is, once again, a cross-correlation function composed of two terms: one symmetric and one antisymmetric. In Fourier space, this corresponds to an imaginary part in the power spectrum.

We can express the contribution of the potential term to the fluctuation in number counts by decomposing it into two components:

$$\Delta^{\mathrm{pot},1} = \frac{1}{\mathcal{H}c}\nabla_r \Psi \cdot \boldsymbol{n}, \tag{3.38}$$

which correspond to the first order contribution in $\mathcal{H}/k$, and

$$\Delta^{\mathrm{pot},2} = \left(\frac{\dot{\mathcal{H}}}{\mathcal{H}^2} + \frac{2c}{\mathcal{H}\chi}\right)\frac{\Psi}{c^2} - \frac{1}{\mathcal{H}c^2}\dot{\Psi}, \tag{3.39}$$

which represents the next-to-leading order contribution in $\mathcal{H}/k$, retaining only the dominant terms. These give rise to dipolar contributions of the form:

$$\langle \xi^{\mathrm{pot},1} \rangle = \frac{3\mathcal{H}}{4c\pi^2}\Omega_{\mathrm{M}}(z)(b_1 - b_2)\int \mathrm{d}k\, k j_1(kr) P(k,z), \tag{3.40}$$

$$\langle \xi^{\mathrm{pot},2} \rangle = -\frac{3\mathcal{H}\mathcal{H}_0}{4\pi^2 c^2}\Omega_{\mathrm{M}}(z)(b_1 - b_2)\left(\frac{\dot{\mathcal{H}}}{\mathcal{H}^2} + \frac{2c}{\mathcal{H}\chi} - f + 1\right)\int \mathrm{d}k\, j_1(kr) P(k,z). \tag{3.41}$$

3.7 Shapiro time delay

The Shapiro time delay is an effect arising from the gravitational potential well surrounding massive structures, as first described by Irwin I. Shapiro in 1964. When a photon travels across a potential well of a massive object, its propagation is delayed relative to the time it would take in an unperturbed, flat spacetime. The net result is effectively as though the photon had "slowed down" along its trajectory. Unlike the local gravitational redshift effects, the Shapiro delay is an integrated effect, since it builds up along the line of sight as the photon travels across gravitational fields.

This propagation delay induces a perturbation in the observed signal and, in particular, in the correction to the observed number count fluctuations. The contribution from the Shapiro delay to the number density contrast is encapsulated by an integral over the gravitational potential along the line of sight, and plays a role, albeit subdominant, in the suit of relativistic corrections to LSS observables:

$$\Delta^{\text{Shapiro}} = \frac{2}{c^2\chi}\int_0^{\chi}(\Phi+\Psi)\,\mathrm{d}\chi', \tag{3.42}$$

where Φ and Ψ are the Bardeen potentials.

When applied to galaxy number counts, especially in cross-correlations between two different populations of galaxies (e.g., bright and faint), the Shapiro delay induces an antisymmetric distortion along the line of sight, generating a non-zero dipole in the 2-point cross-correlation function:

$$\langle\xi^{\text{Shapiro}}\rangle = -\frac{3}{2\pi^2c^2}\Omega_{\text{M}}(z)\mathcal{H}^2(b_1-b_2)\int \mathrm{d}k\, j_1(kr)P(k,z). \tag{3.43}$$

3.8 Integrated Sachs-Wolfe effect

A final effect that must be accounted for is the integrated Sachs-Wolfe (ISW) effect, which arises from time-dependent variations in the gravitational potential experienced by photons along their trajectories.

In a spatially flat Universe, within the linear regime, the gravitational potential remains constant in time at any fixed comoving location. Thus, the detection of a significant ISW signal constitutes compelling evidence for a deviation from this behaviour, potentially indicative of an additional component in the energy content of the Universe, such as a cosmological constant.

This effect is clearly evident on large angular scales, but becomes largely subdominant on smaller angular scales, as fluctuations tend to average out along the line of sight (Afshordi, 2004).

The observed density fluctuation in number counts due to the ISW effect is given by:

$$\Delta^{\rm ISW} = \frac{1}{\mathcal{H}c^2}(\dot{\Phi} + \dot{\Psi}), \tag{3.44}$$

which contributes an anti-symmetric term to the dipole

$$\langle \xi^{\rm ISW} \rangle = \frac{3\mathcal{H}}{2\pi^2 c^2}\Omega_M(z)(b_1 - b_2)(1-f)\int \mathrm{d}k\, j_1(kr)P(k,z). \tag{3.45}$$

3.9 Summary of the dipole modelling

We now summarise all the aforementioned effects in order to determine the total expected number of object counts, accounting for both standard RSD contributions and relativistic corrections. We also include, in the most general form possible, the evolution bias $f_{\rm evo}$ and magnification bias s. Magnification bias quantifies the change in the observed number density of galaxies due to gravitational lensing magnification. Specifically, it measures how the observed galaxy number density responds to the survey luminosity threshold at fixed redshift, making it a survey-dependent quantity. Evolution bias, on the other hand, characterises the physical change in the comoving number density of galaxies relative to the conserved case. It describes how the comoving number density evolves with redshift at a fixed luminosity threshold, depending on the nature of the tracer population. Processes such as galaxy and halo formation, evolution, and mergers lead to a non-conservation of the comoving number density, encapsulated by a non-zero evolution bias (Maartens et al., 2021). Ultimately, evolution bias leads galaxies to merge, reducing their number density and changing the clustering properties of the population as a whole. In the simplest model of structure growth, galaxies are always hosted within haloes whose masses increase with time. In such a picture, evolution bias corresponds to the increase in the average halo mass that hosts a particular galaxy after its formation. Larger values of the evolution bias correspond to faster growth of halo masses, primarily due to continuous merging processes (Percival and Schaefer, 2008).

In the linear regime, we begin by considering a linear mapping between real and redshift space, expressed as

$$\Delta_s = \Delta_r + 1 - |J| \tag{3.46}$$

where J denotes the Jacobian of the transformation from real space to redshift space. As outlined in Challinor and Lewis (2011); Bonvin and Durrer (2011); Bonvin et al. (2014); Tansella et al. (2018); Breton et al. (2019), the contributions to the observed galaxy number counts can be computed as the sum of the following terms:

$$\Delta^{\rm std} = b\delta - \mathcal{H}^{-1}\nabla_r(\boldsymbol{v}\cdot\boldsymbol{n}), \tag{3.47}$$

$$\Delta^{\rm Doppler} = -\left(\frac{\dot{\mathcal{H}}}{c\mathcal{H}^2} + \frac{2-5s}{\mathcal{H}\chi} + 5s - f_{\rm evo}\right)\boldsymbol{v}\cdot\boldsymbol{n}, \tag{3.48}$$

$$\Delta^{\rm acc} = \mathcal{H}^{-1}\dot{\boldsymbol{v}}\cdot\boldsymbol{n}, \tag{3.49}$$

$$\Delta^{\rm lens} = \frac{5s-2}{2\chi c^2}\int_0^{\chi(z)}\frac{(\chi-\chi')}{\chi'}\Delta_\Omega(\Phi+\Psi)d\chi', \tag{3.50}$$

$$\Delta^{\rm pot} = \frac{1}{\mathcal{H}c}\boldsymbol{\nabla}_r\Psi\cdot\boldsymbol{n} + \left(\frac{\dot{\mathcal{H}}}{\mathcal{H}^2} + \frac{(2-5s)c}{\mathcal{H}\chi} + 5s - f_{\rm evo}\right)\frac{\Psi}{c^2} - \frac{1}{\mathcal{H}c^2}\dot{\Psi}, \tag{3.51}$$

$$\Delta^{\rm sh} = \frac{2-5s}{\chi c^2}\int_0^{\chi(z)}{\rm d}\chi'(\Psi+\Phi), \tag{3.52}$$

$$\Delta^{\rm ISW} = \frac{1}{c^3}\left(\frac{\dot{\mathcal{H}}}{\mathcal{H}^2} + \frac{(2-5s)c}{\mathcal{H}\chi} + 5s - f_{\rm evo}\right)\int_0^{\chi(z)}{\rm d}\chi'(\dot{\Phi}+\dot{\Psi}), \tag{3.53}$$

$$\Delta^{\rm LC} = \frac{\boldsymbol{v}\cdot\boldsymbol{n}}{c}, \tag{3.54}$$

where Δ_Ω denotes the angular part of the Laplacian:

$$\Delta_\Omega \equiv \left(\cot\theta\partial_\theta + \partial_\theta^2 + \frac{1}{\sin^2\theta}\partial_\phi^2\right), \tag{3.55}$$

which is linked to the term of the Laplacian perpendicular to the line of sight via $\Delta_\Omega = \chi^2\Delta_\perp$. Here δ represents the matter density contrast in Newtonian gauge, b is the linear bias, $\Delta^{\rm rsd}$ is the standard RSD term, $\Delta^{\rm lens}$ is the lensing term, $\Delta^{\rm q}$ describes the contribution from the accelerated expansion of the Universe, $\Delta^{\rm acc}$ derives from the acceleration of the sources, $\Delta^{\rm pot}$ is the contribution from gravitational potential as the sum of the first order term and the second-order term $(\mathcal{H}/k)$, $\Delta^{\rm sh}$ denotes the Shapiro time-delay contribution, $\Delta^{\rm ISW}$ is the Integrated Sachs-Wolfe term and $\Delta^{\rm LC}$ is the light-cone contribution due to the fact that observed position of sources on the light-cone is different from the position on constant-time hypersurface (Bonvin et al., 2014).

Recalling that the 2-point cross-correlation function between two different populations is given by $\xi = \langle\Delta_1^2\Delta_2^s\rangle$, the terms that generate asymmetries read:

$$\xi^{\rm A} = \sum_i\langle\Delta_1^{\rm std}\Delta_2^{\rm A}\rangle + (1)\leftrightarrow(2), \tag{3.56}$$

where $\Delta^{\rm A}$ is given by Eqs. (3.47) to (3.54).

Using Eq. (3.56), we can write all the different contributions to the dipole, such as:

$$\langle\xi^{\rm LC}\rangle = -f\mathcal{G}\gamma_1^1(r), \tag{3.57}$$

$$\langle \xi^{\rm Doppler} \rangle = \left(\frac{\dot{\mathcal{H}}}{\mathcal{H}^2} + \frac{2c}{\mathcal{H}\chi} \right) f\mathcal{G}\gamma_1^1(r), \tag{3.58}$$

$$\langle \xi^{\rm acc} \rangle = -\left(f^2 + \frac{\dot{f}}{\mathcal{H}} + \frac{\dot{\mathcal{H}} f}{\mathcal{H}^2} \right) f\mathcal{G}\gamma_1^1(r), \tag{3.59}$$

$$\langle \xi^{\rm pot} \rangle = \frac{3\Omega_{\rm M}(z)}{2}\mathcal{G}\left[\gamma_1^1(r) - \frac{\mathcal{H}}{\mathcal{H}_0}\left(\frac{2c}{\mathcal{H}\chi} + \frac{\dot{\mathcal{H}}}{\mathcal{H}^2} - f + 1 \right)\gamma_1^2(r) \right], \tag{3.60}$$

$$\langle \xi^{\rm ISW} \rangle = 3\Omega_{\rm M}(z)\frac{\mathcal{H}}{\mathcal{H}_0}\mathcal{G}(1-f)\gamma_1^2(r), \tag{3.61}$$

$$\langle \xi^{\rm wa} \rangle = -\frac{2}{5}(b_1 - b_2) f \frac{r}{\chi}\gamma_2^0(r), \tag{3.62}$$

$$\langle \xi^{\rm evo1} \rangle = \frac{r}{6}\left[(b_1 - b_2)f' - f(b_1' - b_2') \right]\left[\gamma_0^0(r) - \frac{4}{5}\gamma_2^0(r) \right], \tag{3.63}$$

$$\langle \xi^{\rm evo2} \rangle = \frac{r}{2}(b_1 b_2' - b_1' b_2)\gamma_0^0(r), \tag{3.64}$$

$$\langle \xi^{\rm lens} \rangle = \langle \Delta^{\rm std}\Delta^{\rm lens} \rangle = -\frac{9}{4}\Omega_{\rm M}(z)\frac{r\mathcal{H}}{c}\mathcal{G}\bar{\omega}(r,\overline{z}), \tag{3.65}$$

with

$$\gamma_l^m(r)\frac{1}{2\pi^2}\left(\frac{\mathcal{H}_0}{c} \right)^m \int dk\, k^{2-m} j_l(kr) P(k,\overline{z}), \tag{3.66}$$

$$\bar{\omega}(r,\overline{z})\frac{1}{2\pi}\left(\frac{\mathcal{H}_0}{c} \right)^m \int_{-1}^{1} \mu^2\, {\rm d}\mu \int {\rm d}k\, k J_0\left(kr\sqrt{1-\mu^2} \right) P(k,\overline{z}), \tag{3.67}$$

$$\mathcal{G} = (b_1 - b_2)\frac{\mathcal{H}}{\mathcal{H}_0}. \tag{3.68}$$

A dot represents a derivative with respect to the conformal time, while a prime is a derivative with respect to the comoving distance, f is the linear growth rate, j_l is the l^{th}-order spherical Bessel function and J_0 is the Bessel function of the first kind of 0^{th}-order. Here $P(k,\overline{z})$ is the linear power spectrum at a mean redshift $\overline{z}$:

$$\langle \delta(k,\overline{z})\, \delta(k',\overline{z}) \rangle = (2\pi)^3 P(k,\overline{z})\delta_{\rm D}(k,k'). \tag{3.69}$$

Chapter 4

Results

4.1 Methods

For the present thesis, we made use of data from the RayGalGroupSims suite (RayGalGroupSims), a collection of high resolution N-body simulations that incorporate all first-order relativistic observational effects within the weak-field approximation (Breton et al., 2019). The simulated volume spans $2625\,(\mathrm{Mpc/h})^3$ and is populated with $N_p = 4096^3$ particles, with a mass resolution of approximately $1.88 \times 10^{10}\,h^{-1}\,\mathrm{M}_\odot$. Crucially, full-sky light-cones over the redshift interval $z \sim [0, 0.5]$ are constructed on-the-fly during the simulation run, thereby eliminating the need for replicas.

Haloes within the light-cone have been identified using the parallel Friends-of-Friends (pFoF) algorithm. Their apparent positions on the sky have been computed via a ray-tracing procedure that accounts for all first-order relativistic effects within the weak-field approximation. These include redshift and angular distortions induced by the relativistic Doppler effect, gravitational redshift, the transverse Doppler shift, the Integrated Sachs–Wolfe and Rees–Sciama effects, as well as weak gravitational lensing.

4.1.1 N-body simulations

Numerical simulations in cosmology have become indispensable tools, now widely employed across the field. These simulations offer a dual advantage: they allow for testing cosmological analysis pipelines before applying them on observations, and allow us to predict future outcomes, when closed-form solutions are not available. In particular, N-body simulations provide a means to investigate the evolution of systems comprising an enormous number of particles, systems for which no analytical solutions exist for their dynamics.

To generate this catalogue, the RAMSES code (Teyssier, 2002) was employed. This code combines the particle-mesh (PM) method with adaptive mesh refinement, enabling

both high efficiency and accuracy. In the PM approach, particles are distributed within a specified volume alongside a grid that acts as a regular lattice. The grid elements, or cells, are used to compute the gravitational evolution of particles. Unlike particle–particle codes, which directly compute interactions between individual pairs of particles, PM codes are more computationally efficient, as they treat the gravitational force as an external field. However, due to the limited resolution of the grid, this method is accurate primarily on scales larger than the grid spacing. The main advantage of this approach lies in its physical analogy to a fluid, allowing for a more tractable description of large-scale gravitational dynamics. It is further enhanced by adaptive mesh refinement, which locally increases the grid resolution in regions of high particle density, thereby improving accuracy through the use of finer grids where needed.

4.1.2 Construction of halo catalogues

To construct halo catalogues, all snapshots of the simulation were used. A snapshot refers to a simulated box at fixed cosmic time that contains all particles (and grid cells) within it. The initial conditions are assumed to be Gaussian, as supported by observations of the CMB (Akrami et al., 2020), with the density field defined by the linear power spectrum $P(k, z)$, with the initial redshift set to approximately $z \simeq 50$. This power spectrum can be computed using a Boltzmann solver such as CAMB (Lewis et al., 2000). The simulation starts at $z \simeq 50$. The initial box consists of a periodic, homogeneous, and isotropic particle distribution, with each particle placed at the centre of a cell. The power spectrum is then used to perturb the distribution, yielding the true initial particle positions and velocities corresponding to the initial snapshot. The initial redshift must be chosen carefully to satisfy some constraints. First, it cannot be too high, since the simulation neglects radiation density fluctuations: as a result, the simulation cannot start during the radiation-dominated era. Additionally, starting from excessively high redshifts would lead to the accumulation of numerical errors, which could overwhelm the weak initial density perturbations.

Once the initial conditions are set, the subsequent evolution of the system can be followed using Lagrangian quantities. As described in Teyssier (2002), the main steps for evolving particles in PM codes are as follows. From the particle distribution, the mass density is interpolated onto the grid, in this case using the Triangular Shaped Cloud method, which produces a smoother and more regular density field. The gravitational potential is then computed by solving the Poisson equation, Eq. (1.64), and its gradient is calculated to determine the gravitational acceleration. Acceleration is used to update the particle velocities for the subsequent time step, and the new particle positions are obtained accordingly.

Outputs

During the simulation run, two types of output are produced: snapshots and light cones. As said, a snapshot provides a static view of the simulation box at a fixed cosmic time, with no reference to an observer. In contrast, light cones are constructed from the perspective of a virtual observer, who perceives the past light cone: the greater the distance to the particles, the higher their corresponding redshift.

There are essentially two main approaches to building light cones. The first approach consists of generating the light cone once the simulation has been completed. In this case, the light cone is built as a concatenation of snapshots: at the boundary of a simulation box at a given redshift, the same box at a higher redshift is appended. However, this procedure leads to replicas: the same photon trajectory may intersect identical structures at different redshifts. Although interpolating quantities such as particle positions and velocities onto the light cone from each snapshot can enhance the resolution, it does not eliminate the problem. The second approach, which is the one adopted to generate the catalogues used in this thesis, is known as the onion-shell approach (Fosalba, 2008). In this case, the light cone is constructed on the fly, during the simulation. At each time step, a shell containing particles and cells, within a narrow redshift interval, is saved. The final light cone is then assembled by concatenating all such shells. The light cones generated may be either narrow- or full-sky. Here, the full-sky configuration was used: the distribution of particles is recorded in all directions around the observer, up to a redshift corresponding to half the box length. Once the density field was obtained, the matter power spectrum was computed using POWERGRID (Prunet et al., 2008). Our primary interest lies in the power spectrum at wave numbers $k > 0.3h\,\mathrm{Mpc}^{-1}$, a regime in which relativistic corrections become important, significantly shaping the spectral features alongside nonlinear effects. In Figure 4.1 the comparison between the linear power spectrum and that of the simulation is shown.

At low k, the discrepancy arises primarily due to cosmic variance, resulting from the fact that we are working with a single realisation of a randomly sampled particle distribution. This inherently limits the statistical power of the analyses. Drawing robust conclusions on these scales would require simulations with significantly larger volumes, such as e.g. DEUS-FUR (Alimi et al., 2012). At higher values of k, two main effects, both associated with the finite resolution of the data, modify the shape of the power spectrum. Firstly, shot noise introduces a spurious contribution with an amplitude of $(L/N)^3$, where L is the box size and N the number of cells in the grid per dimension. Secondly, the Nyquist frequency, defined as $k_{\mathrm{Nyq}} = \pi(N/L)$, sets a fundamental upper limit on the recoverable Fourier modes. In practice, to ensure numerical reliability and avoid aliasing artefacts (Colombi et al., 2009), the analysis is typically restricted to modes that satisfy $k \lesssim k_{\mathrm{Nyq}}/2$. Lastly, structure identification is performed using

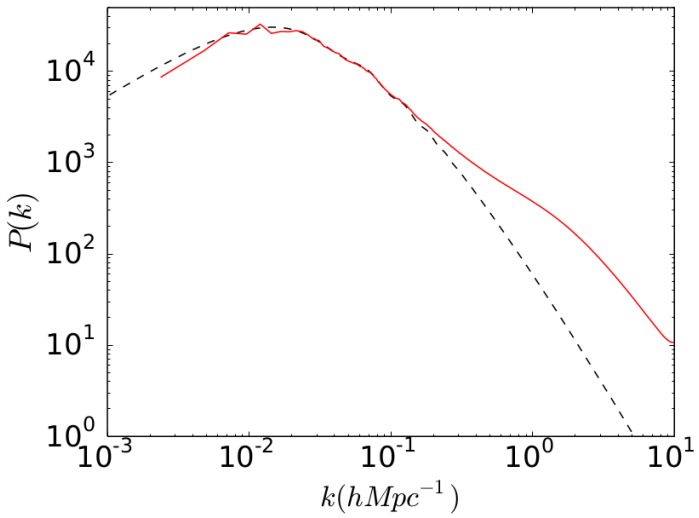

Figure 4.1: Dashed black line: linear power spectrum. Solid red line: linear power spectrum computed from the snapshot. Both are at $z = 0$ (Breton et al., 2019).

the FoF algorithm. This method defines a linking length b, usually set at a fraction (commonly 0.2) of the mean separation between particles. Starting from a given particle, neighbouring particles within this linking length are recursively connected to form a percolating structure. Groups containing more than 100 particles are then identified as dark matter haloes.

4.1.3 Ray-tracing with Magrathea-Pathfinder

To connect sources and the observer, ray-tracing techniques are employed. Several approaches exist for implementing ray tracing in simulations; in this case, the method consists of directly solving the geodesic equations. The simulation used employs Magrathea-Pathfinder, a hybrid Message Passing Interface/pthreads C++11 code. The strength and long-term potential of this approach lie in the fact that, while traditional ray-tracing codes are typically limited to studies of gravitational lensing, Magrathea-Pathfinder provides a unified framework that includes both gravitational lensing and redshift perturbations and, in particular, it implements first-order corrections in metric perturbations. This capability enables, in particular, detailed investigations of relativistic effects in galaxy clustering analyses. A detailed description of the code is provided in Breton and Reverdy (2022); here we summarise the key concepts.

This algorithm integrates the geodesic equations directly,

$$\frac{\mathrm{d}^2 x^\mu}{\mathrm{d}\lambda^2} = -2\mathcal{H}\frac{\mathrm{d}\eta}{\mathrm{d}\lambda}\frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda} - \frac{2}{c^2}\frac{\mathrm{d}\Phi}{\mathrm{d}\lambda}\frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda} + 2\frac{\partial\Phi}{\partial x^\mu}\left(\frac{\mathrm{d}\eta}{\mathrm{d}\lambda}\right)^2, \tag{4.1}$$

using a standard 4^{th}-order Runge–Kutta scheme. In the context of this simulation, ray tracing is performed as a post-processing step following the N-body evolution. The core of the Magrathea-Pathfinder library lies the construction of an N-dimensional tree data structure. This refers to a spatial decomposition technique (such as an octree in 3D), whereby the simulation volume is recursively subdivided into smaller cells or nodes. This hierarchical structure enables efficient localisation of a given point (e.g. a photon's position) within the simulation domain and facilitates the identification of neighbouring cells.

Once the photon's location is determined within the tree, the algorithm searches for surrounding cells up to a given level of refinement, in order to interpolate local Eulerian quantities (such as gravitational potentials and velocities) at that position. The interpolation proceeds as follows: the algorithm first attempts to identify the finest-level cell containing the photon. If eight adjacent cells of the same refinement level are found, interpolation is carried out using their values. If not, the algorithm ascends the tree to a coarser refinement level and repeats the search. This process continues until a valid set of neighbouring cells is located. If no such set is found even at the coarsest level, the integration of the geodesic is aborted, as there is insufficient data to ensure reliable interpolation.

4.1.4 Producing observables

The primary aim is to connect the source to the observer located at the origin of the light cone (i.e., to determine the geodesic linking the two). This is achieved by requiring the angular separation between the source and the photon, at the same comoving radius, to be zero.

To this end, a root-finding algorithm is employed. Consider an observer positioned at the origin of the light cone, observing a photon that has been deflected by gravitational lensing. The observer perceives the source at an apparent angle θ, whereas the true angle, $\beta = (\beta_1, \beta_2)$, corresponds to the comoving coordinates of the source. The tentative photon at the n-th iteration is denoted by ζ_n. Initially, the photon is emitted in the direction β. If the photon fails to intercept the source, the initial angle is adjusted and the process repeated iteratively. This continues until the photon successfully reaches the source, which, in the example illustrated in Figure 4.2, is assumed to occur at the second iteration. The initial angle corresponding to ζ_2 is θ, interpreted as the apparent angle under which the source is seen.

Angles are defined in spherical coordinates. The corresponding normalised comoving

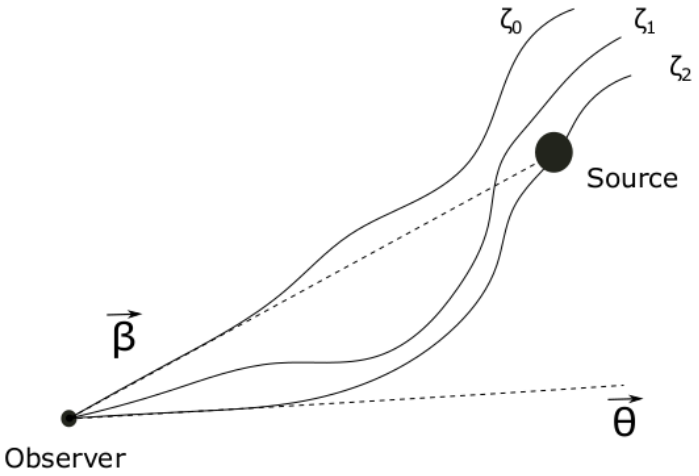

Figure 4.2: Geodesic-finder algorithm. See text for the detailed description (Breton et al., 2019).

coordinates are computed as:

$$\begin{cases} x = \cos(\beta_1)\sin(\beta_2) \\ y = \sin(\beta_1)\sin(\beta_2) \\ z = \cos(\beta_2). \end{cases} \tag{4.2}$$

The same expressions can be used to convert apparent angles into apparent normalised coordinates, by replacing β_i by θ_i.

4.1.5 Redshift terms

The perturbed redshifts z_i are defined as follows:

$$1 + z_0 = \frac{a_0}{a}, \tag{4.3}$$

$$1 + z_1 = \frac{a_0}{a}\left[1 - \frac{(\Phi_s - \Phi_o)}{c^2}\right], \tag{4.4}$$

$$1 + z_2 = \frac{a_0}{a}\left[1 - \frac{(\Phi_s - \Phi_o)}{c^2} + \frac{\boldsymbol{v}\cdot\boldsymbol{n}}{c}\right], \tag{4.5}$$

$$1 + z_3 = \frac{a_0}{a}\left[1 - \frac{(\Phi_s - \Phi_o)}{c^2} + \frac{\boldsymbol{v}\cdot\boldsymbol{n}}{c} + \frac{1}{2}\left(\frac{v}{c}\right)^2\right], \tag{4.6}$$

$$1 + z_4 = \frac{a_0}{a}\left[1 - \frac{(\Phi_s - \Phi_o)}{c^2} + \frac{\boldsymbol{v}\cdot\boldsymbol{n}}{c} + \frac{1}{2}\left(\frac{v}{c}\right)^2 - \frac{2}{c^2}\int_{\eta_s}^{\eta_0} \mathrm{d}\eta\, \dot{\Phi}\right], \tag{4.7}$$

$$1 + z_5 = \frac{(g_{\mu\nu} k^\mu k^\nu)_s}{(g_{\mu\nu} k^\mu k^\nu)_o}. \tag{4.8}$$

Here, the subscripts o and s denote quantities evaluated at the observer and the source position, respectively. The velocity appears without a subscript since the observer velocity is set to zero: our focus is on the dipole that arises around the source. Individual contributions to the total redshift are as follows:

- z_0 corresponds to the cosmological redshift arising from the scale factor;
- z_1 accounts for the gravitational potential;
- z_2 incorporates the relativistic Doppler shift due to the line-of-sight peculiar velocity;
- z_3 introduces the transverse Doppler effect, which is relevant solely on small scales;
- z_4 includes the ISW contribution.

The last term, z_5, is directly inferred from GR. However, in practice, it is almost indistinguishable from z_4.

4.2 CosmoBolognaLib

The CosmoBolognaLib (CBL) has formed the backbone of this thesis. These are *free software* C++ libraries, specifically designed in the context of LSS cosmological computations, a field in which numerical and computational techniques have become not merely advantageous, but utterly indispensable, the "deus ex machina" of contemporary cosmology. The libraries offer a versatile and robust framework for managing both observational and simulated extragalactic catalogues, as well as for constructing the corresponding random catalogues. This infrastructure enables the computation of 1-, 2-, and 3-point statistical estimators in configuration space, thereby facilitating the derivation of constraints on cosmological parameters and offering a rigorous testbed for cosmological models (Marulli et al., 2016). A hallmark feature of the CBL is its bespoke parallel chain-mesh algorithm, purpose-built to optimise the performance of pair and triplet counting across large data sets. In addition, the libraries provide a comprehensive suite of tools for quantifying statistical uncertainties and performing likelihood-based inference. Over the years, their capabilities have been further extended to enable the study of cosmic voids as well (Ronconi and Marulli, 2017). Entirely developed in C++, CBL retains the flexibility to seamlessly integrate into high-level scripting environments such as Python, thereby ensuring both performance and accessibility.

The original contribution of this work is twofold. On the one hand, a pipeline has been developed for computing the 2-point cross-correlation function between two catalogues, specifically tailored for configurations involving two distinct random catalogues, and for the extraction of all multipoles, both even and odd, up to the hexadecapole. On the other hand, dedicated routines have been implemented to model the dipole, building upon the theoretical frameworks discussed in the preceding chapters.

4.3 Computing resources

All the numerical analyses and simulations presented in this thesis were carried out on two machines running the Ubuntu Linux operating system, which provided a stable and efficient environment for scientific computing. The first machine is equipped with 4 logical CPUs (2 physical cores with simultaneous multithreading), with a base frequency ranging from 1.2 GHz up to 2.4 GHz, and support for frequency boosting. The storage device is a 270 GB hard disk drive (HDD). This machine has 16 GB of RAM, enabling the handling of large data sets and memory-intensive numerical computations, such as Monte Carlo integration and estimation of the 2PCF. Additionally, it features a 64 GB swap partition, which serves as additional virtual memory and ensures stable performance during particularly demanding tasks. Although not a high-performance computing platform, this hardware configuration proved sufficient for part of the work presented in this thesis, and was primarily used for analyses involving catalogues with fewer objects.

The second machine used for the analysis is equipped with a 500 GB solid state drive (SSD) storage device, and a 6-core processor (with 12 logical processors via hyperthreading) with a base frequency of 3.50 GHz and a maximum boost frequency of 4.6 GHz, and a swap space of 128 GB. This second machine enabled the analysis of the larger catalogues, allowing for random samples up to 5 times the size of the data catalogue, and more than ten the size of the catalogues for smaller data sets, albeit with very long computation times in the first case. Further improvements will be achievable in the future through the use of supercomputers.

4.3.1 Open Multi-Processing

For pair counting, Open Multi-Processing (OpenMP) is employed. OpenMP is an application programming interface (API) that facilitates shared-memory parallel programming across multiple platforms in C, C++, and Fortran (Dagum, L. and Menon R., 1998). It enables developers to parallelise existing sequential code relatively easily using compiler directives, thereby optimising the utilisation of multi-core processors. OpenMP manages the creation and coordination of multiple threads that execute portions of

the code concurrently, following a fork-join model: the master thread forks a team of worker threads to perform parallel tasks and subsequently joins them once the work is complete. It is widely used in cosmology for estimating cross-correlation functions (Kerscher, 2022), and is currently also employed within the CBL.

The function responsible for counting pairs of objects, such as galaxies or haloes, operates between a primary catalogue and a second catalogue accessed via a chain-mesh structure. This procedure is typically employed in cosmology for estimating 2PCF. The function begins by initialising with the catalogues, a pair object to accumulate counts, and control flags, whilst starting a timer to measure performance. It then retrieves the number of objects in the first catalogue and obtains a reference to the second catalogue from the chain-mesh. Subsequently, the function enters a parallel region created with OpenMP, where the number of threads is set to the maximum available on the system. Each thread acquires a unique identifier and instantiates a private pair object to accumulate counts locally, thereby avoiding concurrent writes to shared memory that could result in data corruption. Within this parallel region, a for-loop is dynamically distributed among the threads. Each thread processes a subset of objects from the first catalogue. For every object processed, the function identifies nearby objects in the second catalogue by querying the chain-mesh, which efficiently restricts the search to spatially proximate candidates. The function then iterates over these neighbouring objects, updating the thread-local pair counts by invoking the appropriate methods.

4.3.2 Computational time

Counting pairs is a fundamental operation in many fields, such as cosmology, where it is often necessary to compute statistics like the 2PCF. The naïve approach to pair counting involves comparing each object in one catalogue with every other object in the same catalogue, or in a different one, resulting in a computational complexity that scales as $\mathcal{O}(N_1 N_2)$, where N_1 and N_2 denote the number of objects in the same catalogue or in different catalogues. This scaling applies to the data–data (D_1D_2), data–random (D_1R_2), random-data (R_1D_2) and random–random (R_1R_2) pair counts, required by estimators such as that of Landy–Szalay (Landy and Szalay, 1993), for auto-correlations or when cross-correlating distinct catalogues, Eq. (2.32). This quadratic scaling rapidly becomes prohibitive for large data sets and their typically even larger associated random catalogues, as the number of pairwise comparisons increases dramatically with sample size. To mitigate this computational challenge, spatial indexing and partitioning techniques are commonly employed. In particular, efficient evaluation of the 2PCF requires skipping pairs separated by more than a chosen maximum distance r_{Max}, since $\xi(r)$ is nearly zero at large scales. One widely used approach is the chain-mesh (or grid-based) method,

which divides the spatial domain into cells of a chosen size. Each object is assigned to a cell, and pair searches are restricted to objects in the same cell or in neighbouring cells within a search radius of $r_{\rm Max}$. This procedure significantly reduces the number of unnecessary distance calculations between distant pairs, whose separation is beyond the scale of interest. The algorithm, thus, reduces the computational cost, which becomes roughly proportional to the number of pairs with $r \leq r_{\rm Max}$. For instance, in the cross-correlation function, the total cost C scales with the number of galaxy pairs as follows:

$$C = f(N_{\rm D_1} N_{\rm D_2} + N_{\rm D_1} N_{\rm R_2} + N_{\rm R_1} N_{\rm D_2} + N_{\rm R_1} N_{\rm R_2}), \tag{4.9}$$

where f is the fraction of pairs with $r \leq r_{\rm Max}$. Since random catalogues are typically much denser than data catalogues, the cost is usually dominated by the $R_1 R_2$ term. Defining the random-to-data ratio as

$$M_r = \frac{N_{\rm R_1}}{N_{\rm D_1}} = \frac{N_{\rm R_2}}{N_{\rm D_2}}, \tag{4.10}$$

the cost expression simplifies to:

$$C = f\, N_{\rm D_1} N_{\rm D_2} (1 + 2M_r + M_r^2). \tag{4.11}$$

Another factor that impacts computational performance is the cell size used in the chain-mesh algorithm. In the implementation of the chain-mesh within the CBL, the cell size is defined as

$$\text{cell\_size} = f \times r_{\rm Max}, \tag{4.12}$$

where f is a dimensionless parameter typically ranges between 0.1 and 0.2. In general, the cell size should be smaller than the minimum scale of interest $r_{\rm Min}$ to ensure that all pairs with separation greater than or equal to $r_{\rm Min}$ are correctly counted. However, in this work, the factor has been optimised and allowed to exceed the value of 0.2. This choice is justified by the following considerations:

- Inclusion of adjacent cells: the chain-mesh algorithm searches not only within the cell containing a given object but also within neighbouring cells. This overlapping search mitigates the risk of missing pairs, even if the cell size is somewhat larger than the minimum scale (Marulli et al., 2016);

- Numerical convergence tests: extensive tests have been performed by varying the value of f and verifying that the measured 2PCF remains stable within the statistical uncertainties. This empirical validation indicates that the increase in cell size does not significantly bias the pair counts;

- Computational efficiency: increasing the cell size reduces the total number of cells and, thus, the overhead associated with neighbour searches, leading to faster computations without significantly compromising accuracy. Although larger cells contain more objects on average, resulting in a greater number of operations within the innermost loops, this cost increases linearly, whereas the number of cells to be examined decreases cubically. As a result, a reduction of approximately 25 to 35% in the computation time of pair counting was achieved. Although this approach proved effective for the present analysis, it is not claimed to be universally optimal, and further testing will be required to assess its robustness in the context of CBL.

Further improvements could be in the direction of splitting random catalogues into a number of subcatalogue of the same size as the data catalgues and excluding pairs across different subcatalogues when computing random-random pairs. This method has been shown to outperform the use of diluted random catalogue and could represent a further direction to improve computational efficiency. In fact, it reduces the computation time by a factor of more than ten for a random catalogue fifty times larger than the data catalogues, without affecting the estimator variance, as shown by Keihänen et al. (2019).

4.4 Codes and data validation

The RayGalGroupSim catalogues have already been extensively validated by Breton et al. (2019) and rasera2021, as have the CBL, e.g. Marulli et al. (2016, 2017a), to name but a few. Nevertheless, as a preparatory step for the subsequent analysis, and in order to verify the integrity of catalogues, the robustness of random samples, and the consistency of pre-existing codes, we performed a series of preliminary tests. These included the measurement of the 2PCF, with particular attention paid to ensuring that the even multipole moments behaved as expected, both in real and redshift space. In Figure 4.3, a representative subsample of the complete data set is shown, identified by its (RA, Dec, z) coordinates.

The corresponding random catalogue was generated to match the properties of the data catalogue, resulting in a complete overlap between the two distributions.

In Figure 4.4, the monopole of the 2PCF computed for this data set is shown in real space. The monopole decreases with increasing scales, approaching zero on large scales. This indicates that haloes cluster more strongly than in a homogeneous distribution, particularly on small scales. On large scales, instead, the Universe becomes increasingly homogeneous, which explains why the monopole tends to vanish. At around $100\,h^{-1}$Mpc, a distinct peak is visible in the monopole: this is the baryon acoustic peak, the configuration-space manifestation of the baryon acoustic oscillations (BAO) observed in Fourier space. These oscillations are relics of the recombination era and serve as

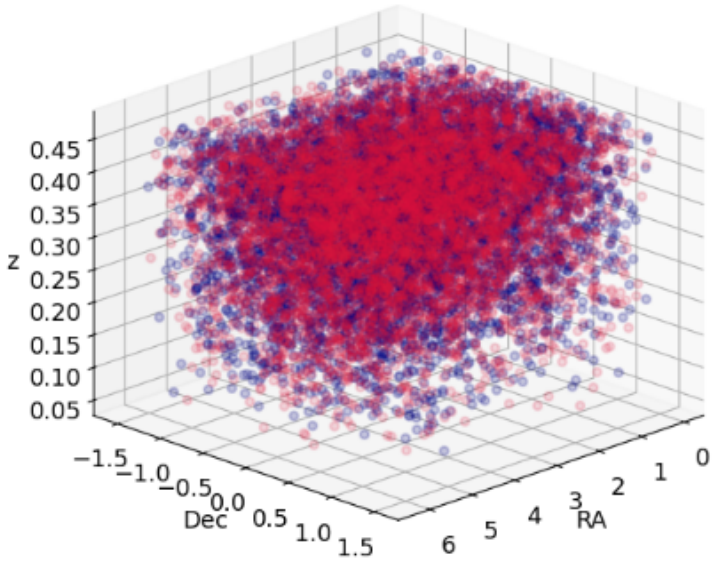

Figure 4.3: A subsample of the catalogue (blue points) and its corresponding random sample (red points), with each object identified by its right ascension, declination, and redshift.

a standard ruler for precise distance measurements, thus enabling a reconstruction of the geometry of the Universe. Furthermore, they provide valuable information on the distribution of matter and the nature of dark energy (Eisenstein et al., 2005; Bassett and Hlozek, 2009; Adam et al., 2025). The superimposed best-fitting model corresponds to the matter correlation function, scaled by the square of the linear bias factor, i.e. is given by:

$$\xi(r) = \xi_{\mathrm{M}}(r)\, b^2. \tag{4.13}$$

The halo bias was computed at a volume-averaged redshift $z = 0.341$, using the model proposed by Castro et al. (2024), which has been recently implemented in the CBL. Error estimation is described in Section 4.5.1.

This model clearly provides a poorer fit on scales smaller than around $25\,h^{-1}$Mpc, as such scales start to be more strongly affected by nonlinear effects. Figure 4.5 shows the first three even multipoles of the redshift-space 2PCF, computed for the full data set. In the linear regime and under the distant observer approximation, the redshift-space 2PCF can be written as:

$$\xi(s) = \xi_0(s)\mathcal{L}_0(\mu) + \xi_2(s)\mathcal{L}_2(\mu) + \xi_4(s)\mathcal{L}_4(\mu), \tag{4.14}$$

where μ is the cosine of the angle between the line of sight and the separation vector s

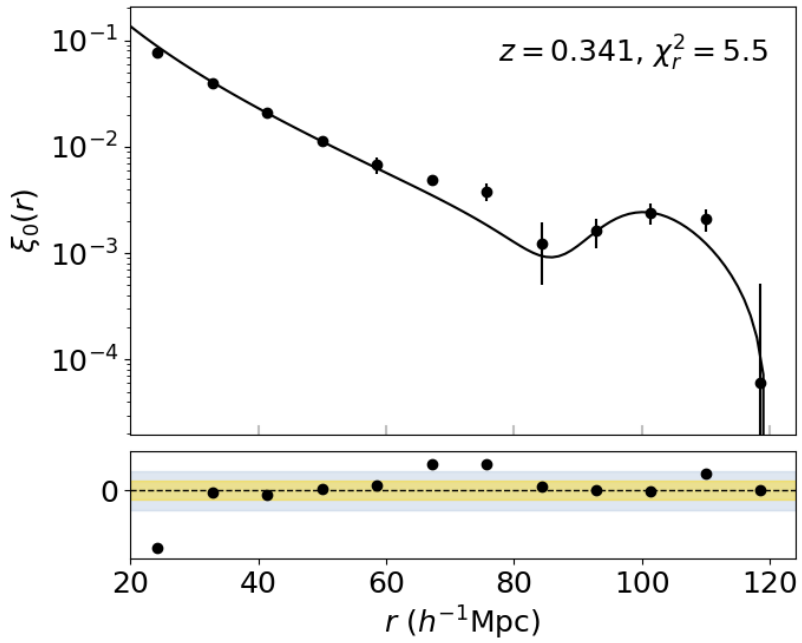

Figure 4.4: Real space monopole of the 2PCF, computed considering all haloes (black circles), with errors estimated using the jackknife resampling method. The fitting model is shown as a black solid line. The lower panel displays the residuals calculated as the difference between data and model, normalised by the error. The yellow and grey bands represent the 1σ and 2σ confidence intervals, respectively.

and $\mathcal{L}_l$ are the Legendre polynomials. The multipoles are given by:

$$\xi_0(s) = \left(1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2\right)\xi(r), \tag{4.15}$$

$$\xi_2(s) = \left(\frac{4}{3}\beta + \frac{4}{7}\beta^2\right)\left[\xi(r) - \overline{\xi}(r)\right], \tag{4.16}$$

$$\xi_2(s) = \frac{8}{35}\beta^2\left[\xi(r) + \frac{5}{2}\overline{\xi}(r) - \frac{7}{2}\overline{\overline{\xi}}(r)\right], \tag{4.17}$$

where $\beta = f/b$ is the linear RSD parameter. The function $\xi(r)$ is the linear real-space correlation function (computed for tracers), and the barred functions are defined as:

$$\overline{\xi}(r) = \frac{3}{r^3}\int dr'\xi(r')r'^2, \tag{4.18}$$

$$\overline{\overline{\xi}}(r) = \frac{5}{r^5}\int dr'\xi(r')r'^4. \tag{4.19}$$

To account for nonlinear effects and small-scale motions (the so-called Fingers of God), an empirical approach known as the dispersion model is commonly adopted

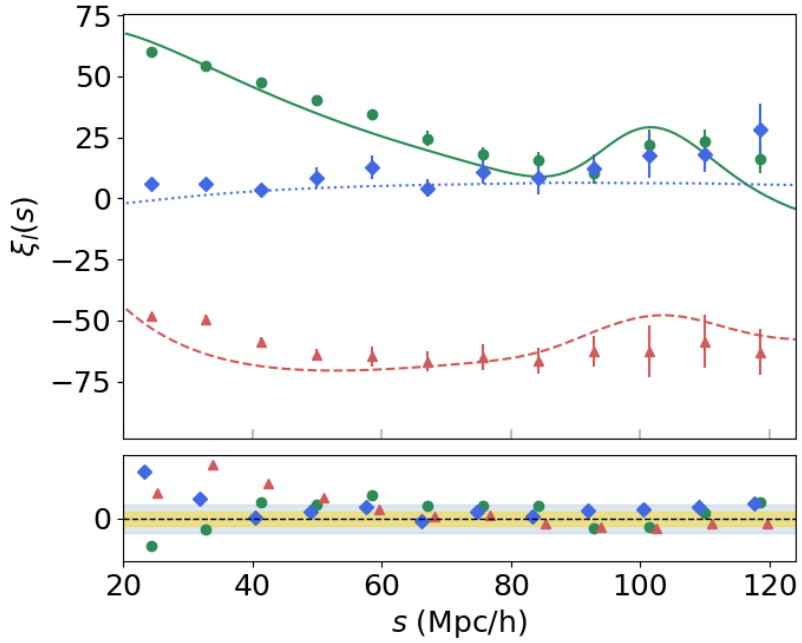

Figure 4.5: Monopole (green circles), quadrupole (red triangles) and hexadecapole (blue diamonds) of the 2PCF in redshift space, computed considering all haloes. Errors are estimated with the jackknife resampling method. Each multipole is fitted with the dispersion model (green solid line for the monopole, red dashed line for the quadrupole and blue dotted line for the hexadecapole). The lower panel displays the residuals calculated as the difference between data and model, normalised by the error. The yellow and grey bands represent the 1σ and 2σ confidence intervals, respectively.

(Peebles, 1980; Peacock and Dodds, 1996). This model has been widely used in the literature to describe RSD, especially in the quasilinear regime (Marulli et al., 2012, 2017b; Nadathur and Percival, 2019). It appears to perform well also in the cases considered here (Figure 4.4 and Figure 4.6).

Nevertheless, several more refined models have been developed to improve the agreement with observations, particularly on intermediate and nonlinear scales. Notable examples include the streaming model with scale-dependent velocity dispersion (Scoccimarro, 2004) and the TNS model, which incorporate higher-order corrections from perturbation theory (Taruya et al., 2010).

Once we verified everything was correct, we proceeded with the 2-point cross-correlation function. For this purpose, the catalogue was partitioned into six logarithmic mass bins, as detailed in Table 4.1. We recall that a halo is defined as a structure containing at least 100 particles, each of mass $1.88 \times 10^{10}\, h^{-1}\mathrm{M}_{\odot}$.

Three further examples of the monopole for the 2PCF in redshift-space are displayed

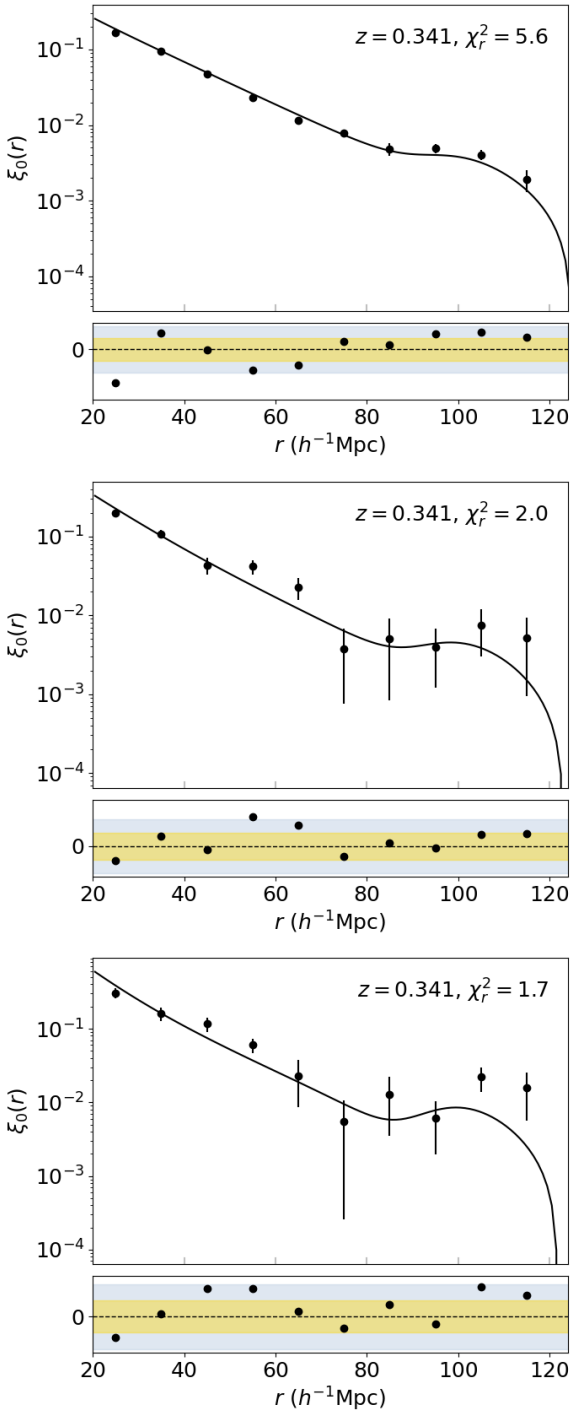

Figure 4.6: Redshift space monopole for the 2PCF of three different data sets: H_{800}, H_{1600} and H_{3200} (from top to bottom panel). Each panel below the plots shows the residuals, calculated as the difference between data and model, normalised by the error. The yellow and grey bands represent the 1σ and 2σ confidence intervals, respectively.

Name	N haloes	n particles per halo	$M\,(\mathrm{M}_\odot \mathrm{h}^{-1})$	bias
H_{100}	5.5×10^6	$100 - 200$	$1.88 \times 10^{12} - 3.76 \times 10^{12}$	1.12
H_{200}	3.4×10^6	$200 - 400$	$3.76 \times 10^{12} - 7.52 \times 10^{12}$	1.25
H_{400}	1.9×10^6	$400 - 800$	$7.52 \times 10^{12} - 1.50 \times 10^{13}$	1.45
H_{800}	1.0×10^6	$800 - 1600$	$1.504 \times 10^{13} - 3.008 \times 10^{13}$	1.71
H_{1600}	4.0×10^5	$1600 - 3200$	$3.008 \times 10^{13} - 6.016 \times 10^{13}$	2.09
H_{3200}	2.0×10^5	$3200 - 6400$	$6.016 \times 10^{13} - 1.203 \times 10^{14}$	2.63

Table 4.1: Sub-sample characteristics: name, number of haloes, number of particles per halo, range of masses in each sub catalogue and bias.

in Figure 4.6. By comparing them with the monopole computed in real space (Figure 4.4), we observe that the redshift space monopole exhibits a broader and flatter shape, as also noted by previous studies, and tends to remain approximately 10–20% higher than its real space counterpart (Valageas and Clerc, 2012). These not only served to validate the codes and models employed, but also provided the basis for a subsequent a-posteriori analysis using MCMC fitting, which will be presented and discussed below.

4.5 The 2-point cross-correlation function

As introduced in the beginning of this section, part of this work involved developing a code for the subsequent analysis of simulated data from the RayGalGroupSims. Firstly, it was necessary to extend the existing code for computing the cross-correlation function to allow its use with two distinct random catalogues. This was implemented within the class `cbl::measure::twopt::TwoPointCorrelationCross`, employing the Landy–Szalay estimator, Eq. (2.32). Subsequently, a dedicated class, `cbl::measure::twopt::TwoPointCorrelationCross_multipoles`, was created to enable the extraction of multipoles, including odd multipoles. After constructing the catalogues in (RA, Dec, z) and generating the corresponding random catalogues (one for each data catalogue), and after instantiating the object responsible for measuring the multipoles of the cross-correlation, the method `measure` allows selection between two different uncertainty estimation techniques, Poisson and jackknife, with provisions for future implementation of the bootstrap method. These methods will be described in detail in the next section. If the position of each object of a pair is given by $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ we choose $\boldsymbol{x} = (\boldsymbol{x}_1 + \boldsymbol{x}_2)/2$, $\boldsymbol{r} = \boldsymbol{x}_2 - \boldsymbol{x}_1$ and $\mu = \hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{r}}$, where a hat denotes a unit vector. The angle defined this way is symmetric under exchange of pairs (Gaztanaga et al., 2017). The pair counts in each radial and angular bin are stored and subsequently used to compute the multipoles via Eqs. (2.35) to (2.39).

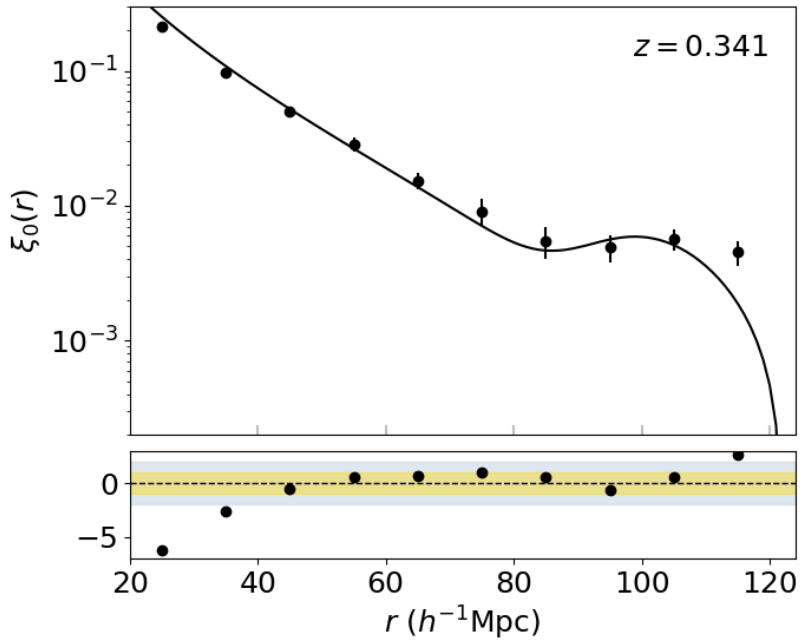

Figure 4.7: Monopole of the 2-point cross-correlation function between H_{3200} and H_{1600} in real space. The model is represented by a black solid line. Errors were estimated with the jackknife resampling method. Lower panels display the residuals calculated as the difference between data and model, normalized by the error. The yellow and grey bands represent the 1σ and 2σ confidence intervals, respectively.

In Figures 4.7 and 4.8, as in the case of the 2PCF, we show examples of the monopole in real space and of the even multipoles in redshift space, for representative cases of cross-correlations between two distinct catalogues. The agreement between the measurements and theoretical predictions for the even multipoles is generally good, particularly at intermediate scales. However, at small scales the discrepancy becomes significant, as the analysis enters the quasi-linear regime, where more refined models and appropriate corrections are required. On large scales, the monopole also appears to remain higher than in the corresponding 2PCF shown in Figure 4.4, indicating that the more massive tracer population contributes to an enhanced clustering signal on large scales.

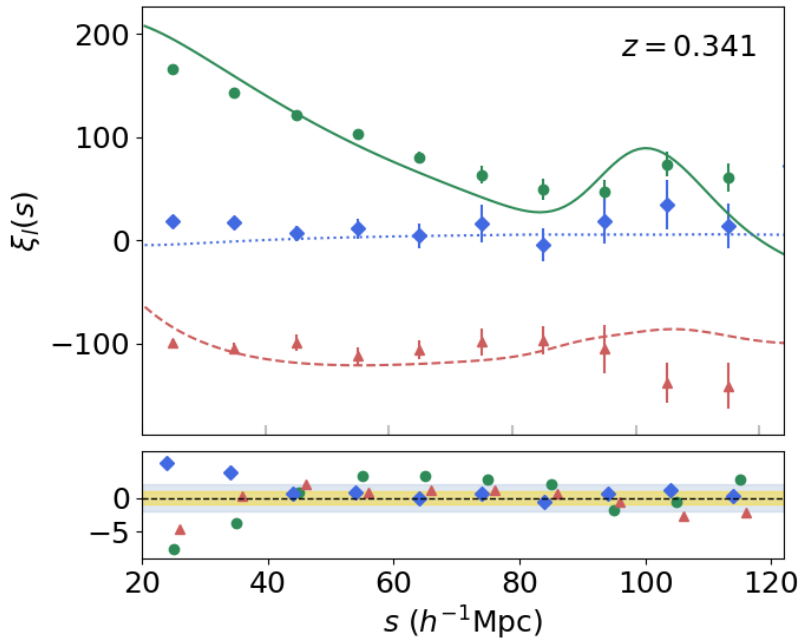

Figure 4.8: Monopole (green circles), quadrupole (red triangles) and hexadecapole (blue diamonds) of the 2-point cross-correlation function between H_{3200} and H_{1600} in redshift space. Each multipole is fitted with a dispersion model (green solid line for the monopole, red dashed line for the quadrupole and blue dotted line for the hexadecapole). Errors were estimated with the jackknife resampling method. Lower panels display the residuals calculated as the difference between data and model, normalized by the error. The yellow and grey bands represent the 1σ and 2σ confidence intervals, respectively.

4.5.1 Evaluation of measurement uncertainties

In the CBL framework, there are three standard approaches for estimating statistical uncertainties associated with 2-point clustering measurements, particularly in the context of the galaxy auto-correlation function. These methods are:

- Poisson errors;
- The jackknife resampling technique;
- The bootstrap method.

The Poisson error is the most straightforward statistical estimate that one can adopt when dealing with discrete counting processes. If we register a number of counts N, the variance is simply given by:

$$\sigma^2(N) = \langle (N - \overline{N})^2 \rangle = \overline{N}. \tag{4.20}$$

The standard deviations associated with the data-data (DD) and random-random (RR) pair counts are typically taken to be $\sqrt{DD}$ and $\sqrt{RR}$, respectively. However, Poisson statistics provide a reasonable approximation to the uncertainty only under the assumption that the events (or galaxy pairs) are independent and uncorrelated. In fact, this assumption breaks down in more realistic cosmological scenarios, particularly when correlated structures and systematic effects come into play. Sample variance constitutes a major source of non-Poissonian uncertainty, reflecting the intrinsic limitation inherent to any observation of large-scale density fluctuations, as we are confined to observing only a finite region of the Universe. This effect is particularly relevant in the context of deep surveys (Moster et al., 2011) and highly clustered tracers (Somerville et al., 2004), which, despite probing large comoving volumes in redshift space, typically span only a limited area of the sky. As a result, the variance among different realisations, driven by the underlying LSS, can be substantial, and is not captured by simple Poisson error estimates. For this reason, this type of error has not been adopted in this work, despite having been implemented for 2-point cross-correlation measurements.

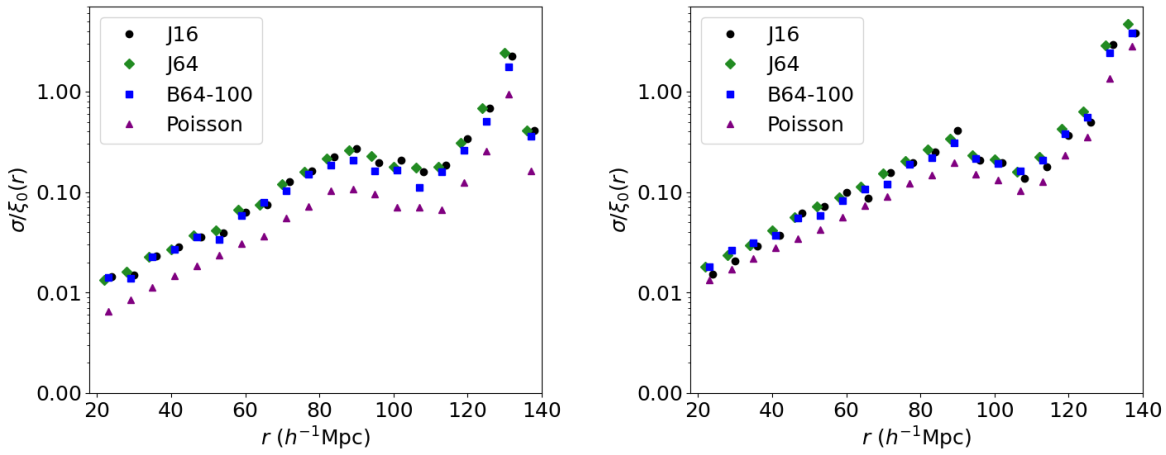

Figure 4.9: Comparison between different error estimates for the monopole of the 2PCF. Left panel: 2PCF of H_{800}*, with the random catalogue containing 3 times more objects than the data catalogue. Right panel: 2PCF of* H_{3200}*, with the random catalogue containing 10 times more objects than the data catalogue. Black circles and green diamonds represent the jackknife estimates with 16 and 64 subregions, respectively. Blue squares show errors estimated using the bootstrap resampling method with 100 mocks. Purple triangles correspond to Poisson errors.*

Both the jackknife and bootstrap methods are based on resampling techniques. In the jackknife approach, the data set is divided into N subvolumes. Consider the cross-correlation function between two catalogues, i and j. For each of these, the 2-point statistic, specifically the cross-correlation function ξ_{ij}, is measured by systematically

excluding one subvolume at a time. At a given redshift $\overline{z}$, the covariance matrix is defined as:

$$\mathrm{Cov}_{ij}^{\mathrm{JK}} = \frac{N-1}{N} \sum_{n=1}^{N} \left[\xi_{ij}^{n}(\overline{z}, r_i) - \overline{\xi}_{ij}(\overline{z}, r_i)\right] \left[\xi_{ij}^{n}(\overline{z}, r_j) - \overline{\xi}_{ij}(\overline{z}, r_j)\right], \tag{4.21}$$

where, in case the dipole moment is considered, ξ_{ij} is replaced by ξ_1, as defined in Eq. (2.36). This error estimation method has been validated on scales greater than $25\,h^{-1}$Mpc for 2-point clustering statistics; however, it is not well suited for the highly nonlinear regime, for which a Monte Carlo approach would be more appropriate (Norberg et al., 2009).

In contrast to the jackknife, the bootstrap method generates a large number of synthetic samples by randomly selecting subvolumes with replacement. The target statistic is recalculated for each realisation, thereby constructing an empirical distribution from which uncertainties can be estimated. Although the bootstrap appears particularly well-suited to nonlinear regimes, where it can capture non-Gaussian features, it has been shown that it tends to systematically overestimate uncertainties across all scales (Arnalte-Mur and Norberg, 2014). For what concerns our analyses, Figure 4.9 shows that, while the Poisson error tends to be lower than the others, especially on large scales, there are no significant differences between the jackknife and bootstrap methods. For this reason, the jackknife resampling technique was adopted and implemented in the present analysis of the 2PCF dipole, also because it is less computationally burdensome.

4.6 Modelling the dipole in the CosmoBolognaLib

As anticipated, the second part of this work required the implementation of the functions necessary to model the dipole on linear scales. This phase involved the development of a dedicated routine for setting the required parameters of the models. In order to compute these models, some preexisting functions had to be adapted to allow the evaluation of the integrals involved in the Eqs. (3.66) and (3.67). Additionally, specific routines were implemented for computing the necessary derivatives, namely, the derivatives of the linear growth rate f and the Hubble parameter $\mathcal{H}$ with respect to conformal time, as well as the derivatives of the bias parameters of the two populations and of the linear growth rate f with respect to comoving distance. With the aim of ensuring the highest possible generality, each contribution to the model is kept separable.

In the previous chapter, we analysed each of these terms (and their internal contributions) from a theoretical perspective, providing the rationale for why they cannot be neglected in dipole analyses. We also presented the equations of the models used to describe the dipole on linear scales.

Here we provide only a brief summary, intended to reframe the subsequent discussion on the various terms entering the models. In what follows, we neglect magnification bias, since we are working with a simulation of dark matter haloes. Moreover, in many practical applications, evolution bias is also often neglected because its contribution, entering primarily through relativistic Doppler terms, is typically subdominant on the scales of interest. This is especially true for galaxy samples whose comoving number density evolves slowly with redshift. Additionally, given the modelling uncertainties associated with galaxy evolution, omitting this term avoids introducing further systematic errors. Nevertheless, the general formalism has been introduced to provide a comprehensive framework, since this dark matter simulation (as well as other simulations) may be populated with galaxies in the future, and to preserve completeness as much as possible with a view to future surveys such as DESI or *Euclid*.

Beyond standard RSD contributions, certain additional terms, of geometric or relativistic origin, must be taken into account, as they induce an antisymmetry in the cross-correlation function, which in turn gives rise to a dipole moment. Modelling these contributions individually is not merely an academic exercise; rather, it is a preparatory step for the analysis of current and future surveys, which are expected to achieve the precision and accuracy, along with the large volume, required to disentangle and model the different effects.

In particular, among the relativistic contributions, in addition to the relativistic Doppler effect and the gravitational potential, one must account for gravitational lensing, the ISW effect, Shapiro time delay, and acceleration effects. Furthermore, some purely geometric contributions, such as the wide-angle effect and the light-cone effect, also play a role in shaping the dipole signal, alongside evolutionary effects.

4.6.1 Wide angle

Wide-angle corrections arise from the need to include additional terms beyond standard RSD for more accurate predictions, when relaxing the distant-observer approximation. This means that if the lines of sight to the two galaxies in a pair cannot be approximated as parallel and introduces an antisymmetric contribution to the 2-point cross-correlation, generating a dipole signal. The exact form depends on how the angle between galaxy pairs and the observer's line of sight is defined. Two common definitions exist: one treats the galaxies symmetrically, while the other centers on one galaxy, tipically the brighter (or the most massive one), to better captures relativistic effects like gravitational redshift (Gaztanaga et al., 2017). At leading order, the dipole signal is independent of the angle choice; differences only appear at higher orders and are negligible on the scales considered.

Dam and Bonvin (2025) demonstrated that wide-angle corrections to the even multipoles, compared to the distant-observer approximation, become significant at large separations, and that the monopole requires greater care than the quadrupole when modeling wide-angle effects. This is because the quadrupole is dominated by pairs more aligned with the line of sight, where wide-angle effects are smaller.

Regarding the wide-angle contribution to the dipole, Figure 4.10 presents the isolated wide-angle effect's contribution and explores how its shape varies with different cosmological and model parameters, namely the redshift z and the growth rate f. If the first population of galaxies is brighter than the second, the wide-angle effect induces a negative dipole contribution, typically of the order of 10^{-4}. This contribution increases in magnitude for larger values of $\Delta b = b_1 - b_2$, that is, for populations with more pronounced differences in luminosity or mass.

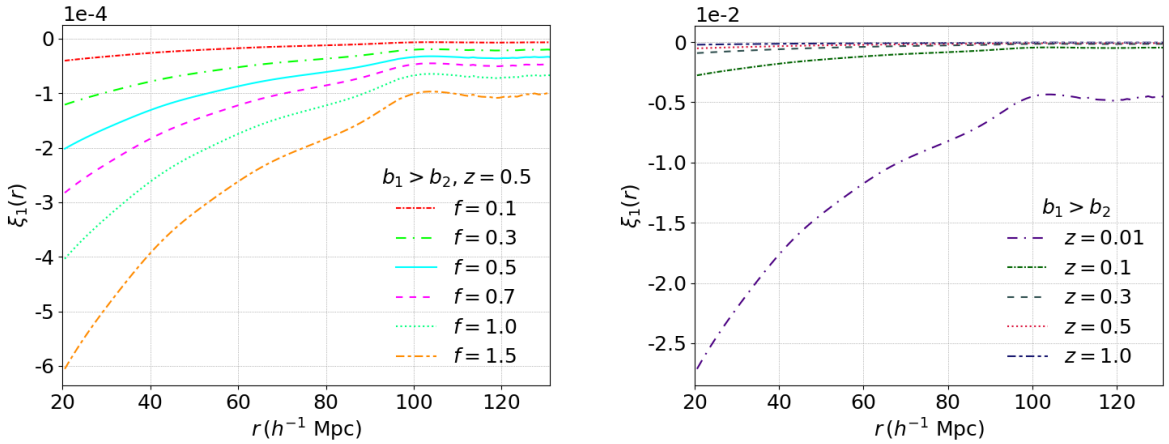

Figure 4.10: Wide-angle contribution to the dipole across different parameter regimes. Left panel: the biases of both populations and the redshift are kept constant; here, the variation lies solely in the linear growth rate. The value of f corresponding to ΛCDM *in this plot can be roughly represented by the curve at* 0.7*. Bottom panel: the model is shown for different values of redshift, with the galaxy biases fixed throughout. In this context, changes in redshift are inherently accompanied by changes in the growth rate, as predicted by the* ΛCDM *model.*

The growth rate f explicitly enters the expression for the contribution of the wide-angle to the dipole. As f increases at fixed redshift, so does the amplitude of the wide-angle term, in step with the faster growth of density perturbations and, consequently, of cosmic structures. In the context of modified gravity theories, deviations from GR can alter the growth rate f. Such modifications affect the peculiar velocities and clustering dynamics of galaxies, which in turn impact the wide-angle corrections to the dipole signal. Therefore, the amplitude and scale dependence of the wide-angle term in the

dipole could serve as sensitive probes to distinguish between different gravitational models. Unlike some other contributions to the dipole, where the BAO feature can be diminished or even absent, the geometric nature of the wide-angle term preserves the BAO peaks. This characteristic renders the wide-angle effect particularly significant when analysing the dipole at scales close to the BAO scale, even though a purely linear treatment may be insufficient in this regime Castorina and White (2018).

As for the redshift dependence, we observe that the wide-angle contribution is most prominent at low redshift, while at around $z \sim 0.3$ it remains at the level of a few parts in 10^{-4}, confirming its subtle yet non-negligible impact in that regime. This should come as no surprise, since the comoving distance increases with redshift, while the angle between the line of sight and the galaxy pair separation vector narrows. As a result, the wide-angle contribution becomes progressively suppressed at higher redshift, effectively placing us back on firm ground within the distant-observer approximation.

The wide-angle effect is on the order of 10^{-4} across all scales considered at the average redshift of the simulations used in this thesis ($\overline{z} \sim 0.34$). However, if one wishes to investigate quasi-linear or nonlinear regimes, especially at low redshift and within modified gravity contexts that require higher values of the linear growth rate f, one should expect the wide-angle contribution to be not negligible, up to the order of 10^{-2}.

4.6.2 Light cone and evolution term

When quantitatively analysing clustering of high-redshift objects, properly accounting for light-cone effects is essential. Cosmological observations, in fact, are not made on a surface of constant time, but rather along the observer's past light cone, and all events are observed via light signals, which propagates through a perturbed spacetime. This introduces a subtle correction to the observed number density of objects (Bonvin et al., 2014).

While estimating the distance between two galaxies, we are, in fact, reconstructing the spatial gap between two events whose light reaches the observer at the same instant. Since the galaxies are positioned at different depths along the line of sight, the light from each must have originated at a different time in the past. If the galaxies are moving with respect to the observer, their locations change between those two emission times. This temporal mismatch leads to a subtle modification in the inferred separation.

Moreover, since sources evolve over time (and this produces, for examples, changes in bias), this evolution contaminates observed data, although such effects are subdominant at linear scales.

In Figure 4.11, we show the contribution of the light-cone term and the ξ^{evo1} term to the dipole. The evolution term arises from the explicit redshift dependence of the galaxy bias and the growth rate. This term contributes at the level of 10^{-5} to the dipole

signal, and becomes increasingly relevant at higher redshift, where the time derivatives of cosmological functions are larger and evolution effects are more pronounced.

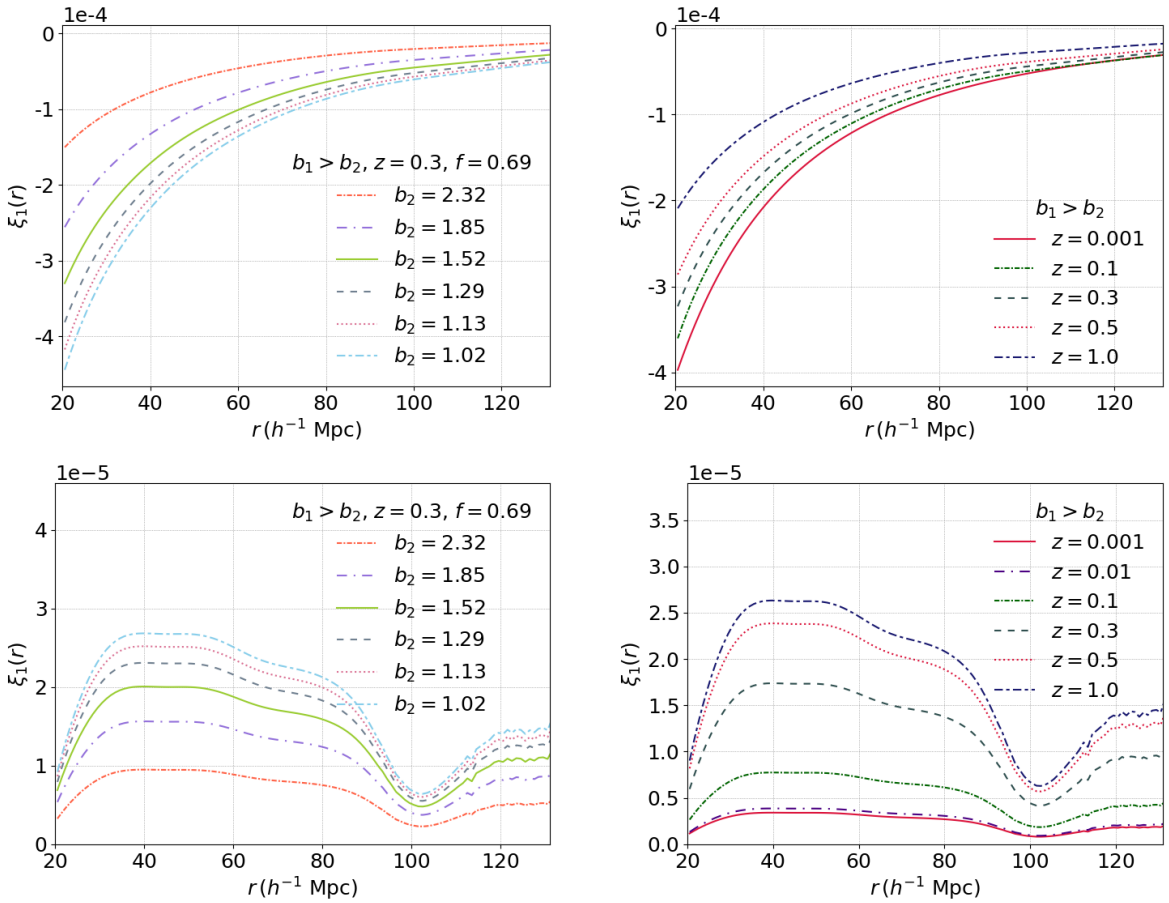

Figure 4.11: Contribution terms to the dipole from the light-cone effect (top panels) and from the evolution term ξ^{evo1} (bottom panels) with varying $\Delta b = b_1 - b_2$ (left) and z (right) in the ΛCDM model.

The contributions arising from the light cone are of the order of 10^{-5} at large scales, for all values of Δb and z considered. On smaller scales, however, these effects become more prominent and require greater attention, especially since at these scales they appear to be of the same order of magnitude as the contribution from the gravitational potential, as we shall see shortly.

Figure 4.12 illustrates the contribution to the dipole induced by light-cone effects, for different values of the growth rate f, while keeping all other cosmological and bias parameters fixed. The amplitude of the dipole increases with f, which is physically expected: a higher growth rate implies a faster evolution of density perturbations and therefore larger peculiar velocities. These velocities, projected along the line of sight,

contribute to distortions in redshift space that are asymmetric under the exchange of galaxy positions, thereby inducing a net dipole signal.

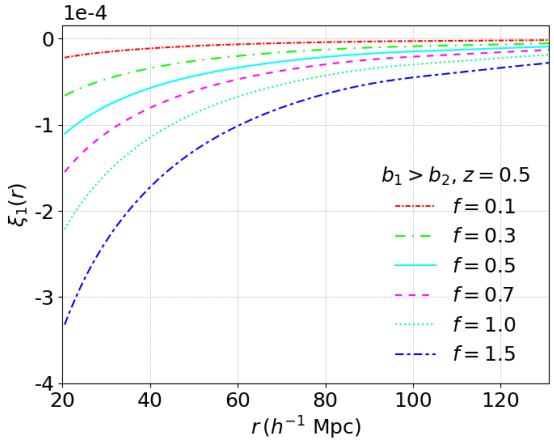

Figure 4.12: Contribution to the dipole from the light-cone term with varying linear growth rate. The value of f predicted by ΛCDM *is approximately represented by the curve at* 0.7.

Modified gravity models often predict deviations from the standard ΛCDM growth history, typically modifying the growth rate f, and, thus the growth of cosmic structures, especially at low redshift (Brax et al., 2010). This increase leads to stronger peculiar velocities and hence a more pronounced dipole contribution from the light-cone term. As a result, the dipole signal becomes a sensitive probe of deviations from GR, making accurate modelling of these subleading contributions, essential for robust constraints.

4.6.3 Lensing

Gravitational lensing occurs when light from distant sources is deflected by gravitational potentials encountered along its path, resulting in a discrepancy between the observed angular position and the source's true location. This phenomenon alters the apparent distribution of background galaxies and becomes particularly relevant in cross-correlation analyses involving two galaxy populations at different redshifts, typically a foreground population and a background (lensed) one. The mass distribution of the foreground galaxies magnifies or demagnifies the background population, inducing a modulation in the observed galaxy number counts.

Here lies the physical origin of the dipole signal: foreground galaxies act as lenses that alter the photon paths of more distant sources, enhancing the apparent number of galaxies behind the lens due to magnification, while leaving the foreground unaltered. This leads to an antisymmetric contribution to the correlation function. The amplitude

of this contribution depends on the difference in galaxy bias between the two populations and on the projected matter power spectrum.

Although generally subdominant when compared to other contributions to the dipole, the lensing effect must be accounted for in high-precision measurements, particularly at high redshifts.

In Figure 4.13, the lensing contribution to the dipole is illustrated for various values of the bias of the background (fainter) population and across different redshifts. The amplitude of the effect increases with the bias difference, but remains nearly scale-independent, exhibiting similar behaviour across a wide range of separations. Conversely, the redshift dependence shows a marked suppression of the lensing contribution at higher redshifts, while, for redshifts below $z \sim 0.1$, the curves become almost indistinguishable.

At low redshifts gravitational potentials are deeper, due to the more evolved structure formation, which enhances the deflection angle of photons. As a result, the lensing contribution to the dipole can become non-negligible at low redshift.

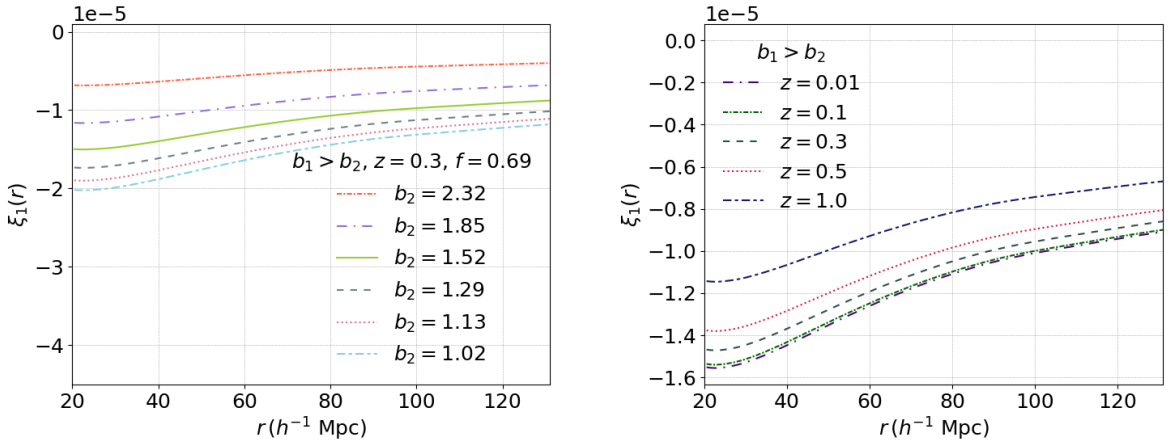

Figure 4.13: Contribution to the dipole from the lensing effect. The different lines are computed for different values of the bias of the fainter population (left panel) and of the redshift (right panel).

4.6.4 Acceleration of objects

Another effect which contributes to the origin of a dipole signal derives from the variation in velocity of the observed objects along the line of sight. This includes bulk acceleration of galaxies, but also local effects, which may dominate this term for $z \lesssim 0.1$.

In Figure 4.14, the dipole term is shown for different values of the bias of the fainter population, redshift, and linear growth rate. It is interesting to note that, in the upper-right panel, the dipole changes sign at around redshift $z \sim 0.1$. This sign inversion may originate from local effects dominating over the bulk acceleration of galaxies. In the

lower panel, a similar behaviour is observed for values of $f < 0.7$, that is, when the growth of structure is slower and local acceleration terms may dominate over global contributions associated with the LSS growth.

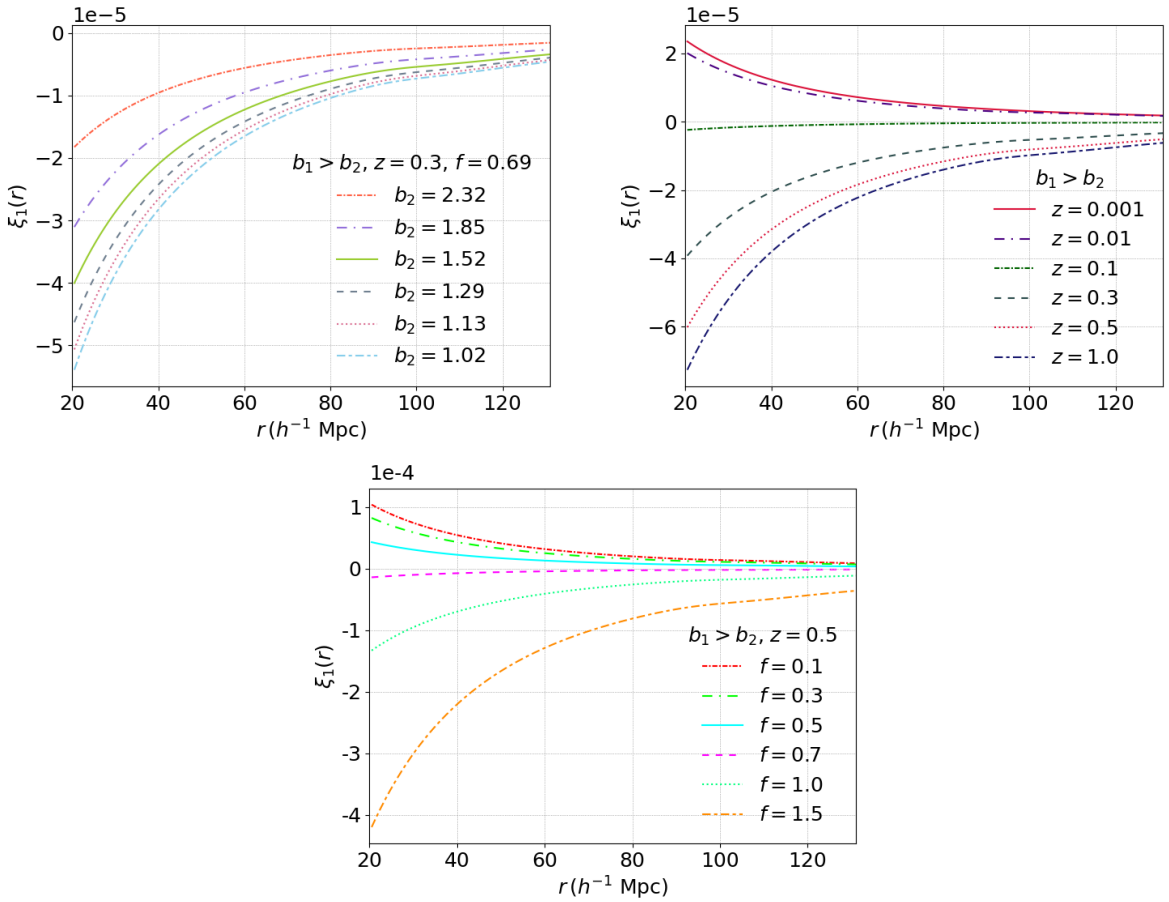

Figure 4.14: Contribution to the dipole from the acceleration term, shown for various values of the bias of the fainter population (top left panel), redshift (top right panel), and linear growth rate (bottom panel). The ΛCDM expectation for f is roughly indicated by the curve at 0.7.

This term could become extremely important in the context of modified gravity, as an increase in the linear growth rate f causes it to grow rapidly and become comparable in magnitude to other contributions, such as the gravitational potential, which we shall now discuss. However, this is particularly true on small scales, where nonlinear effects are known to play a role (Breton et al., 2019; Dam and Bonvin, 2025). For this reason, with upcoming high-precision observations, it may become promising to attempt to isolate this contribution.

4.6.5 Relativistic Doppler

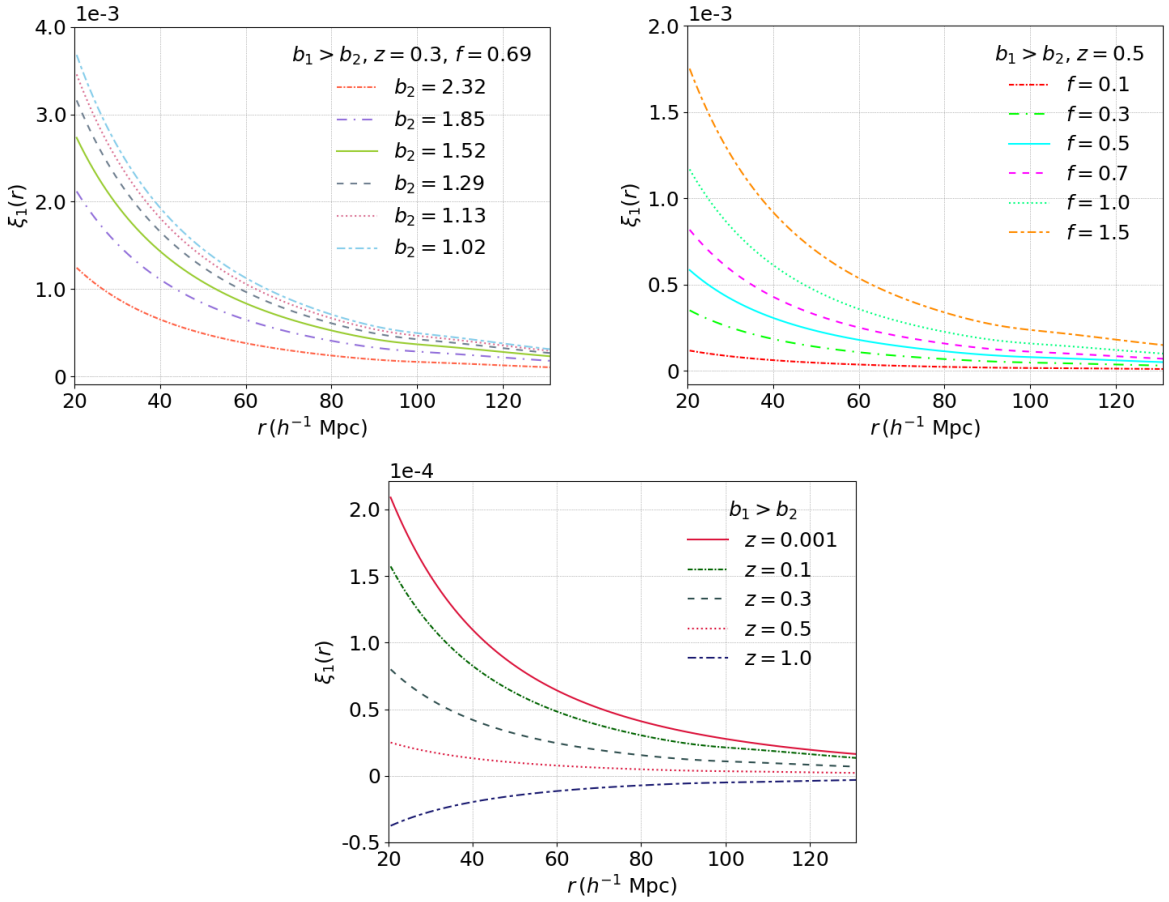

Figure 4.15: Top panels: contribution to the dipole from the relativistic Doppler term as a function of varying Δb (top) and linear growth rate f (bottom), where the ΛCDM prediction for f is approximately traced by the curve at 0.7. Bottom panel: Contribution to the dipole from the term associated solely with the accelerated expansion of the Universe.

A key relativistic effect is the relativistic Doppler contribution, which also affects fluctuations in number density, and observed redshift. The major contribution comes from the the line of sight velocity divergence effect, meaning that volumes moving towards the observer appear denser, while those moving away appear less dense. This geometric effect becomes the dominant contributor to the dipole signal on large scales. Though geometric in origin, the relativistic Doppler effect naturally fits within the relativistic treatment of redshift distortions in a perturbed spacetime and is thus part of the full relativistic corrections to galaxy number counts. The complete relativistic

Doppler effect on the dipole is shown in the top panels of Figure 4.15. The dipole contribution increases with the bias difference between the two populations and with the linear growth rate, as a larger growth rate amplifies the peculiar velocity field. With a keen eye, one can also discern a faint signal from the BAO at around $100\,h^{-1}$Mpc. The contribution to the dipole from the Universe accelerated expansion, with its dependence on redshift, is shown in the left bottom panel of Figure 4.15. As we approach the present epoch, this contribution becomes more pronounced; nevertheless, it remains clearly subdominant compared to the full relativistic Doppler signal, smaller by one to two orders of magnitude. Quantitatively, the main relativistic Doppler term is on the order of 10^{-3}, while the contribution related to the accelerated expansion of the Universe is around $10^{-4} - 10^{-5}$. As we will see later from the analysis of the dipole in the simulated data, the relativistic Doppler term largely dominates the dipole contribution across all the scales considered, except for scales approaching the nonlinear regime.

4.6.6 Gravitational potential

Gravitational redshift effects also contribute significantly to the dipole, although it is a subdominant contribution at large scales with respect to the relativistic Doppler term, as we will see in the following. As photons climb out of gravitational potential wells, they lose energy and experience redshift. Different populations, varying in mass or luminosity, undergo differing amounts of gravitational redshift, breaking the isotropy of the observed galaxy distribution. While standard RSD symmetrically affect galaxy positions along the line of sight, gravitational redshift introduces an asymmetry: the central, more massive galaxy experiences a larger redshift than its companions, causing a front-back asymmetry in their relative positions (Bonvin et al., 2014). This results in a cross-correlation function with both symmetric and antisymmetric components (and thus a dipole signal).

In Figure 4.16 (top panels) the total contribution from the potential term $\xi^{\mathrm{pot},1} + \xi^{\mathrm{pot},2}$ is shown. The overall contribution is of the order of 10^{-4} for a wide range of $\Delta b = b_1 - b_2$ and of redshifts. The bottom panel shows the contribution from the second-order terms-only. A comparison with the previously described figures clearly shows how subdominant this term is, relative to the total contribution from gravitational redshift.

From the top-left plot, we can observe the behaviour of the dipole term arising from the gravitational potential as redshift varies. At leading order, this term depends on the matter density parameter $\Omega_m(z)$, and its behaviour therefore mirrors that of this parameter. This explains at leading order why these curves are expected to lie at lower values.

However, at low redshift, the contribution of the potential to the dipole signal is not solely controlled by $\Omega_m(z)$. An additional correction appears in the full expression,

involving the conformal Hubble parameter $\mathcal{H}$, the speed of light c and an inverse dependence on the comoving distance χ. At low z, χ is small and this term becomes very large, leading to a negative contribution in the dipole signal, Eq. (3.60)).

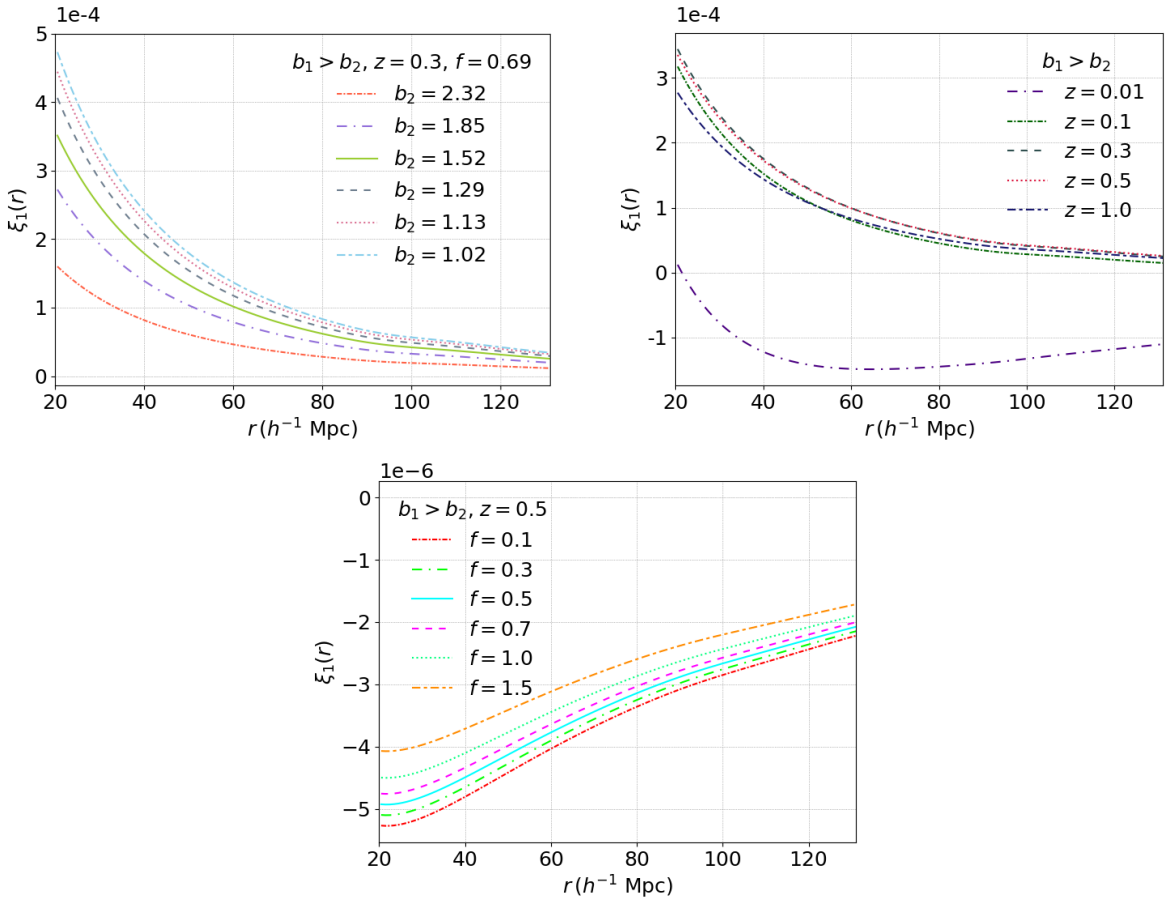

Figure 4.16: Gravitational redshift contribution to the dipole for different values of bias of the fainter population (top left panel) and redshifts (top right panel). Bottom panel: gravitational redshift contribution to the dipole from the leading-only terms in $(\mathcal{H}/k)^2$ for a broad range of the linear growth rate. The ΛCDM value of f is roughly illustrated by the curve at 0.7 in this plot.

We can also observe a number of interesting features: the contribution from the gravitational potential is smaller than that of the relativistic Doppler term, yet it is entirely comparable to several other contributions, such as the light cone and wide-angle effects. Naturally, our treatment is accurate on large, linear scales, while it breaks down in the nonlinear regime. Nevertheless, a preliminary analysis makes it clear that, if one aims to isolate the gravitational potential contribution, the effects of the light-cone and wide-angle terms cannot be neglected. Therefore, this needs to be properly modelled

and compared in a thorough analysis across all scales, as also confirmed by the recent study of Dam and Bonvin (2025).

4.6.7 Shapiro time-delay

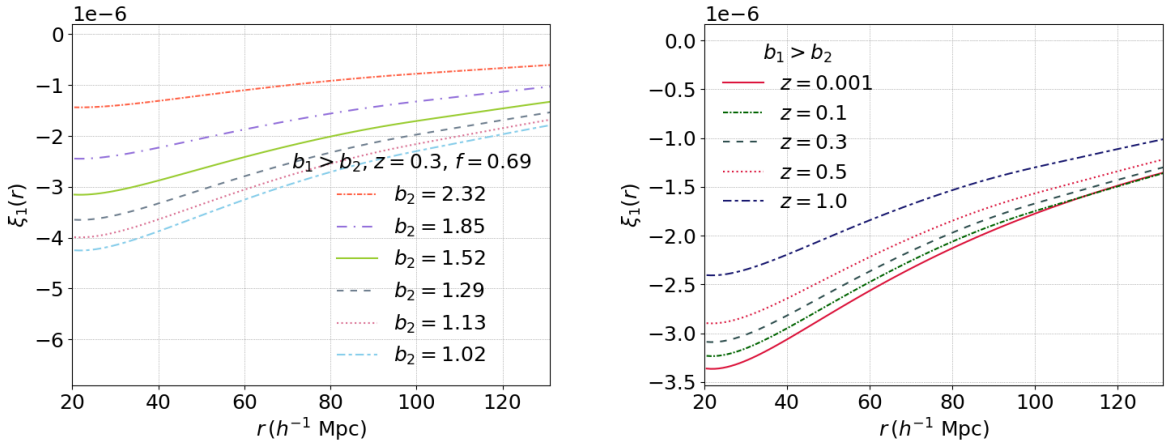

Figure 4.17: Contribution to the dipole from the Shapiro time delay.

The Shapiro time delay is a relativistic effect caused by gravitational potential wells of massive structures, which cause delaying in photon propagation compared to unperturbed flat spacetime. Unlike local gravitational redshift, it is an integrated effect accumulated along the line of sight. This delay perturbs observed galaxy number counts, contributing a subtle, but non-zero, relativistic correction. When cross-correlating different galaxy populations, the Shapiro delay introduces an antisymmetric distortion along the line of sight, generating a dipole in the 2-point cross-correlation function.

In Figure 4.17 the contribution to the dipole from the Shapiro time delay is illustrated, for different values of $\Delta b = b_1 - b_2$ and redshift.

Specifically, in the top-left plot we observe the behaviour of the dipole, arising from the time delay effect, for increasing values of bias difference between the two populations. The contribution tends to increase with the bias difference. A similar behaviour is evident in the plot on the right. Higher redshift values correspond to a stronger dipole signal from the time delay effect (which is little wonder, given that this is an integrated effect accumulating along the photons' path). The contribution from the Shapiro time delay becomes progressively less significant at larger scales, albeit with only a slight decreasing trend in magnitude. This is due to the fact that variations in the gravitational potential are less pronounced on large scales, as the Universe tends to be homogeneous and isotropic, and, as a result, the time delay effect is less prominent at these scales.

In summary, although the Shapiro time delay is highly subdominant compared to other relativistic contributions such as relativistic Doppler or gravitational redshift, it plays a crucial role in breaking the symmetry of the correlation function when cross-correlating distinct tracers of LSS. Its contribution is most significant on large radial separations and can be probed via the dipole of the cross-correlation, offering an additional relativistic signature to test GR on cosmological scales. Although this contribution has been implemented in the CBL and its behaviour analysed, it has not been taken into account in the analysed simulation, and we will also neglect it from now on.

4.6.8 Integrated Sachs-Wolfe

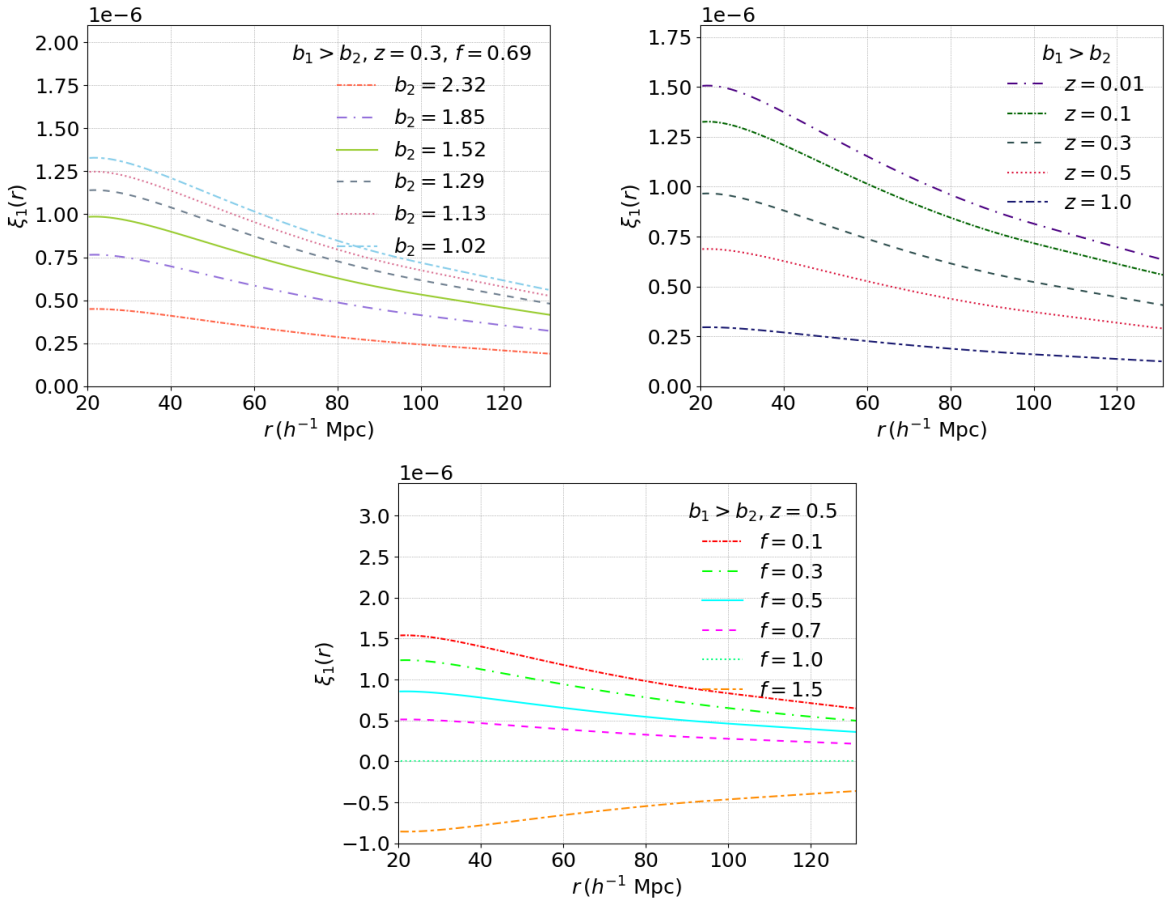

Figure 4.18: Contribution to the dipole from the ISW effect. Bottom panel: the ΛCDM *prediction is for approximately* $f \simeq 0.7$.

The ISW effect arises from time-dependent variations in the gravitational potential experienced by photons along their path. In a matter-dominated Universe, the gravitational

potential remains nearly constant over time at large scales, leading to a negligible ISW effect. However, when a component such as a cosmological constant or dark energy drives an accelerated expansion, the gravitational potentials evolve with time, giving rise to a non-zero ISW signal. Thus, the detection of a significant ISW effect provides indirect evidence for the presence of dark energy. Moreover the ISW leave an imprint in the dipole of the cross-correlation function.

The ISW effect on the dipole contributes a signal that decreases at larger scales, is proportional to the bias difference between the populations considered, being stronger for larger Δb, and diminishes as redshift increases, given the nature of an integrated effect.

In particular, the lower panel shows that for values of $f > 1$, the contribution is opposite in sign to that for values within the range $0 < f < 1$. Linear growth rates exceeding unity make density perturbations to grow faster than the cosmic expansion rate, implying an increase in the gravitational potential over time. This leads to photons losing energy, resulting in negative fluctuations in the observed number counts. However, such high values are incompatible with the standard ΛCDM model and instead would necessitate alternative scenarios that involve modified gravity.

4.6.9 Closing remarks

	$r > 20\,h^{-1}\,\mathrm{Mpc}$		$z \sim 0.3$	
	$z \lesssim 0.1$	$0.1 < z < 1$	$r \lesssim 50\,h^{-1}\,\mathrm{Mpc}$	$r > 50\,h^{-1}\,\mathrm{Mpc}$
ξ^{wa}	$\lesssim 10^{-2}$	$\gtrsim 10^{-2}$	$10^{-3}/10^{-4}$	$\lesssim 10^{-4}$
ξ^{LC}	$\lesssim 10^{-4}$	$\lesssim 10^{-4}$	$\sim 10^{-4}$	$\lesssim 10^{-4}$
ξ^{evo}	$\sim 10^{-6}$	$\sim 10^{-5}$	$\sim 10^{-5}$	$\sim 10^{-5}$
ξ^{lens}	$\sim 10^{-5}$	$\sim 10^{-5}/10^{-6}$	$\sim 10^{-5}$	$\sim 10^{-5}$
$\xi^{\mathrm{Rel,Doppler}}$	$10^{-5}/10^{-4}$	$\sim 10^{-5}$	$\sim 10^{-3}$	$\sim 10^{-4}$
ξ^{acc}	$\lesssim 10^{-5}$	$\lesssim 10^{-5}$	$\sim 10^{-5}$	$< 10^{-5}$
ξ^{pot}	$\lesssim 10^{-4}$	$\lesssim 10^{-4}$	$\sim 10^{-4}$	$< 10^{-4}$
ξ^{Shapiro}	$\sim 10^{-6}$	$\sim 10^{-6}$	$\sim 10^{-6}$	$\sim 10^{-6}$
ξ^{ISW}	$\sim 10^{-6}$	$\sim 10^{-7}$	$\lesssim 10^{-6}$	$\sim 10^{-7}$

Table 4.2: Summary of all the dipole contributions.

In Table 4.2, a summary of all contributions to the dipole is shown, indicating the order of magnitude of individual terms, to clearly visualise which effect dominates on which scale, for redshifts up to unity and linear scales from 20 to $140\,h^{-1}$Mpc. We choose $z \sim 0.3$ as a representative value, as it is the mean redshift value of the haloes

considered in this thesis. As can be seen at this redshift, the dominant relativistic contribution to the dipole arises from the relativistic Doppler term, while at larger scales the contributions from the relativistic Doppler effect and gravitational redshift are comparable.

	ξ^{Doppler}	ξ^{pot}	ξ^{WA}	ξ^{ISW}	ξ^{lens}	ξ^{LC}	ξ^{evo1}	ξ^{evo2}
(β, z_0)						×		×
(θ, z_0)					×			
(β, z_1)		×						
(β, z_2)	×		×				×	
(β, z_3)								
(β, z_4)				×				
(θ, z_5)	×	×	×	×	×		×	

Table 4.3: Terms to be considered when modelling the dipole for different angle and redshift combinations.

All these contributions are summarised in Table 4.3. A cross means that the terms should be added when predicting the dipole for different combinations of angles and specific redshifts of the simulation, each including different contributions.

4.7 Dipole extraction and analysis

We have already described in detail how the RayGalGroupSims simulations are constructed. However, it is essential to provide some additional cosmological information to fully contextualise the data. This supplementary information allows for a clearer understanding of the underlying assumptions and parameters that shape the simulations, ensuring the robustness of the analysis. The simulation consists of 13 million dark matter haloes (Breton, 2018), with the full-sky version of the simulation set covering a redshift range of $[0, 0.5]$. This set of simulations employ CAMB to compute the initial matter power spectrum, assuming a ΛCDM cosmology with the following parameters:

$$\begin{aligned} h &= 0.72, \\ \Omega_m &= 0.25733, \\ \Omega_b &= 0.04356, \\ \Omega_r &= 8.076 \times 10^{-5}, \\ n_s &= 0.963, \\ \sigma_8 &= 0.801, \end{aligned}$$

where Ω_m, Ω_b and Ω_r represent the matter, baryon and radiation density parameters, respectively, n_s is the slope of primordial power spectrum and σ_8 is a parameter used to normalise the amplitude of fluctuation density, referred to a scale of $8\,h^{-1}$Mpc. The initial conditions were generated using second-order Lagrangian perturbation theory (2LPT), starting from a redshift of $z \sim 50$.

The subdivision into sub-samples, considered in this work, is reported in Table 4.1. Each catalogue was constructed using halo observed coordinates, (RA, Dec), derived from the two angular coordinates provided by the simulation (the azimuthal angle and the colatitude), and the redshift, directly extracted from the simulation data.

The corresponding random catalogues were generated on the same light cones, preserving the same angular and redshift distribution (RA, Dec, z) as the data catalogues. To preserve the clustering information along the line of sight, the redshift distribution is smoothed by dividing it into 10 redshift bins. The selected redshift interval, ranging from 0.05 to 0.465, ensures an accurate angular reconstruction by excluding sources too close to the observer and avoids edge-related artefacts near the boundaries of the light cone.

The selected haloes span a mass range between 1.88×10^{12} $h^{-1}\mathrm{M}_\odot$ and 2.2×10^{14} $h^{-1}\mathrm{M}_\odot$. The total number of haloes considered is approximately 1.2×10^7, corresponding to a mean number density of 5.4×10^{-4} Mpc^{-3}.

Each catalogue, labelled as H_N, includes dark matter haloes composed of a number of particles between N and $2N$. The mass resolution of the simulation is 1.88×10^{10} $h^{-1}M_\odot$. The radial separation range was divided into 12 bins spanning from 20 to $140\,h^{-1}$Mpc. This interval was chosen to focus on intermediate to large scales: these scales are of primary interest for investigating the linear dipole signal, while still allowing for a broader perspective that includes the transition to the quasi-linear regime at scales below $30\,h^{-1}$Mpc. This choice provides an opportunity to assess to what extent the theoretical models remain valid or begin to break down. Scales below $20\,h^{-1}$Mpc are strongly affected by nonlinear effects and shot noise, while scales beyond $140\,h^{-1}$Mpc are typically limited by sample variance. A bin width of approximately $10\,h^{-1}$Mpc represents a good compromise between resolution and statistical robustness, ensuring that the contribution to the dipole signal is adequately sampled, without introducing excessive noise.

In the angular domain, 208 bins in $\mu = \cos(\theta)$[1] were considered, with μ allowed to vary between -1 and 1. This fine angular resolution is essential to capture the anisotropies induced by RSD and to ensure a reliable extraction of the dipole signal from the data. A large number of angular bins reduces numerical noise and improves the

[1]We recall that the angle θ considered here is defined by the separation vector between the two objects and the line of sight, which is taken to be the vector pointing to the midpoint between the two objects. This is not, however, the only possible choice (de la Torre et al., 2025).

accuracy of the discrete summation performed in the dipole calculation. We analysed the dipole signal arising from various cross-correlations. To do this, we implemented within the CBL both the functions to extract the multipoles of the cross-correlation from the data, and the functions to model the different terms contributing to the dipole, as described in detail in Chapter 3.

Throughout the analysis, we consider the first population to be the more massive one, such that $b_1 > b_2$. The coupling between potential and relativistic Doppler effects, as well as the full relativistic dipole, including potential, relativistic Doppler, wide-angle, lensing, evolution, and ISW effects, on large scales, are shown. Superimposed models are represented by solid black lines. The dipole models were tested by evaluating their χ^2 values, in order to quantify how well the theoretical predictions reproduce the observed signal. The chi-squared statistic is widely used to assess the agreement between observed data and a theoretical model, and it quantifies the goodness of the fit (Plackett, 1983). The χ^2, for independent data, is defined as follows:

$$\chi^2 = \sum_{i=1}^{N} \frac{(x_{\mathrm{n},i} - x_{\mathrm{th},i})^2}{\sigma_i^2}, \tag{4.22}$$

where $x_{\mathrm{n},i}$ is the observed value, $x_{\mathrm{th},i}$ is the theoretical value, σ_i^2 is the variance of the observed value and N is the total number of data points. Here we computed the reduced χ_r^2, which corresponds to the χ^2, normalised by the number of degrees of freedom. As expected, in cases where the observational errors are large compared to the amplitude of the signal, the reduced χ^2 becomes less informative and cannot reliably discriminate whether the model is good or necessitates to be refined. In cases where the χ^2 value is relatively high, this may point to a tension between the theoretical prediction and the observed dipole signal. Such a result indicates that the model does not fully capture the features present in the data, possibly due to missing contributions or approximations made in the theoretical treatment.

4.7.1 Contributions from the relativistic Doppler and gravitational potential terms

The relativistic Doppler and gravitational potential contributions to the dipole in the cross-correlation between differently biased tracers are presented in Figure 4.19. To facilitate comparison, the y-axis scale is kept consistent across all plots, emphasizing the differences among the various cross-correlations. The bias values of the populations involved are indicated in the figure, with the bias difference Δb ranging from 0.38 to 1.51. The observed dipole signal extracted from the data exhibits a general increasing trend with the growing bias difference between the two catalogues.

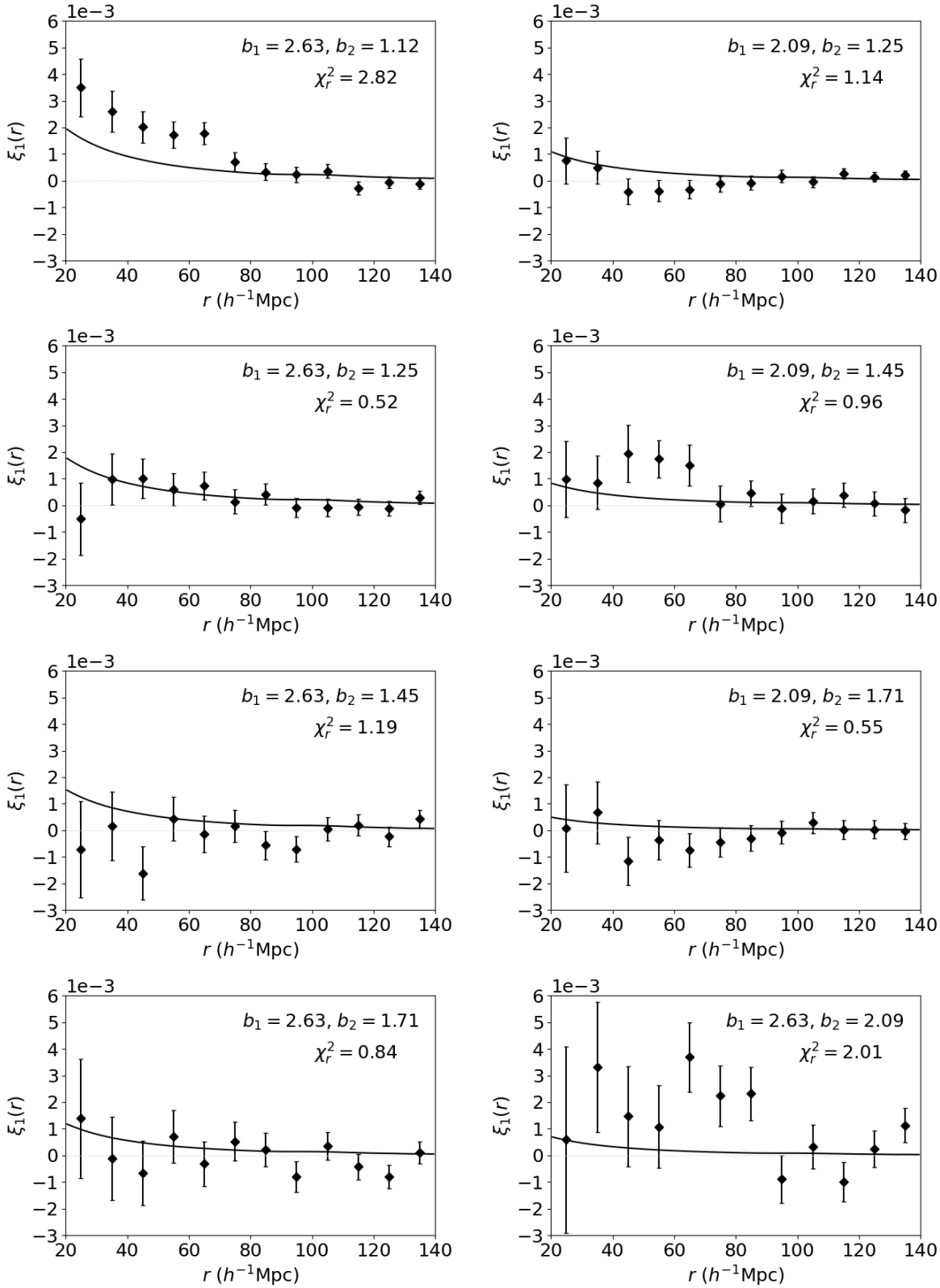

Figure 4.19: Relativistic Doppler+potential term of the dipole of the cross-correlation function between different data sets. Models are shown as black solid lines.

Correspondingly, the theoretical models scale proportionally with this bias difference, as discussed in Chapter 3.

The cross-correlations shown involve catalogues with progressively increasing differences in halo mass, and therefore in galaxy bias, and specifically include cross-correlations between the H_{3200} and H_{1600} catalogues and the other populations. This setup allows us to examine how the dipole depends both on the individual bias values of the tracers, and on the bias difference between them. Significant scatter is present at small and intermediate scales, while the signal becomes smoother and tends to approach zero at larger scales. The curves representing the linear theory fits for these data show a dipole behaviour that tends to increase at small scales for the chosen bias values. The measurements from simulations are in overall agreement with the expectations, except for minor discrepancies at small scales; this is slightly visible e.g. in the second panel on the left from the top. Recent studies appear to confirm this latter behaviour at small scales. This could be a real effect, or could be due to sampling issues. The dipole signal arises from perturbations in the observed redshift that incorporate both relativistic Doppler and gravitational potential effects. Our model is cumulative, comprising several contributions: terms associated with the cosmic expansion, the divergence term and the acceleration term (which are properly identified as the relativistic Doppler effect), along with the gravitational potential term. Additionally, evolution effects and wide-angle corrections are included, although the latter two are subdominant.

While these components originate from an interplay between relativistic Doppler shifts and gravitational potentials, the relativistic Doppler term overwhelmingly dominates the dipole signal, being roughly one to two orders of magnitude larger than the contribution from gravitational potentials alone. This dominance will also be evident in later comparisons isolating these individual effects. The radial distance scale employed matches that previously discussed for the even multipoles. It is important to acknowledge that the models are developed within the linear perturbation regime, where we expect a good description of the data. Nevertheless, restricting the analysis only to large scales would be limiting, so we also explore the dipole behaviour on small and intermediate scales. The number of radial bins is chosen to strike a balance between achieving sufficient statistical robustness and adequate sampling within each bin. For the cross-correlation involving the two most massive populations (bottom right panel), the data show significant scatter around the model prediction. Although a faint dipole signal appears at small and intermediate scales, the high uncertainties do not allow a definitive interpretation nor any firm conclusion about the model accuracy.

In contrast, the other cross-correlations generally exhibit a reasonable agreement between model and data. However, as also noted by Breton et al. (2019), the small to intermediate scales (approximately 30 to 60 h^{-1}Mpc) remain the most challenging. This difficulty arises because nonlinear effects, which are neglected in our current linear

framework, become relevant at these scales. These discrepancies highlight directions for future work aimed at improving the model. With the exception of the case mentioned above, where the number of particles in the catalogues is limited, the χ^2_r values computed for the fits are generally close to unity. This is noteworthy given that the analysis also includes data from small scales where the model is not expected to perform well.

The signal-to-noise ratio (S/N) for all the catalogues considered lies between 3.0 and 4.1, which indicates a statistically meaningful detection of the dipole signal. Only in one case the S/N reaches a value of approximately 8, which may point to either an anomaly in the analysis of that specific catalogue, such as an underestimated error or a particular configuration amplifying the signal, or to a genuinely enhanced signal in that configuration. Further investigation will be required to clarify the origin of this deviation.

Given the limited volume of the simulations and the relatively small number of tracers, these values demonstrate the robustness of the estimator and the viability of measuring the dipole with future, larger data sets.

Increasing the number of realisations and improving the modelling at intermediate scales will help to stabilise and enhance the significance of the detection. In this context, forthcoming wide-area spectroscopic surveys such as *Euclid* are expected to provide the volume, depth, and statistical power necessary to achieve a higher S/N, allowing for a robust detection of the relativistic contribution to the dipole.

Although the contributions show in the figure arise from a coupling between the relativistic Doppler effect and the gravitational potential, the relativistic Doppler term is the one to dominate, as it is one to two orders of magnitude larger than the dipole induced by the gravitational potential, as previously discussed. We will also see it later on by the comparison between this two isolated effects.

4.7.2 Potential-only term

A comparison between the relativistic Doppler term and the gravitational potential-only contribution to the dipole, for three representative cross-correlation configurations, is shown in Figure 4.20. These cases were selected as illustrative examples covering different ranges of halo population properties.

The top panels display the dipole signal for the cross-correlation between the two most massive halo populations, H_{3200} and H_{1600}, which also have the lowest number densities. The left-hand panel shows the relativistic Doppler contribution, while the right-hand panel shows the isolated gravitational potential term. The potential contribution in this configuration is extremely difficult to isolate, mainly due to poor statistics stemming from the small sample size. S/N ratio is 3.1 and 0.29 for the relativistic Doppler-only and potential-only contribution, respectively, and we will discuss the implications in

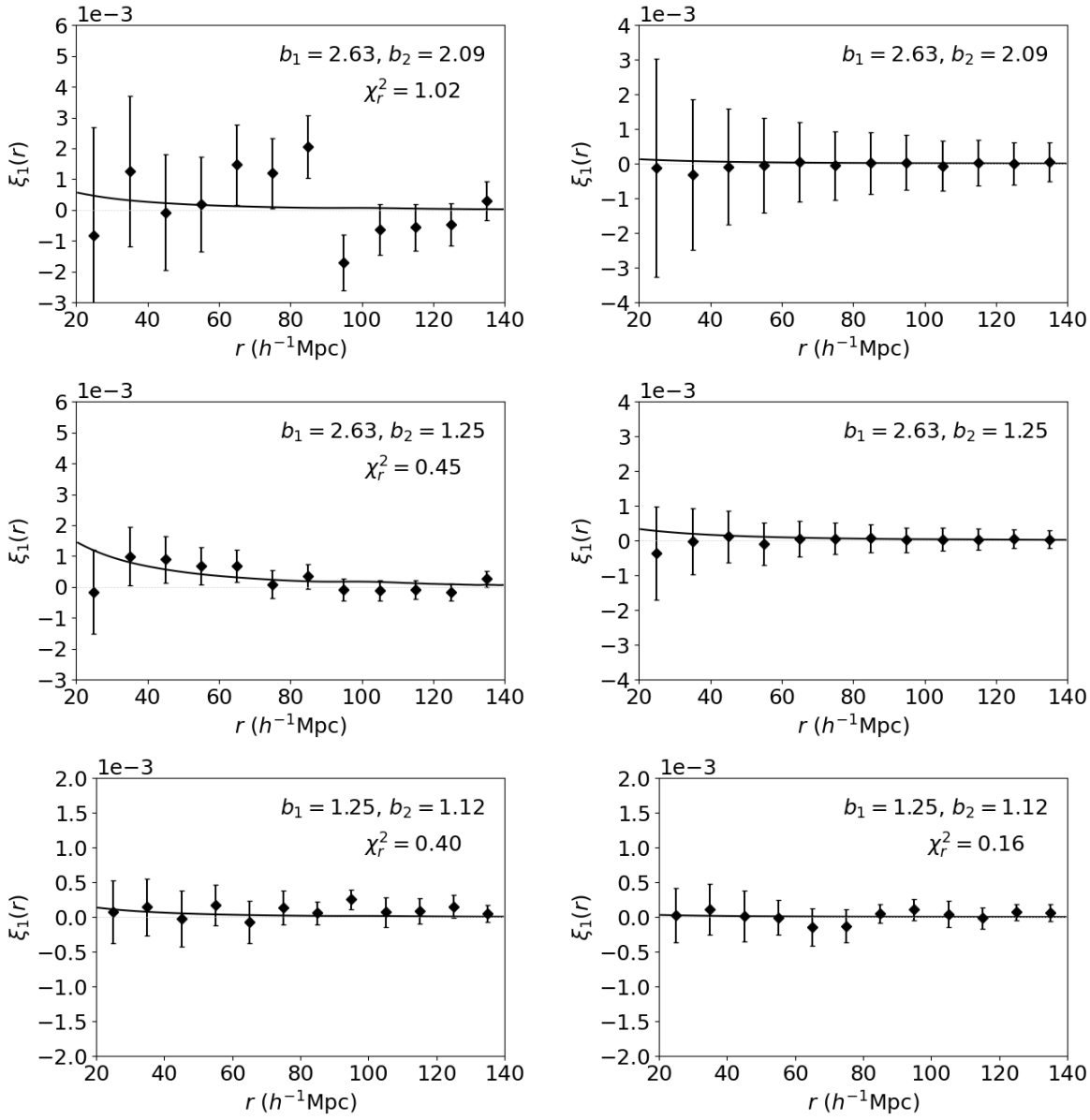

Figure 4.20: Top panels: cross-correlation between H_{3200} and H_{1600}. Middle panels: cross-correlation between H_{3200} and H_{200}. Bottom panels: cross-correlation between H_{200} and H_{100}. Comparison between relativistic Doppler only contribution (left panels) and potential only contribution (right panels) to the dipole.

what follows.

The middle panels refer to the cross-correlation between H_{3200} and H_{200}, which represents a case with a large bias difference between the two tracers. Again, the relativistic Doppler term is shown on the left and the potential term on the right. In this configuration, uncertainties are significantly reduced, though still sizeable on small scales. While these error bars prevent us from drawing definitive conclusions on the gravitational potential term, the dominance of the relativistic Doppler contribution is evident. This is especially clear when comparing both the amplitude and shape of the isolated effects with the corresponding total signal in Figure 4.19, as well as with the analogous plots in the top panels. In particular, the relativistic Doppler contribution remains approximately one order of magnitude larger than the potential one on small and intermediate scales, in agreement with the theoretical predictions. These findings are consistent with expectations: the relativistic Doppler term is enhanced as the bias difference increases, while the gravitational potential contribution remains subdominant and more sensitive to noise. Here S/N settles around 3.1 for the relativistic Doppler contribution and 0.5 for the potential contribution.

The bottom panels correspond to the cross-correlation between the two least massive halo populations, H_{100} and H_{200}, both characterised by low bias values and a small bias difference. Unlike the previous cases, this configuration does not show a clear dominance of either contribution across the entire range of scales, and it is confirmed also by S/N (1.9 for the relativistic Doppler term and 1.4 for gravitational potential only). Here, the two effects appear to be comparable within the uncertainties. Due to the significantly higher number of objects in both catalogues (approximately an order of magnitude more than in the massive halo samples), the error bars are smaller, enabling a more stable estimation of the χ^2_r for the potential-only fit, which is included for this case only. The resulting values are similar to those found for the joint contribution from the relativistic Doppler and potential terms, shown in the previous section, and are deemed acceptable, especially considering the known limitations of linear models on small scales.

As for the S/N: for the cross-correlations involving H_{3200}, with H_{1600} and H_{200}, the relativistic Doppler-only signal yields values around 3, as also confirmed by Lepori et al. (2025), although with studies involving small scales. In contrast, the potential-only case remains at or below 0.5 in both configurations. Despite the high level of noise due to the limited number of objects, the relativistic Doppler signal remains clearly detectable. However, no firm conclusion can be drawn regarding the gravitational potential contribution at these scales, primarily due to the large uncertainties stemming from the low number density in the two catalogues.

In the cross-correlation between the less massive and more numerous halo catalogues, H_{100} and H_{200}, the overall signal is substantially lower, including the relativistic Doppler contribution, due to the small bias difference between the tracers. In this case, the S/N

is comparable for both effects (relativistic Doppler-only: 1.9; potential-only: 1.4), which prevents us from drawing any definitive conclusions.

Nevertheless, the analysis and comparison of these three cross-correlation configurations offer broader insights and valuable indications for future studies. Firstly, the relativistic Doppler effect emerges as dominant, particularly on large scales, which has hindered our ability to adequately isolate the gravitational potential contribution. Theoretical considerations suggest that future investigations should prioritise catalogue selections with pronounced bias differences, as these enhance the dipole signal. On the data side, it is clear that very large and highly precise surveys, both simulated and real, will be necessary, especially to attempt the detection of the weaker dipole contributions on large scales.

In this respect, future wide-area spectroscopic surveys, such as *Euclid*, will be essential to improve statistical significance, test relativistic predictions, and potentially isolate the gravitational potential contribution in galaxy clustering.

4.7.3 Full dipole

In Figure 4.21, the total dipole signal is presented, which includes the relativistic Doppler, gravitational potential, ISW, lensing, evolution, wide-angle and light-cone contributions.

Comparing this with the corresponding plots including only the relativistic Doppler and potential terms (Figure 4.19), it is clear that the total signal is largely dominated by the relativistic Doppler effect, across all scales considered, with an amplitude of the order of 10^{-3}.

The S/N are as follows: 2.98 (H_{3200}–H_{1600}), 2.50 (H_{3200}–H_{800}), 3.09 (H_{3200}–H_{400}), 3.94 (H_{3200}–H_{200}), 8.09 (H_{3200}–H_{100}), 3.07 (H_{1600}–H_{800}), 1.71 (H_{1600}–H_{400}), and 3.05 (H_{1600}–H_{200}). Most measurements reach a significance level of about 3σ, as confirmed by other studies on the dipole in the literature (Lepori et al., 2025), except for two cases: one exhibits a very high S/N, which, as previously discussed, could either be a genuine physical signal, a particular feature of that cross-correlation, or simply an artefact stemming from statistical fluctuations. The other case, H_{1600}–H_{400}, shows a noticeably lower value.

Regarding this latter discrepancy, the lower S/N for the H_{1600}–H_{400} cross-correlation may be attributed to several factors:

- Sample variance: given the finite size of the catalogues, statistical noise might have been reduced the effective significance;
- Systematic effects or incomplete modelling: certain methodological systematic errors may affect this particular cross-correlation more strongly, and this could thus represent an indication for models to require further refinement.

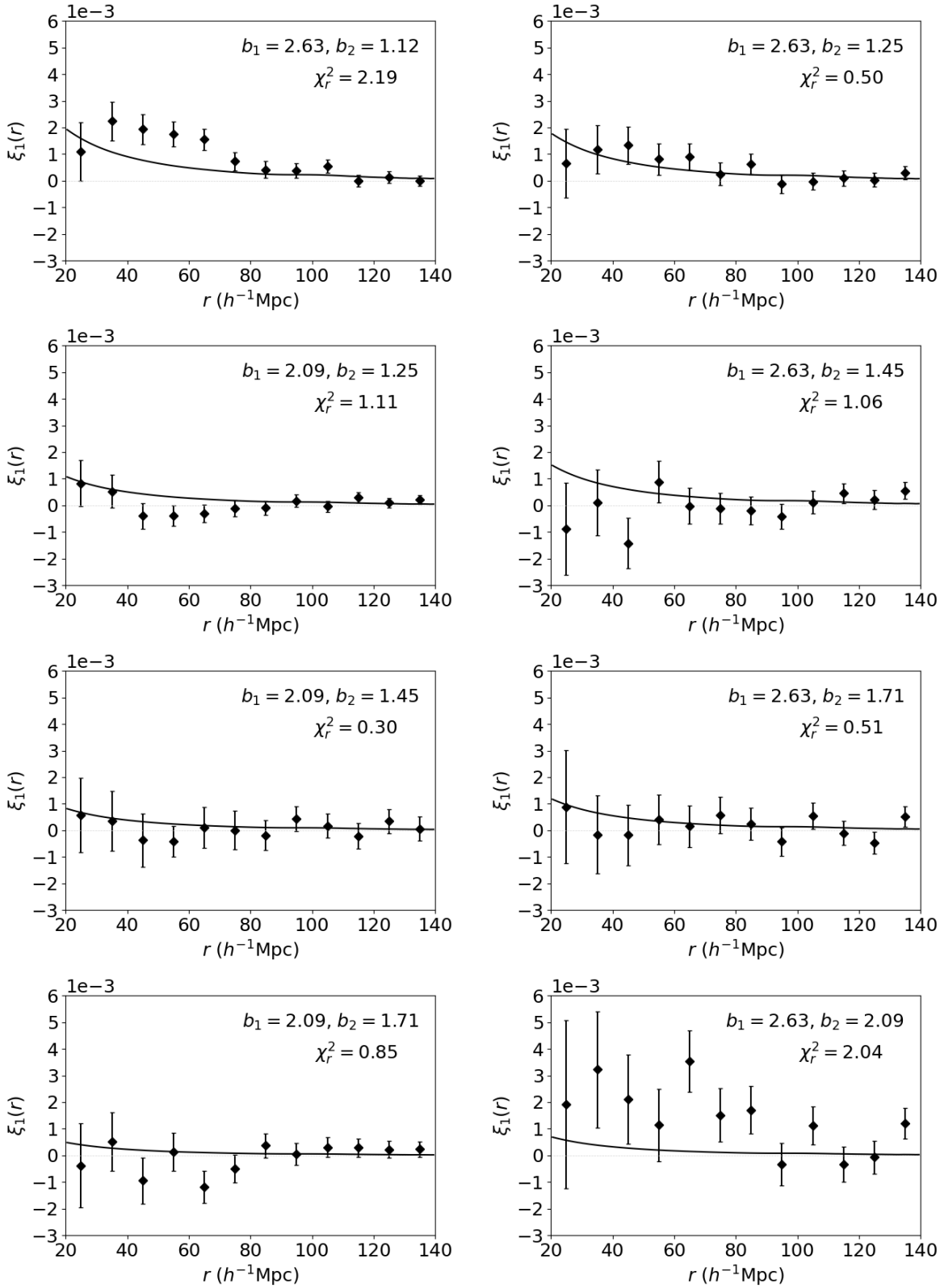

Figure 4.21: Full dipole of the cross-correlation function between different data sets. Models are shown as black solid lines.

As for the model: the χ_r^2 values deteriorate when considering the full dipole model compared to the simpler relativistic Doppler plus gravitational potential case. This is unsurprising, as the increased complexity from adding further relativistic terms introduces more uncertainties in modelling, which can lead to a poorer fit to the current data. Accurately modelling these additional effects will likely require further theoretical refinement to be compared to potentially higher-precision data.

As observed in all previous cases, there is a general trend of increasing dipole signal in magnitude on small and intermediate scales, as well as with increasing bias difference between the populations, as also shown by Bonvin et al. (2014) and Alam et al. (2017a). These trends indicate promising directions for the analysis of future data sets from surveys such as *Euclid*.

Disentangling the other contributions proves challenging, since their induced dipole signals are one to several orders of magnitude smaller than the dominant relativistic Doppler effect. The same considerations that apply to the gravitational potential term also hold for ISW, evolution, wide-angle, and lensing effects, all of which remain subdominant in the total dipole signal at large scales.

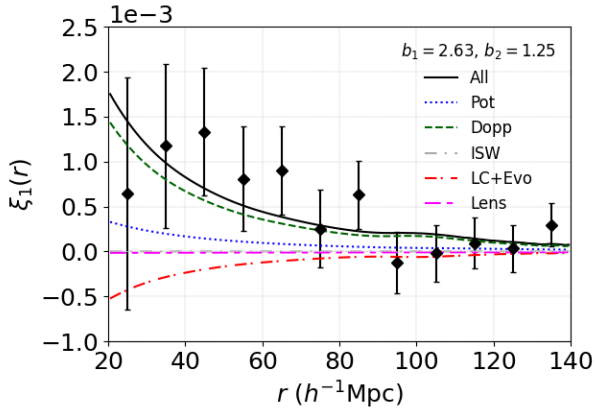

Figure 4.22: Dipole of the cross-correlation function between H_{3200} and H_{200}. The total model, resulting from the sum of all contributions to the dipole, is shown as a black solid line. The other curves represent the individual contributions: relativistic Doppler (green dashed line), gravitational potential (blue dotted line), ISW (grey line), lensing (magenta line), and the combined effect of light-cone and evolution terms (red line).

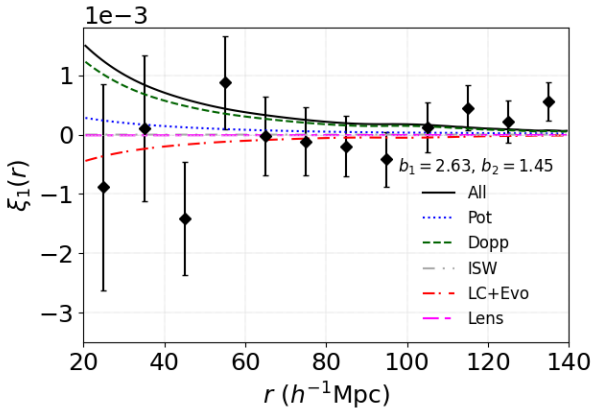

Figure 4.23: Dipole of the cross-correlation function between H_{3200} and H_{400}. The total model, resulting from the sum of all contributions to the dipole, is shown as a black solid line. The other curves represent the individual contributions: relativistic Doppler (green dashed line), gravitational potential (blue dotted line), ISW (grey line), lensing (magenta line), and the combined effect of light-cone and evolution terms (red line).

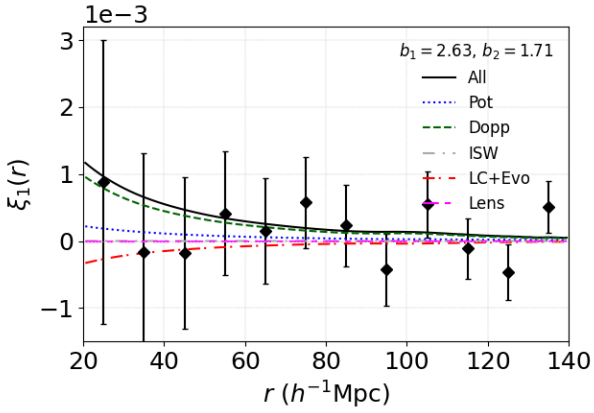

Figure 4.24: Dipole of the cross-correlation function between H_{3200} and H_{800}. The total model, resulting from the sum of all contributions to the dipole, is shown as a black solid line. The other curves represent the individual contributions: relativistic Doppler (green dashed line), gravitational potential (blue dotted line), ISW (grey line), lensing (magenta line), and the combined effect of light-cone and evolution terms (red line).

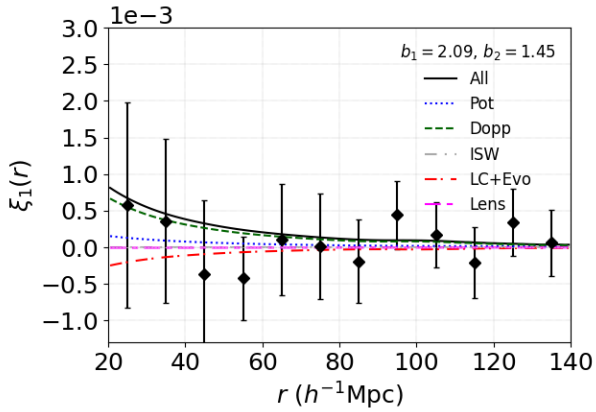

Figure 4.25: Dipole of the cross-correlation function between H_{1600} and H_{400}. The total model, resulting from the sum of all contributions to the dipole, is shown as a black solid line. The other curves represent the individual contributions: relativistic Doppler (green dashed line), gravitational potential (blue dotted line), ISW (grey line), lensing (magenta line), and the combined effect of light-cone and evolution terms (red line).

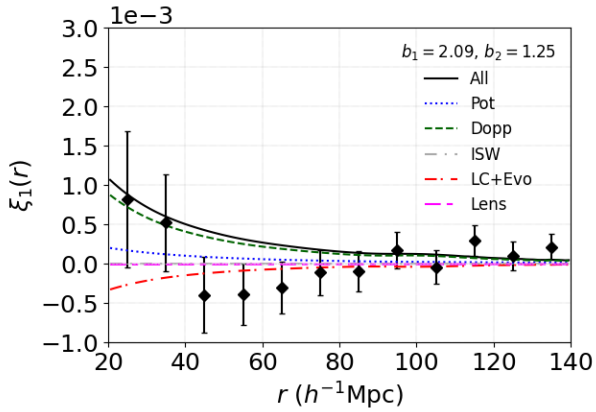

Figure 4.26: Dipole of the cross-correlation function between H_{1600} and H_{200}. The total model, resulting from the sum of all contributions to the dipole, is shown as a black solid line. The other curves represent the individual contributions: relativistic Doppler (green dashed line), gravitational potential (blue dotted line), ISW (grey line), lensing (magenta line), and the combined effect of light-cone and evolution terms (red line).

In Figures 4.22 to 4.26, we present the cross-correlations with a $S/N \sim 3$, for the cases discussed above. The plots are shown in descending order of the bias difference, Δb, between the population considered. The total model (black solid lines), which includes the sum of all contributions, is displayed alongside the individual components: the relativistic Doppler effect (green dashed lines), gravitational potential (blue dotted

lines), ISW effect (grey lines), lensing (magenta lines), and the combination of light-cone and evolutionary effects (red lines).

As clearly illustrated in these plots, the total dipole signal is dominated by the relativistic Doppler contribution across all the scales considered. The ISW and lensing terms remain significantly subdominant at all scales.

From these plots, in which we have zoomed in on the y-axis scale, we can also more clearly see the imprint left by the BAO in the dipole of the cross-correlation function, at around $100\,h^{-1}$Mpc

Although the models employed are linear, and therefore known to be inaccurate on small scales, they nonetheless provide an important indication: at small separations, the gravitational potential and light-cone effects can no longer be neglected in precise modelling of the dipole. In particular, the contributions from the potential and the light-cone become comparable in amplitude at these scales.

4.8 Dipole signal on small scales

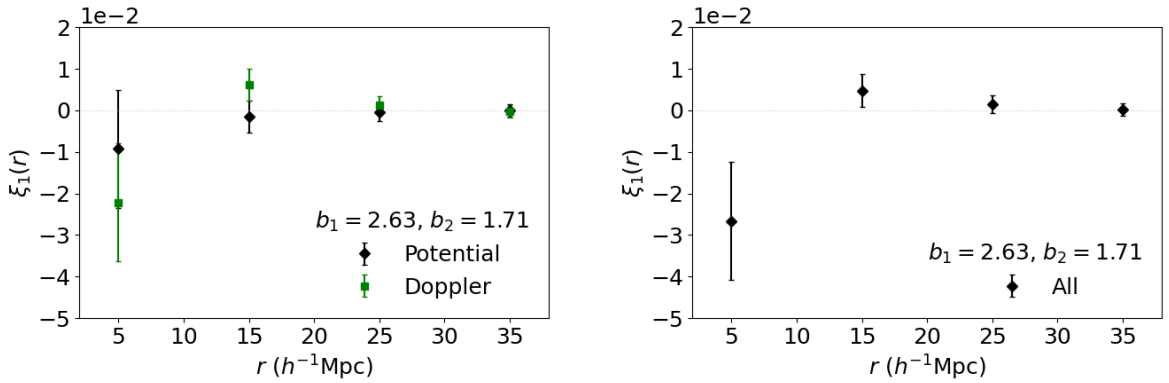

Figure 4.27: Different contribution to the dipole for the cross-correlation between H_{3200} and H_{800} at low scales. Left panel: potential only (green square) and relativistic Doppler-only (black diamonds) term contributions. Right panel: full dipole, thus including potential, relativistic Doppler, ISW, wide-angle, lensing and evolution effects.

Everything we have presented so far, and which, it bears recalling, holds true on large scales, within the linear regime, confirms the state of the art regarding the dipole signal and its main contributors (Breton et al., 2019). Namely, the signal is very weak and predominantly dominated by relativistic Doppler effects. These analyses also concur that the models presented here are unable to accurately describe the dipole signal on very small scales. Although our primary goal was to study the behaviour of the dipole on large scales, gaining a comprehensive picture of its behaviour across all scales proves

crucial. In Figure 4.27 we show the dipole signal for the cross-correlation between H_{3200} and H_{800}. The left panel displays the potential and relativistic Doppler effects separately. As we can see, the dipole tends to approach zero[2] on scales larger than $\sim 25\,h^{-1}$Mpc. Around scales of $15 - 20\,h^{-1}$Mpc, the dipole signal remains above zero and appears to increase going at smaller scales. Below these scales, we observe a sign reversal of the dipole, where the nonlinear regime becomes significant. The comparison with the plot on the right drives this point home even further. Linear models simply fail to predict or explain this behaviour.

Furthermore, at scales below $\sim 20\,h^{-1}$Mpc, the signal arising from the gravitational potential no longer appears negligible, when compared to the relativistic Doppler term, at least within the error bars, and can even become the dominant contribution. This finding is also supported by the current state of the literature. Breton et al. (2019) showed that the dipole exhibits a sign reversal below approximately $25\,h^{-1}$Mpc. Similarly, Lepori et al. (2025) have demonstrated how the dipole turns negative on small scales, and how light-cone effects cannot simply be brushed aside at these scales, clearly at odds with linear models.

Both the relativistic Doppler and gravitational contributions exhibit this behaviour, which is well documented in the literature and clearly visible in Figure 4.27.

Regarding nonlinear effects on small scales, it is known that non-Gaussian features arise from structure formation. One of the most prominent nonlinear phenomena is the FoG effect, caused by the peculiar motions of objects within virialised gravitational potential wells. While linear theory assumes that peculiar velocities and gravitational potential perturbations remain small, on smaller scales the motions of objects, such as galaxies within clusters, are driven by far stronger gravitational interactions than those at large scales. The contribution of these peculiar velocities to the observed redshift can easily outweigh that due to cosmic expansion.

A very recent study by Dam and Bonvin (2025) also shows that this sign inversion is indeed present, and that neither light-cone effects nor wide-angle effects can be neglected when analysing the dipole signal on small scales. These authors primarily focus on improving the modelling of the dipole originating from the gravitational potential at these scales. In this regard, their study attempts to model the dipole down to very small scales, and validate their predictions on the RayGalGroupSims.

Their model succeeds in explaining the sign inversion in the dipole on these tiny scales quite well, although it still fails in fully capturing the dipole trend around the $20\,h^{-1}$Mpc scale, where the signal appears to remain positive.

A more accurate modelling of the potential on small scales calls for, at least, the inclusion of an additional component, known as the one-halo term, which arises from the

[2]Recall that the dipole due to the relativistic Doppler effect, on large scales, is of the order of, or less than, 10^{-3}, and that from the potential is of the order of, or less than, 10^{-4}.

internal dynamics of dark matter haloes and becomes significant at these scales. This contribution effectively boosts the large-scale potential (which is described by spacetime perturbation theory and valid in the linear regime). The LSS provides a smooth, weaker potential, upon which local deviations, caused by the potentials of the dark matter haloes themselves, are superimposed. When observing the LSS, this local contribution cancels out, since the local potentials of widely separated objects are uncorrelated and thus do not contribute to the correlation function. However, the situation can quickly become quite complex at small scales, and as these authors point out, further refinements of the models will be necessary in the near future.

4.9 A statistical analysis of the dipole

4.9.1 Likelihood

Inference is the process by which we can deduce the characteristics of a population through the observation and analysis of a sub-sample, thereby reconstructing the probability distribution based on observed events and quantifying the degree of plausibility of a hypothesis (or model) in light of the experimental data. We consider a probability distribution as a function of the model parameters for a fixed experimental outcome. This distribution is referred to as the likelihood function. Let A and B be two independent subsets of the event space, such that

$$\mathcal{P}(A \cap B) = \mathcal{P}(A)\mathcal{P}(B), \tag{4.23}$$

where $P(A, B) = \mathcal{P}(A \cap B)$ is referred to as the joint probability. Let us consider an observable random variable X which follows a probability distribution $f(X, \theta)$, depending on a set of unknown theoretical parameters $\theta = (\theta_1, ..., \theta_M)$: this function is the conditional probability of observing the variabile X given the parameters θ. If the measurement is repeated N times, yielding N independent observations X_1, ..., X_N, the joint probability of observing $X_1 \in \mathrm{d}X_1$, $X_2 \in \mathrm{d}X_2$, and so on, is given by:

$$\mathcal{P}(X_1, X_2, ..., X_N) = \prod_{i=1}^{N} f(X_i|\theta)\,\mathrm{d}X_i. \tag{4.24}$$

Since inference is carried out after the data analysis has been completed, the X_i are considered fixed, and the likelihood depends only on the model parameters. It is then simply written as the joint probability distribution, thus

$$\mathcal{L}(X, \theta) = f(X_1, \theta)\, f(X_2, \theta) \ldots f(X_N, \theta). \tag{4.25}$$

It is logical to assume as the most plausible values for the set of parameters $\theta = (\theta_1, \theta_2, ..., \theta_M)$ the ones that maximise the likelihood. The general method is therefore based on solving the following system of equations:

$$\frac{\partial \mathcal{L}(\theta)}{\partial \theta_j} = 0. \tag{4.26}$$

Let us suppose we have a random variable X, whose Gaussian (normal) probability distribution is

$$G(\overline{x}, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X-\overline{x})^2}{2\sigma^2}}, \tag{4.27}$$

where $\overline{x}$ e σ^2 represent the mean value and the variance, respectively.

Thus, taking N Gaussian distributed variables X, the likelihood takes the following form:

$$\mathcal{L}(X_1, ..., X_N, \theta) \propto \prod_{i=1}^{N} \exp\left\{-\frac{[X_i - \overline{x}_i(\theta)]^2}{2\sigma_i^2}\right\}. \tag{4.28}$$

It is often convenient to maximise the logarithm of the likelihood: this is entirely valid because the logarithm is a monotonically increasing function. Taking the logarithm of the expression above yields:

$$-2\ln \mathcal{L}(X_1, ..., X_N, \theta) \propto -2\sum_{i=1}^{N}\left\{-\frac{[X_i - \overline{x}_i(\theta)]^2}{2\sigma_i^2}\right\} = \chi^2(\theta). \tag{4.29}$$

Therefore, in the case where the Gaussian distribution can be assumed, maximising the likelihood is equivalent to minimising the function χ^2.

4.9.2 Bayesian statistics

Let us interpret the probability as the degree of plausibility of a proposition. Bayesian statistics involves inferring the probability distribution from observed frequencies, essentially the inverse of the classical (frequentist) approach. Fundamentally, the Bayesian framework relies on specifying, a priori, a (subjective) probability for a given event based on prior knowledge, whether empirical or theoretical, and then updating the credibility of that hypothesis in light of the data, leading to the so-called posterior probability. In practical terms, one assumes that a given model, described by a set of parameters θ, is valid, and seeks to determine the probability distribution over this parameter space in light of the observed data X.

Bayes' theorem (Bayes, 1763) states that the posterior probability can be written as

$$\mathcal{P}(\theta|X) = \frac{\mathcal{P}(X|\theta)\,\mathcal{P}(\theta)}{\mathcal{P}(X)}, \tag{4.30}$$

where $\mathcal{P}(X|\theta) = \mathcal{L}(X,\theta)$ is the likelihood. The function $\mathcal{P}(\theta)$ is known as the prior probability and represents prior knowledge. The term $\mathcal{P}(X)$ is a normalisation factor, also called evidence, which reads

$$\mathcal{P}(X) = \int \mathcal{P}(X|\theta)\,\mathcal{P}(\theta)\mathrm{d}\theta. \tag{4.31}$$

A possible agnostic choice consists of adopting uniform priors, that is, priors of the form $\mathcal{P}(\theta) = \text{const}$ in the full domain, which assign equal probability to all values of θ within a given interval. In this case, the posterior becomes proportional to the likelihood, $P(\theta|X) \propto \mathcal{L}(\theta)$, and maximising the posterior is therefore equivalent to maximising the likelihood.

4.9.3 Markov Chain Monte Carlo

Although Bayes' theorem theoretically enables the application of Bayesian statistics in all circumstances, its analytical solution is often intractable, except in cases where the likelihood takes a Gaussian form, see Eq. (4.28). However, with the exponential growth of computational power and numerical simulations, several techniques have been developed to tackle Bayesian inference problems. One of the most widely adopted approaches, particularly in astrophysical and cosmological applications, is the MCMC method. This class of algorithms generates sequences of points (the so-called "chains") in the parameter space, whose density is proportional to the posterior probability distribution of interest, $P(\theta|X)$. The method relies on two fundamental concepts:

- Markovianity: the next state depends only on the current one;
- Ergodicity: the empirical distribution of the samples converges to the target distribution.

We are interested in Markov chains that admit a unique invariant distribution, namely the equilibrium distribution. This is typically reached after an initial transient phase, commonly referred to as the *burn-in* period.

One of the most widely used MCMC algorithms is the Metropolis–Hastings algorithm. Given a current state θ_t, it proposes a new state θ' using a proposal distribution $q(\theta'|\theta_t)$, and accepts it with probability

$$\alpha = \min\left(1,\, \frac{\mathcal{P}(\theta'|X)q(\theta_t|\theta')}{\mathcal{P}(\theta_t|X)q(\theta'|\theta_t)}\right). \tag{4.32}$$

If accepted, $\theta_{t+1} = \theta'$, otherwise, the chain remains at θ_t (Hastings, 1970). In the original Metropolis algorithm, the proposal is symmetric, $q(\theta'|\theta_t) = q(\theta_t|\theta')$, simplifying the acceptance ratio (Metropolis et al., 1953).

After the burn-in period, the remaining samples are assumed to be representative of the posterior. These can be used to estimate expectations and other statistical quantities under the posterior model.

4.9.4 Testing GR with MCMC sampling

Finally, we perform a posterior analysis using MCMC, trying to constrain the linear growth rate f and the two population biases from the dipole signal driven by relativistic Doppler, gravitational potential and light-cone effects (Figure 4.31), even if these last two terms are negligible at large scales with respect to the relativistic Doppler contribution and compare the results with those obtained from the monopole fit. Testing GR on large scales has become, in fact, one of the foremost objectives in modern cosmology, alongside probing models of modified gravity.

A key test in this context concerns the linear growth rate, f, and its well-known dependence predicted by GR (Linder, 2005): this allows a model-independent quantification of gravitational modifications through the introduction of the growth index, γ, which captures any deviations from the GR prediction.

Defining the priors

For the monopole of the 2PCF of the H_{3200}, H_{1600}, and H_{800} catalogues, the model considered is the dispersion model (Kaiser, 1987; Peacock and Dodds, 1996). The free parameters in this model are $f\sigma_8$, $b\sigma_8$ and σ_v, namely the linear growth rate and the bias of the population multiplied by σ_8, and the velocity dispersion, respectively.

Parameters	Uniform Prior
$f\sigma_8$	[0, 1]
$b_{\mathrm{H}_{3200}}\sigma_8$	[0, 2]
$b_{\mathrm{H}_{1600}}\sigma_8$	[0, 2]
$b_{\mathrm{H}_{800}}\sigma_8$	[0, 2]
σ_v	[0, 1000] $\mathrm{km\,s^{-1}}$

Table 4.4: Adopted priors on the sampling parameters for the monopole fitting.

Instead, in the case of the dipole models, the free parameters are directly the linear growth rate f and the biases of the two populations, b_1 and b_2. In all cases, the priors are set to be flat and allowed to vary over a broad range of values. Specifically, Table 4.4

shows prior values for the monopole fit with the dispersion model, while Table 4.5 shows prior values for the dipole fit with the linear model including light cone, evolution, gravitational potential and relativistic Doppler effects.

Parameters	Uniform Prior
f	[0.2, 1.2]
$b_{\mathrm{H}_{3200}}$	[1, 3]
$b_{\mathrm{H}_{1600}}$	[1, 3]
$b_{\mathrm{H}_{800}}$	[0.5, 2.5]

Table 4.5: Adopted priors on the sampling parameters for the dipole fitting.

Regarding the priors for the dipole fit, these are again flat and allowed to vary between 0.2 and 1.2 for f, and between 1 and 3 for the biases of the two populations.

Setting the likelihood

Once the model parameter priors are defined, we aim to compare the observed or simulated data with our theoretical predictions. For this purpose, it is necessary to specify a likelihood function, that is, the probability of observing the experimental data given a particular set of model parameters. In cosmology, as in our case, the likelihood is often assumed to be Gaussian, and there are several good reasons for this choice:

- Gaussianity of large-scale density fluctuations: on large cosmological scales, the density field fluctuations are well described by a Gaussian random field, which further justifies the use of a Gaussian likelihood.
- Central limit theorem: if the observed data (for example, the value of the dipole or the correlation function at a given scale) is obtained as an average or sum of a large number of independent, stochastic contributions, as happens in galaxy clustering, where many effects add up, then the central limit theorem ensures that the resulting distribution tends towards a Gaussian, regardless of the initial distribution of the individual contributions.

The consequence of this assumption is that maximising a Gaussian likelihood is computationally straightforward and corresponds to minimizing the χ^2, as discussed earlier.

It is important to note that the covariance should be calculated only once, based on the fiducial cosmology, as done in CBL, and not recalculated during the MCMC[3],

[3]This is beacuse doing so result in Fisher information content that violates the Cramér-Rao bound.

due to independent, but artificial, information arising from the parameter-dependent covariance matrix (Carron, 2013), which tends to cause an overestimation of the amount of information present in the data.

Maximising and sampling the posterior

MCMC algorithms are widely used for sampling from complex probability distributions. However, as the dimensionality of the parameter space increases, the computational cost of thoroughly exploring the space becomes significant. For this reason, we first perform a maximum a posteriori (MAP) estimation. This allows us to start the sampling close to the peak of the posterior distribution, thereby avoiding long burn-in periods if the initial positions are chosen sensibly.

MAP optimisation in the CBL uses the Nelder–Mead simplex method, which operates on a set of $k + 1$ points, that form a simplex in an k-D space: for example, a triangle in 2D or a tetrahedron in 3D. The method seeks a local optimum by iteratively replacing the worst vertex with a new point reflected through the centroid of the remaining vertices. If the reflection improves the function value significantly, the algorithm attempts to stretch the simplex further along that direction. If the reflection isn't an improvement, the simplex is drawn closer to better points to avoid overshooting tight valleys.

Termination is usually based on the standard deviation of function values at the simplex vertices falling below a tolerance, or when the simplex shrinks sufficiently.

Although the Nelder–Mead algorithm is an effective first step in locating the peak, a full exploration of the posterior distribution, including parameter uncertainties, requires MCMC methods.

MCMC sampling is performed following the MAP estimation. The sampling strategy incorporates the following elements:

- Thinning: Since each new sample depends on the previous one, successive points can be highly correlated, often resulting in very similar values. Discarding a certain number of samples helps reduce this auto-correlation, thereby yielding more robust parameter estimates. However, it is generally not recommended to discard too many samples, as this inevitably leads to some loss of information. In this work, we adopt a conservative thinning factor of 10.

- Burn-in: this involves discarding the initial portion of the chain to allow it sufficient time to converge. At first, the chain begins from an arbitrary position, often far from the maximum of the posterior distribution, and requires some iterations to "slow down" and reach the region of interest. However, performing a MAP estimate beforehand mitigates this issue considerably. In our case, a burn-in length of 200 proved sufficient to ensure that the chain was not excessively long.

Validation of General Relativity constraints

We now present a final discussion in light of the results obtained after MCMC sampling. We performed a fit of the monopole by considering jointly the three different catalogues: H3200, H1600, and H_{800}. These results are compared with the one obtained by the dipole fitting. It may also be of interest to compare the dipole fitting with monopole fitting for the cross-correlation, using a Kaiser term adapted for two populations with different biases, to strengthen the reliability of the analysis, as shown by Gaztanaga et al. (2017).

From theoretical expectations, we know that[4]

$$f\sigma_8 \simeq 0.47$$

$$\sigma_8 \simeq 0.801$$

$$b_{\mathrm{H}_{3200}} \simeq 2.63$$

$$b_{\mathrm{H}_{1600}} \simeq 2.09$$

$$b_{\mathrm{H}_{800}} \simeq 1.71$$

The plots with contour lines represent regions of equal posterior probabilities distributions of each pair of model parameters estimated from MCMC sampling. The histograms along the diagonal show the 1D marginal distributions of each individual parameter, i.e., the probability of that parameter regardless of the others. Specifically, the contours outline the regions of probability density containing 68% and 95% of the joint probability. These plots help reveal correlations or degeneracies between parameters: if the contours are elongated or tilted, it means that the two parameters are correlated. All theoretical predictions, shown as blue lines in Figures 4.28 to 4.30, lie within the 95% confidence region, indicating a satisfactory level of agreement between the theoretical predictions and the data. The parameter σ_v, on the other hand, is not directly predicted, and is considered as a nuisance parameter. In fact, our primary focus lies on the value of $f\sigma_8$, which consistently falls within 2σ in all configurations considered, supporting the overall consistency of the measurements with theoretical expectations.

[4]All values are computed at redshift $z = 0.341$.

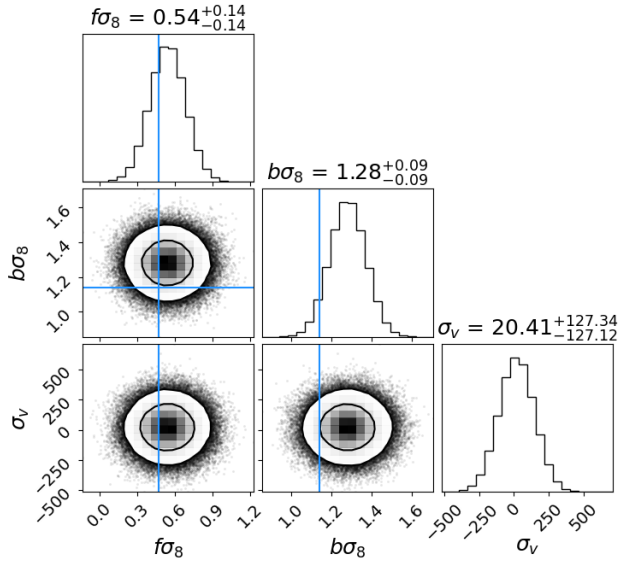

Figure 4.28: The 1- and 2-dimensional posterior distribution of the derived parameters $f\sigma_8$, $b\sigma_8$ and σ_v for the monopole of the 2PCF of H_{800}, using the dispersion model. The contour areas correspond respectively to the 68% and 95% joint two-parameter confidence levels.

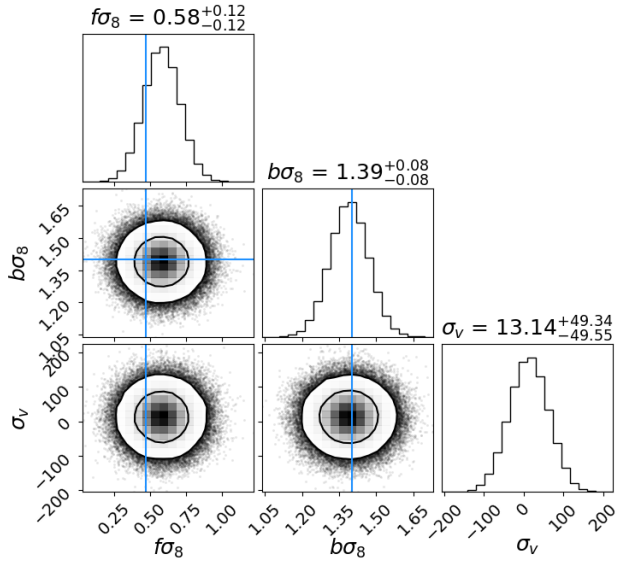

Figure 4.29: The 1- and 2-dimensional posterior distribution of the derived parameters $f\sigma_8$, $b\sigma_8$ and σ_v for the monopole of the 2PCF of H_{1600}, using the dispersion model. The contour areas correspond respectively to the 68% and 95% joint two-parameter confidence levels.

Subsequently, we performed a fit of the dipole signal, accounting for contributions from the relativistic Doppler effect, gravitational potential, light cone, and evolution effects (although the latter three are clearly subdominant), on large scales, considering two representative cross-correlations: H_{3200}–H_{1600} and H_{3200}–H_{800}. In Figure 4.31, we show the data points (diamonds), overlaid with the theoretical model, computed by fixing all parameters, as done in previous sections (black solid lines), alongside the corresponding model predictions obtained by allowing the parameters to vary (purple lines).

The known issues affecting the model on small and intermediate scales have already been discussed and will not be repeated here; nevertheless, their presence cannot be overlooked, as they inevitably weaken the robustness of our conclusions. The best-fit models do not significantly improve the agreement with the data (in the first case, the purple and black curves are almost indistinguishable), thus they yield estimates for f and for the biases of the two populations that are fully consistent with the fiducial cosmological model, as shown in Figure 4.31. All recovered values fall well within one standard deviation of the theoretical predictions, underscoring their excellent agreement with the fiducial model.

These results should be regarded as preliminary, providing a proof of concept for what might be achievable, with real data sets, possibly covering larger volumes of the Universe and featuring precise redshift measurements.

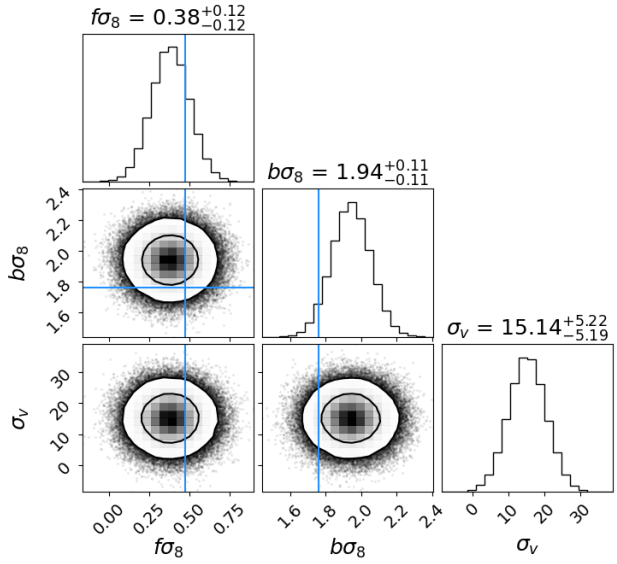

Figure 4.30: The 1- and 2-dimensional posterior distribution of the derived parameters $f\sigma_8$, $b\sigma_8$ and σ_v for the monopole of the 2PCF of H_{3200}, using the dispersion model. The contour areas correspond respectively to the 68% and 95% joint two-parameter confidence levels.

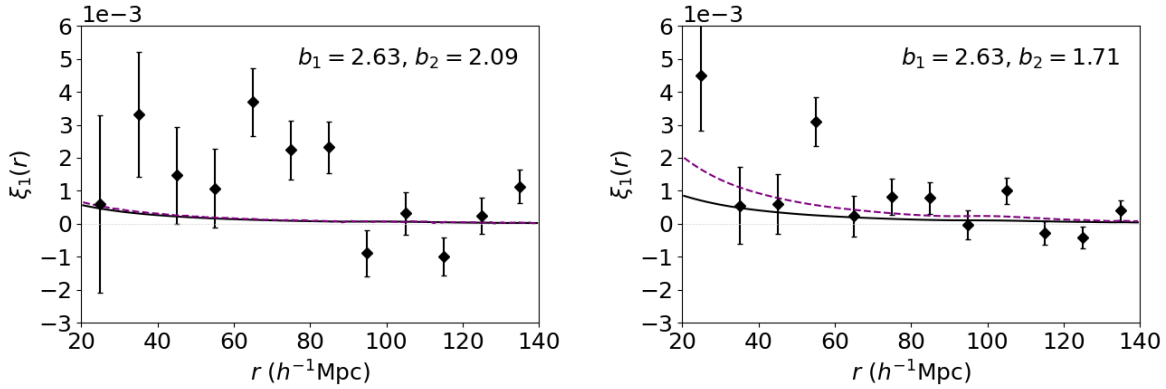

Figure 4.31: Contribution to the dipole from the coupling between gravitational potential, relativistic Doppler and light-cone effect, for the cross-correlation between H_{3200} and H_{1600} (top panels) and H_{3200} and H_{800} (bottom panels). Black solid lines show the model as derived from theory, while purple dashed lines show the fitting curve after a MCMC analysis.

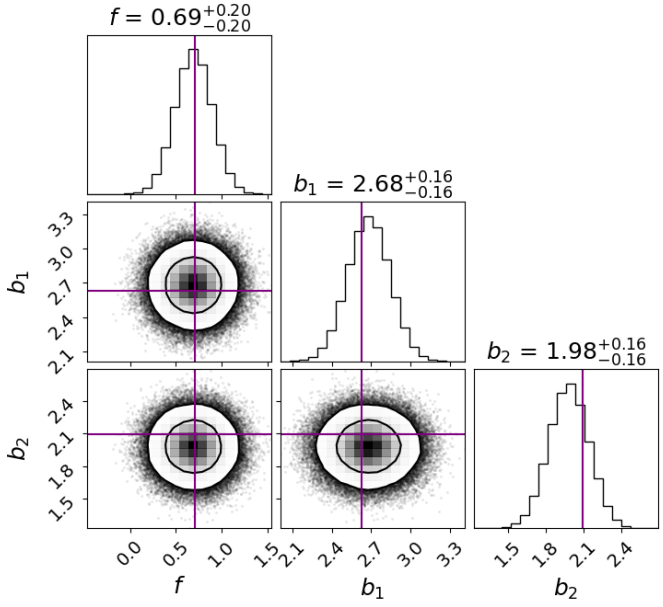

Figure 4.32: The 1- and 2-dimensional posterior distribution of the derived parameters $f\sigma_8$, $b\sigma_8$ and σ_v for the dipole of the 2-point cross-correlation of H_{3200}-H_{800}, using the linear model including light-cone, evolution, gravitational potential and relativistic Doppler term. The contour areas correspond respectively to the 68% and 95% joint two-parameter confidence levels.

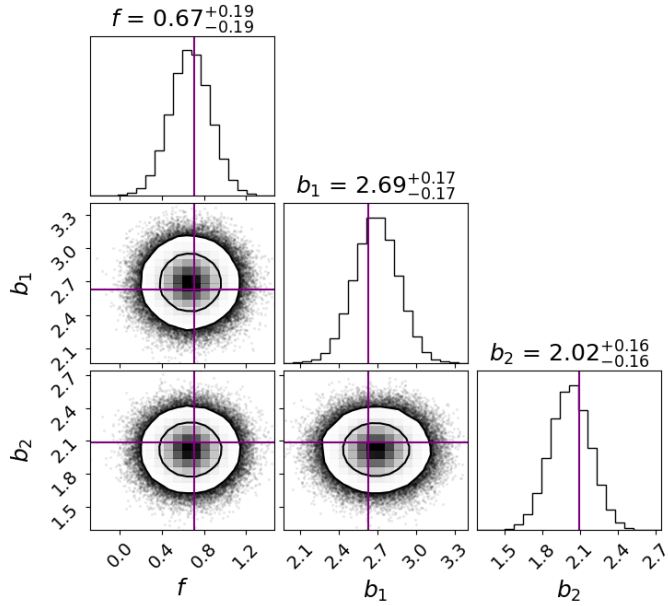

Figure 4.33: The 1- and 2-dimensional posterior distribution of the derived parameters $f\sigma_8$, $b\sigma_8$ and σ_v for the dipole of the 2-point cross-correlation of H_{3200}-H_{1600}, using the linear model including light-cone, evolution, gravitational potential and relativistic Doppler term. The contour areas correspond respectively to the 68% and 95% joint two-parameter confidence levels.

4.10 Additional considerations

Now that we have reached the conclusion of this thesis, it is pertinent to take stock of the main challenges encountered or that may arise in such an analysis. When attempting to detect a very subtle signal, such as the dipole in the large scale 2-point cross-correlation function, there are numerous and diverse factors that must be carefully considered.

- One of the principal challenges lies in the necessity of acquiring data over large volumes, whether real or simulated. A large data set is crucial to achieve robust statistics and to maximise the S/N, particularly when aiming to isolate subdominant contributions on large scales. However, handling such amounts of data inevitably entails significant computational costs, that must be carefully managed. In this context, it is worth noting that we have observed the dipole signal to be more pronounced in cross-correlations between populations with a greater difference in bias. Therefore, it may be advantageous, at least initially, to focus on these configurations to optimise the analysis.

- We have also considered, given that our work is based on a simulation, redshift

values free of observational errors. Naturally, this is an idealisation, that does not hold in real galaxy surveys, because redshift measurements are subject to uncertainties arising from instrumental limitations, and other systematic effects. Such errors can blur the signal and introduce biases, thereby complicating the interpretation of clustering statistics and the extraction of cosmological parameters.

- Another critical issue stems from the limitations of the models on small scales. The inability of linear models to accurately capture the behaviour of the dipole at these scales is well documented in the literature. Despite ongoing advancements, modelling this regime remains challenging, particularly in the quasi-nonlinear region where the dipole appears to linger slightly above zero (Dam and Bonvin, 2025).

- Our analysis has been based on dark matter N-body simulations. While this approach is convenient for a first validation of the pipelines, it proves insufficient when confronting with real observational data. In such cases, a range of additional factors, entirely neglected here, must be carefully considered. For instance, one should consider reliable catalogue completenesses and purities. These quantify the fraction of genuine astronomical objects successfully detected, and the fraction of detections that correspond to real sources, respectively, thereby directly affecting the reliability of any derived statistical measurements. A catalogue with low completeness risks omitting a substantial portion of the underlying population, thereby biasing number counts, redshift distributions, and ultimately the inferred cosmological parameters, if not properly modelled. Conversely, low purity indicates contamination by spurious detections or misclassified sources, which can dilute the signal and distort measured correlations. Furthermore, selection effects, both implicit and explicit, play a pivotal role in shaping the observed sample. These encompass instrumental limitations, survey geometry, magnitude thresholds and redshift-dependent detection efficiencies. If neglected, such effects can imprint artificial patterns onto the data or suppress real cosmological signals. It is therefore crucial to quantify and, where feasible, correct for these factors. This typically involves constructing completeness and purity selection functions, validating mock catalogues, and incorporating such selection functions into theoretical models. These foundational considerations have been extensively discussed in cosmological and astrophysical literature, e.g. Aguena and Lima (2018) and Datrier and Hendry (2025).

All these aspects will, of course, need to be thoroughly addressed, to ensure greater robustness in future conclusions. Only by carefully accounting for these challenges can we hope to extract reliable cosmological information from forthcoming data sets. This

will ultimately enable more stringent tests of fundamental physics, including GR on the largest scales.

Chapter 5

Discussion and conclusions

In this chapter, we revisit the work undertaken, its underlying premises, the methods employed, the results obtained, and the final conclusions, including some reflections with a view to future developments.

5.1 Overview

One of the key scientific goal of modern cosmology is to detect anisotropies in the LSS with high precision, thereby providing a stringent test of GR on cosmological scales. Clustering statistics serve as an indispensable tool in this endeavour, enabling tighter constraints on GR by probing the LSS of the Universe (Yoo, 2010; Bonvin and Durrer, 2011). While the bulk of analyses concentrate on even multipoles dominated by standard RSD, relativistic, wide-angle, and evolutionary effects introduce an antisymmetric component, manifesting as odd multipoles in the cross-correlation between different populations of objects.

Incorporating relativistic RSD effects is essential, as conventional RSD alone cannot produce a dipole signal (Bonvin et al., 2014; Gaztanaga et al., 2017; Breton et al., 2019). Though subtle and challenging to model, these effects pave the way for more robust tests of GR, and offer deeper insights into the nature of dark matter and dark energy. The dipole signal arises from several contributions: some are geometric in origin, such as wide-angle effects, while others require a fully relativistic treatment at linear order, including the gravitational potential, relativistic Doppler shifts, lensing, and the ISW effect. Hence, it is crucial to adequately model and isolate these effects on large scales. Crucially, these relativistic contributions depend on the linear growth rate of cosmic structures, which is intimately connected to GR through the growth index parameter, γ.

This thesis aimed to validate cosmological analyses beyond classical dynamical and geometric distortions, by incorporating the full range of effects, including relativistic contributions, which affect redshift measurements and manifest as a dipole in the 2PCF

(Bonvin and Durrer, 2011). In particular, we tested models at large and intermediate scales which will be used in the near future to constrain GR and investigate alternative gravity models.

5.2 Main findings

To carry out our analysis, we used a set of dark matter halo simulations from the RayGalGroupSims, comprising approximately 1.2×10^7, objects within a volume of $2625\,(h^{-1}\mathrm{Mpc})^3$, in order to assess the dipole contribution, and to test the concordance between theoretical models and the measured signal. The RayGalGroupSims data were binned logarithmically by mass, enabling the cross-correlation of distinct catalogues.

The work presented in this thesis involved the implementation of codes to compute the multipoles of the cross-correlation function, including the odd multipoles, using two separate random catalogues, as required by the Landy-Szalay estimator (Landy and Szalay, 1993).

Poissonian errors and uncertainty estimates using the jackknife technique were also implemented. Both the codes and the RayGalGroupSims were subsequently validated through auto-correlation tests.

The second part of the thesis involved the implementation of linear models to describe the dipole signal on large scales. These models incorporate terms accounting for relativistic Doppler effects, light-cone effects, wide-angle contributions, gravitational potential, the ISW effect, Shapiro time delay, lensing, and evolutionary effects. Each contribution is implemented as an individual term, to ensure maximum flexibility for future applications.

The main results obtained are as follows:

- We found that the dipole is overwhelmingly dominated by the relativistic Doppler term at all scales and for all the cross-correlation considered, while all other contributions are subdominant, both in the full signal, and with respect to other contribution such as the gravitational potential.

- Nevertheless, we attempted to isolate the contributions from the gravitational potential for three cross-correlation configurations. To perform this analysis on these catalogues, it was necessary to use random catalogues more than 10 times times larger than the data catalogues. We was not able anyway to detect a distinct signal from gravitational potential term at large scales.

- S/N remains at approximately the 3σ level in nearly all configurations considered, except for two cases where it is respectively significantly higher and lower than this value, likely due to systematic errors in the analysis.

- The agreement between models and measurements for the total dipole is fairly good, with a reduced $\chi^2_r \sim 1$, although further refinements are necessary. Isolating the contribution of the gravitational potential proved to be particularly challenging, and no definitive conclusions can be drawn at these scales.

- The model–data agreement worsen on scales below $60\,h^{-1}$Mpc, as also confirmed by Breton et al. (2019).

- At smaller scales (below $30\,h^{-1}$Mpc), the agreement further deteriorates, mainly due to the increasing importance of nonlinear effects, as also outlined by Giusarma et al. (2017) among others. Moreover, at these scales, the gravitational potential contribution is no longer subdominant and must be accounted for: even if for the cross-correlation at nonlinear scales shown here, it seems to be of the same order of the relativistic Doppler term, for many configurations, it has shown to dominate over it at nonlinear scales (Lepori et al., 2025; Dam and Bonvin, 2025).

- A preliminary MCMC fit of the dipole model, including light-cone effects, evolutionary contributions, gravitational potential, and the relativistic Doppler effect, demonstrates, at an early-stage, agreement between the data and theoretical predictions from the standard ΛCDM model. Although we are not yet in a position to draw definitive conclusions, as the models require significant refinement, particularly on intermediate and small scales, testing GR on large scales is, indeed, feasible.

5.3 Future perspectives

As we have seen, the path is complex but opens up numerous possible avenues. Several promising directions exist for future work. Firstly, refining theoretical modelling on nonlinear scales is essential. For example, Breton et al. (2019) demonstrated that nonlinear mapping effects contribute to the dipole signal at the order of 10^{-4}, a level that cannot be ignored when attempting to isolate the gravitational potential term.

Increasing the number of objects and improving simulation resolution would also help isolate otherwise elusive dipole contributions. Extending the fitting procedure to various population configurations and different dipole components would enable more reliable and stringent tests of GR.

Additionally, developing simulations within modified gravity frameworks could provide valuable insights into model-data agreement under alternative theories, testing the validity of ΛCDM models over $f(R)$ modified gravity theory.

Furthermore, populating dark matter haloes with galaxies or clusters would yield a more robust framework better suited for comparison with forthcoming observations from

Euclid and to test the baryon effects on clustering statistics and on the dipole signal of cross-correlation function in particular. Accurately simulating baryon dynamics is challenging, as it involves complex fluid processes such as collapse and structure formation, along with feedback mechanisms from AGN, galactic winds, and supernovae. These effects are likely to influence mainly small to intermediate scales. While computationally demanding to achieve the necessary precision, this area represents surely a worthwhile target for future investigation.

One natural direction for future work involves producing forecasts for ongoing or upcoming large-scale galaxy surveys, such as *Euclid* (Mellier et al., 2025), the Rubin Observatory Legacy Survey of Space and Time (LSST; Ivezić et al., 2019), or the prospective Wide-Field Spectroscopic Telescope (WST; Mainieri et al., 2024).

In summary, this study highlights both the potential and the challenges of using cross-correlation dipoles as probes of relativistic effects in the cosmic LSS. While several limitations remain, particularly concerning modelling accuracy at small scales and the extraction of subdominant contributions, this framework offers a promising basis for future applications to high-precision cosmological surveys.

To conclude, an important avenue for future work would be the application of the techniques, implemented and validated here, to actual observational data, for instance from *Euclid*. Analysing real data with the statistical tools, presented in this work, will offer an opportunity to directly test relativistic corrections and assess their significance in the context of contemporary cosmology.

Bibliography

T. M. C. Abbott et al. Dark Energy Survey Year 3 results: Cosmological constraints from galaxy clustering and weak lensing. *Phys. Rev. D*, 105(2):023520, 2022. doi: 10.1103/PhysRevD.105.023520.

A. G. Adam et al. DESI 2024 VI: cosmological constraints from the measurements of baryon acoustic oscillations. *JCAP*, 02:021, 2025. doi: 10.1088/1475-7516/2025/02/021.

P. A. R. Ade et al. Planck 2013 results. *A&A*, **571**:A1, 2014. doi: 10.1051/0004-6361/201321529.

S. A. Adil et al. Revisiting the concordance ΛCDM model using Gamma-Ray Bursts together with supernovae Ia and Planck data. *JCAP*, **08**:015, 2024. doi: 10.1088/1475-7516/2024/08/015.

N. Afshordi. Integrated sachs-wolfe effect in cross - correlation: The observer's manual. *Phys. Rev. D*, **70**:083536, 2004. doi: 10.1103/PhysRevD.70.083536.

N. Aghanim et al. Planck 2018 results. VI. Cosmological parameters. *Astron. Astrophys.*, 641:A6, 2020. doi: 10.1051/0004-6361/201833910. [Erratum: Astron.Astrophys. 652, C4 (2021)].

M. Aguena and M. Lima. Effects of Completeness and Purity on Cluster Dark Energy Constraints. *Phys. Rev. D*, 98(12):123529, 2018. doi: 10.1103/PhysRevD.98.123529.

Y. Akrami et al. Planck 2018 results. IX. Constraints on primordial non-Gaussianity. *Astron. Astrophys.*, 641:A9, 2020. doi: 10.1051/0004-6361/201935891.

S. Alam et al. Relativistic distortions in the large-scale clustering of SDSS-III BOSS CMASS galaxies. *Mon. Not. Roy. Astron. Soc.*, 470(3):2822–2833, 2017a. doi: 10.1093/mnras/stx1421.

S. Alam et al. The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample. *MNRAS*, 470(3):2617–2652, 2017b. doi: 10.1093/mnras/stx721.

C. Alcock and B. Paczynski. An evolution free test for non-zero cosmological constant. *Nature*, **281**:358–359, 1979. doi: 10.1038/281358a0.

J. M. Alimi et al. DEUS Full Observable ΛCDM Universe Simulation: the numerical challenge, 2012.

L. Amendola et al. Cosmology and fundamental physics with the Euclid satellite. *Living Rev. Rel.*, 21(1):2, 2018. doi: 10.1007/s41114-017-0010-3.

J. Antoniadis et al. A Massive Pulsar in a Compact Relativistic Binary. *Science*, 340: 6131, 2013. doi: 10.1126/science.1233232.

P. Arnalte-Mur and P. Norberg. Internal errors: a valid alternative for clustering estimates? In *Proceedings of the International Astronomical Union, IAU Symposium*, volume 306, pages 247–248. Cambridge University Press, 2014. doi: 10.1017/S1743921314013817.

J. M. Bardeen. Gauge invariant cosmological perturbations. *Phys. Rev. D*, **22**:1882–1905, 1980. doi: 10.1103/PhysRevD.22.1882.

J. D. Barrow and S. P. Growth of large-scale structure with a cosmological constant. *MNRAS*, **262**(3):717–725, 1993. doi: 10.1093/mnras/262.3.717.

B. A. Bassett and R. Hlozek. Baryon Acoustic Oscillations. 10 2009.

J. E. Bautista et al. The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: measurement of the BAO and growth rate of structure of the luminous red galaxy sample from the anisotropic correlation function between redshifts 0.6 and 1. *MNRAS*, 500(1):736–762, 2020. doi: 10.1093/mnras/staa2800.

T. Bayes. LII. An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, 53:370–418, 1763. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton.

E. Bertschinger. Cosmological dynamics: Course 1. In *Les Houches Summer School on Cosmology and Large Scale Structure (Session 60)*, pages 273–348, 1993.

F. Beutler et al. The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Testing gravity with redshift-space distortions using the power spectrum multipoles. *MNRAS*, 443(2):1065–1089, 2014. doi: 10.1093/mnras/stu1051.

C. Blake et al. The WiggleZ Dark Energy Survey: mapping the distance-redshift relation with baryon acoustic oscillations. *MNRAS*, 418:1707–1724, 2011. doi: 10.48550/arXiv.1108.2635.

C. Bonvin and R. Durrer. What galaxy surveys really measure. *Phys. Rev. D*, **84**: 063505, 2011. doi: 10.1103/PhysRevD.84.063505.

C. Bonvin, L. Hui, and E. Gaztanaga. Asymmetric galaxy correlation functions. *Phys. Rev. D*, **89**(8):083535, 2014. doi: 10.1103/PhysRevD.89.083535.

C. Bonvin et al. A case study for measuring the relativistic dipole of a galaxy cross-correlation with the Dark Energy Spectroscopic Instrument. *MNRAS*, 525(3):4611–4627, 2023. doi: 10.1093/mnras/stad2567.

P. Brax et al. Modifying Gravity at Low Redshift. *JCAP*, 04:032, 2010. doi: 10.1088/1475-7516/2010/04/032.

M. A. Breton. *Constructing observables in cosmology : towards new probes of the dark sector.* 2018.

M. A. Breton and V. Reverdy. Magrathea-pathfinder: a 3d adaptive-mesh code for geodesic ray tracing in n-body simulations. *Astron. Astrophys.*, 662:A114, 2022. doi: 10.1051/0004-6361/202142661.

M. A. Breton et al. Imprints of relativistic effects on the asymmetry of the halo cross-correlation function: from linear to non-linear scales. *Mon. Not. Roy. Astron. Soc.*, **483**(2):2671–2696, 2019. doi: 10.1093/mnras/sty3206.

J. Carron. On the assumption of Gaussianity for cosmological two-point statistics and parameter dependent covariance matrices. *A&A*, 551:A88, 2013. doi: doi.org/10.1051/0004-6361/201220538.

J. Casado. Linear expansion models vs. standard cosmologies: a critical and historical overview. *Astrophysics and Space Science*, **365**, 2020. doi: 10.1007/s10509-019-3720-z.

S. Castello et al. Testing the equivalence principle across the Universe: a model-independent approach with galaxy multi-tracing. 12 2024. doi: 10.1103/1my7-zklj.

E. Castorina and E. di Dio. The observed galaxy power spectrum in General Relativity. *JCAP*, 01(01):061, 2022. doi: 10.1088/1475-7516/2022/01/061.

E. Castorina and M. White. The Zeldovich approximation and wide-angle redshift-space distortions. *Mon. Not. Roy. Astron. Soc.*, 479(1):741–752, 2018. doi: 10.1093/mnras/sty1437.

T. Castro et al. Euclid preparation: L. calibration of the halo linear bias in Λ(v)cdm cosmologies. *Astron. Astrophys.*, 691:A62, 2024. doi: 10.1051/0004-6361/202451230.

A. Challinor and A. Lewis. The linear power spectrum of observed source number counts. *Phys. Rev. D*, **84**:043516, 2011. doi: 10.1103/PhysRevD.84.043516.

S. Colombi et al. Accurate estimators of power spectra in n-body simulations. *MNRAS*, 393:511, 2009. doi: 10.1111/j.1365-2966.2008.14176.x.

Dagum, L. and Menon R. OpenMP: an industry standard API for shared-memory programming. *Computational Science and Engineering*, 5(1):44–55, 1998. doi: 10.1109/99.660313.

L. Dam and C. Bonvin. Gravitational redshift from large-scale structure: nonlinearities, anti-symmetries, and the dipole. 6 2025.

L. Datrier and M. Hendry. Implementing a Robust Test of Galaxy Catalogue Completeness for Dark Siren Measurements of the Hubble Constant. 2 2025.

S. de la Torre et al. Euclid preparation. 3-dimensional galaxy clustering in configuration space. Part I. 2-point correlation function estimation. 1 2025.

F. W. Dyson et al. A Determination of the Deflection of Light by the Sun's Gravitational Field, from Observations Made at the Total Eclipse of May 29, 1919. *Philosophical Transactions of the Royal Society of London. Series A*, 220:291–333, 1920. doi: 10.1098/rsta.1920.0009.

A. Einstein and W. de Sitter. On the relation between the expansion and the mean density of the universe. *PNAS*, **18**(3):213–214, 1932. doi: 10.1073/pnas.18.3.213.

D. J. Eisenstein et al. Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies. *Astrophys. J.*, 633:560–574, 2005. doi: 10.1086/466512.

M. Y. Elkhashab et al. Euclid preparation. LXIX. The impact of relativistic redshift-space distortions on two-point clustering statistics from the Euclid wide spectroscopic survey. *Astron. Astrophys.*, 697:A85, 2025. doi: 10.1051/0004-6361/202452480.

J. L. Feng. Dark matter candidates from particle physics and methods of detection. *Ann. Rev. Astron. Astrophys.*, **48**:495–545, 2010. doi: 10.1146/annurev-astro-082708-101659.

D. J. Fixsen. The temperature of the cosmic microwave background. *ApJ*, **707**:916–920, 2009. doi: 10.1088/0004-637X/707/2/916.

P. a. a. Fosalba. The onion universe: all sky light-cone simulations in shells. *MNRAS*, 391:435, 2008. doi: 10.1111/j.1365-2966.2008.13910.x.

C. Franco et al. Probing large-scale structures with the 2-point function and the power spectrum: insights into cosmic clustering evolution. 2 2025.

A. Friedmann. Über die krümmung des raumes, 1922. Translated and republished as: On the Curvature of Space, *Proceedings of the Royal Society A*.

X. J. Gao. Gravitational lensing and shadow by a schwarzschild-like black hole in metric-affine bumblebee gravity. *Eur. Phys. J. C*, **84**(9):973, 2024. doi: 10.1140/epjc/s10052-024-13338-9.

E. Gaztanaga et al. Measurement of the dipole in the cross-correlation function of galaxies. *JCAP*, **01**:032, 2017. doi: 10.1088/1475-7516/2017/01/032.

E. Giusarma et al. Relativistic asymmetries in the galaxy cross-correlation function. 9 2017.

M. Hameeda et al. The large-scale structure formation in an expanding universe. *Int. J. Mod. Phys. A*, **37**(15), 2022. doi: 10.1142/S0217751X22500865.

A. J. S. Hamilton. Measuring Omega and the Real Correlation Function from the Redshift Correlation Function. *Astrophysical Journal Letters*, 365:L5, 1 1992. doi: 10.1086/186264.

A. J. S. Hamilton. Linear redshift distortions: a review. In *The evolving universe*, pages 185–275. Kluwer, Dordrecht, 1997. doi: 10.1007/978-94-011-4960-0_17.

A. J. S. Hamilton. Formulae for growth factors in expanding universes containing matter and a cosmological constant. *MNRAS*, **322**:419–425, 2001. doi: 10.48550/arXiv.astro-ph/0006089.

W. K. Hastings. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. 57(1):97–109, 1970. doi: 10.2307/2334940.

J. Hu and F. Wang. Hubble tension: The evidence of new physics. *Universe*, 2023. doi: 10.48550/arXiv.2302.05709.

E. Hubble. A relation between distance and radial velocity among extra-galactic nebulae. *PNAS USA*, **15**:168–173, 1929. doi: 10.1073/pnas.15.3.168.

Ivezić et al. LSST: From Science Drivers to Reference Design and Anticipated Data Products. *ApJ*, 873:44, 2019. doi: 10.3847/1538-4357/ab042c.

J. H. Jeans. The stability of a spherical nebula. *Philos. Trans. R. Soc. Lond. Series A*, **199**:1–53, 1902. doi: 10.1098/rsta.1902.0012.

N. Kaiser. Clustering in real space and in redshift space. *MNRAS*, **227**:1–21, 1987. doi: 10.1093/mnras/227.1.1.

N. Kaiser. Measuring gravitational redshifts in galaxy clusters. *MNRAS*, **435**:1278, 2013. doi: 10.1093/mnras/stt1370.

E. Keihänen et al. Estimating the galaxy two-point correlation function using a split random catalog. *Astron. Astrophys.*, 631:A73, 2019. doi: 10.1051/0004-6361/201935828.

M. Kerscher. Improving the accuracy of estimators for the two-point correlation function. *Astron. Astrophys.*, 666:A181, 2022. doi: 10.1051/0004-6361/202243632.

A. Kozmanyan et al. Deriving the hubble constant using planck and xmm-newton observations of galaxy clusters. *A&A*, **621**, 2019. doi: 10.1051/0004-6361/201833879.

S. D. Landy and A. S. Szalay. Bias and variance of angular correlation functions. *ApJ*, **412**:64–71, 1993. doi: 10.1086/172900.

G. Lemaître. Un univers homogène de masse constante et de rayon croissant, rendant compte de la vitesse radiale des nébuleuses extragalactiques, 1927. Translated and republished as: A Homogeneous Universe of Constant Mass and Increasing Radius Accounting for the Radial Velocity of Extra-Galactic Nebulae, *MNRAS*, **91**:483–490, 1931.

F. Lepori et al. Euclid: Relativistic effects in the dipole of the two-point correlation function. *Astron. Astrophys.*, 694:A321, 2025. doi: 10.1051/0004-6361/202452531.

J. Lesgourgues. The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview. 4 2011.

A. Lewis and A. Challinor. CAMB: Code for Anisotropies in the Microwave Background. Astrophysics Source Code Library, Feb. 2011.

A. Lewis, A. Challinor, and A. Lasenby. Efficient computation of CMB anisotropies in closed FRW models. *Astrophys. J.*, **538**:473–476, 2000. doi: 10.1086/309179.

E. V. Linder. Cosmic growth history and expansion history. *Phys. Rev. D*, 72:043529, 2005. doi: 10.1103/PhysRevD.72.043529.

A. Lue et al. Probing Newton's constant on vast scales: DGP gravity, cosmic acceleration and large scale structure. *Phys. Rev. D*, 69:124015, 2004. doi: 10.1103/PhysRevD.69.124015.

R. Maartens et al. Magnification and evolution biases in large-scale structure surveys. *JCAP*, 12(12):009, 2021. doi: 10.1088/1475-7516/2021/12/009.

V. Mainieri et al. The Wide-field Spectroscopic Telescope (WST) Science White Paper. 3 2024.

F. Marulli, A. Veropalumbo, and M. Moresco. Cosmobolognalib: C++ libraries for cosmological calculations. *Astron. Comput.*, **14**:35–42, 2016. doi: 10.1016/j.ascom.2016.01.005.

F. Marulli et al. Cosmology with clustering anisotropies: disentangling dynamic and geometric distortions in galaxy redshift surveys. *MNRAS*, 426:2566–2580, 2012. doi: 10.48550/arXiv.1203.1002.

F. Marulli et al. Redshift-space distortions of galaxies, clusters and AGN: testing how the accuracy of growth rate measurements depends on scales and sample selections. *Astron. Astrophys.*, 599:A106, 2017a. doi: 10.1051/0004-6361/201526885.

F. Marulli et al. Redshift-space distortions of galaxies, clusters, and AGN. Testing how the accuracy of growth rate measurements depends on scales and sample selections. *A & A*, 599:A106, 2017b. doi: 10.48550/arXiv.1505.01170.

T. Matsubara. The gravitational lensing in redshift-space correlation functions of galaxies and quasars. *Astrophys. J. Lett.*, **537**:L77, 2000. doi: 10.1086/312762.

P. McDonald. Gravitational redshift and other redshift-space distortions of the imaginary part of the power spectrum. *JCAP*, 2009. doi: 10.1088/1475-7516/2009/11/026.

Y. Mellier et al. Euclid. I. Overview of the Euclid mission. *Astron. Astrophys.*, 697:A1, 2025. doi: 10.1051/0004-6361/202450810.

N. Metropolis et al. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.*, 21:1087–1092, 1953. doi: 0.1063/1.1699114.

F. G. Mohammad et al. The VIMOS Public Extragalactic Redshift Survey (VIPERS). An unbiased estimate of the growth rate of structure at $\langle z \rangle = 0.85$ using the clustering of luminous blue galaxies. *Astron. Astrophys.*, 610:A59, 2018. doi: 10.1051/0004-6361/201731685.

B. P. Moster et al. A cosmic variance cookbook. *The Astrophysical Journal*, 731(2):113, 2011. doi: 10.1088/0004-637X/731/2/113.

S. Nadathur and W. J. Percival. An accurate linear model for redshift space distortions in the void-galaxy correlation function. *MNRAS*, 483(3):3472–3487, 2019. doi: 10.1093/mnras/sty3372.

P. Norberg et al. Statistical analysis of galaxy surveys - i. robust error estimation for 2-point clustering statistics. *MNRAS*, 396:19, 2009. doi: 10.1111/j.1365-2966.2009.14389.x.

J. A. Peacock and S. J. Dodds. Non-linear evolution of cosmological power spectra. *MNRAS*, 280:L19–L26, 1996.

J. A. Peacock et al. A measurement of the cosmological mass density from clustering in the 2df galaxy redshift survey. *Nature*, **410**:169–173, 2001. doi: 10.1038/35065528.

P. J. E. Peebles. *Large-Scale Structure of the Universe*. Princeton University Press, 1980.

P. J. E. Peebles and M. G. Hauser. Statistical analysis of catalogs of extragalactic objects. *ApJ Supplement Series*, 1974. doi: 10.1086/190312.

W. J. Percival and B. M. Schaefer. Galaxy peculiar velocities and evolution-bias. *Mon. Not. Roy. Astron. Soc.*, 385:78, 2008. doi: 10.1111/j.1745-3933.2008.00437.x.

R. L. Plackett. Karl Pearson and the Chi-Squared test. *International Statistical Review*, 51:59–72, 1983. doi: 10.2307/1402731.

R. V. Pound and G. A. Rebka. Apparent Weight of Photons. *Phys. Rev. Lett.*, 4:337–341, Apr 1960. doi: 10.1103/PhysRevLett.4.337.

S. Prunet et al. Initial conditions for large cosmological simulations. *Astrophys. J. Suppl.*, 178:179, 2008. doi: 10.1086/590370.

A. G. Riess et al. Observational evidence from supernovae for an accelerating universe and a cosmological constant. *AJ*, **116**:1009–1038, 1998. doi: 10.48550/arXiv.astro-ph/9805201.

T. Ronconi and F. Marulli. Cosmological exploitation of cosmic void statistics - new numerical tools in the cosmobolognalib to extract cosmological constraints from the void size function. *Astron. Astrophys.*, 607:A24, 2017. doi: 10.1051/0004-6361/201730852.

R. Scoccimarro. Redshift-space distortions, pairwise velocities and nonlinearities. *Phys. Rev. D*, 70:083007, 2004. doi: 10.1103/PhysRevD.70.083007.

A. Sneppen et al. Measuring the hubble constant with kilonovae using the expanding photosphere method. *A&A*, **678**, 2023. doi: 10.48550/arXiv.2306.12468.

R. S. Somerville et al. Cosmic variance in the great observatories origins deep survey. *The Astrophysical Journal Letters*, 600(2):L171–L174, 2004. doi: 10.1086/381541.

V. Tansella et al. The full-sky relativistic correlation function and power spectrum of galaxy number counts. part i: theoretical aspects. *JCAP*, **03**:019, 2018. doi: 10.1088/1475-7516/2018/03/019.

A. Taruya, T. Nishimichi, and S. Saito. Baryon Acoustic Oscillations in 2D: Modeling Redshift-space Power Spectrum from Perturbation Theory. *Phys. Rev. D*, 82:063522, 2010. doi: 10.1103/PhysRevD.82.063522.

R. Teyssier. Cosmological hydrodynamics with adaptive mesh refinement: a new high resolution code called RAMSES. *Astron. Astrophys.*, **385**:337–364, 2002. doi: 10.1051/0004-6361:20011817.

H. Totsuji and T. Kihara. The correlation function for the distribution of galaxies. *Publications of the Astronomical Society of Japan*, **21**:221–229, 1969.

C. G. Tsagas et al. Relativistic cosmology and large-scale structure. *Phys. Rept.*, 465: 61–147, 2008. doi: 10.1016/j.physrep.2008.03.003.

P. Valageas and N. Clerc. Redshift-space correlation functions in large galaxy cluster surveys. *Astron. Astrophys.*, 547(A100):19, 2012. doi: 10.1051/0004-6361/201219646.

D. Walsh, R. F. Carswell, and R. J. Weymann. 0957+561 A, B: twin quasistellar objects or gravitational lens? *Nature*, **279**:381–384, 1979. doi: 10.1038/279381a0.

R. Wojtak et al. Gravitational redshift of galaxies in clusters as predicted by general relativity. *Nature*, **477**:567–569, 2011. doi: 10.1038/nature10445.

J. Yoo. General Relativistic Description of the Observed Galaxy Power Spectrum: Do We Understand What We Measure? *Phys. Rev. D*, 82:083508, 2010. doi: 10.1103/PhysRevD.82.083508.

Y. B. Zeldovich. A hypothesis, unifying the structure and the entropy of the universe. *MNRAS*, **160**:1P–3P, 1972. doi: 10.1093/mnras/160.1.1P.