

UNIVERSIDADE DE SÃO PAULO  
INSTITUTO DE FÍSICA

# Estimadores do Espectro de Potência em Levantamentos de Galáxias

Lucas Frozza Secco

**Orientador: Prof. Dr. Luis Raul Weber Abramo**

---

Dissertação apresentada ao Instituto de Física da Universidade de São Paulo para a obtenção do título de Mestre em Ciências.

---

**Banca Examinadora:**

Prof. Dr. Luis Raul Weber Abramo (IF-USP)

Prof. Dr. Laerte Sodré Jr. (IAG-USP)

Prof. Dr. Miguel Boavista Quartin (IF-UFRJ)

São Paulo

2015

## FICHA CATALOGRÁFICA

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

Secco, Lucas Frozza

Estimadores do espectro de potência em levantamentos de galáxias. São Paulo, 2015.

Dissertação (Mestrado) — Universidade de São Paulo.  
Instituto de Física. Depto. Física Matemática

Orientador: Prof. Dr. Luis Raul Weber Abramo

Área de concentração: Cosmologia

Unitermos: 1. Física; 2. Cosmologia; 3. Astronomia; 4.  
Galáxias.

USP/IF/SBI-065/2015

UNIVERSIDADE DE SÃO PAULO  
INSTITUTO DE FÍSICA

# Power Spectrum Estimators in Galaxy Surveys

Lucas Frozza Secco

**Advisor: Prof. Dr. Luis Raul Weber Abramo**

---

Dissertation presented to the Institute of Physics of the University of São Paulo as a partial requirement for the title of Master of Science.

---

**Examination Committee:**

Prof. Dr. Luis Raul Weber Abramo (IF-USP)

Prof. Dr. Laerte Sodré Jr. (IAG-USP)

Prof. Dr. Miguel Boavista Quartin (IF-UFRJ)

São Paulo

2015



# Resumo

No presente trabalho, revisamos o bem estabelecido formalismo de formação de estruturas no universo e a fundamentação teórica do espectro de potência da matéria. Revisamos então a literatura atual sobre estimação do espectro de potência, focando nos métodos de análise de Fourier que possuem conexão com o formalismo da matriz de Fisher. Com a mesma motivação, desenvolvemos um conjunto de estimadores para levantamentos de galáxias com múltiplas espécies e testamos eficientemente nossas previsões em catálogos artificiais simplificados. Esse método é adequado para tratar levantamentos contendo diferentes espécies de traçadores da estrutura em larga escala, como galáxias, quasares, sistemas de absorção de Lyman- $\alpha$ , etc. Encontramos boa concordância entre nossas expectativas e os resultados numéricos, e uma notável superioridade do nosso método quando comparado a outros pré-existentes.



# Abstract

In the present work, we review the well established formalism of structure formation in the universe and the theoretical foundation for the matter power spectrum. We then review the current literature on power spectrum estimation, focusing on Fourier methods that have a close connection to the Fisher matrix formalism. With the same motivation, we develop a set of estimators for surveys with multiple galaxy species and efficiently test our predictions in simple mock catalogs. This method is suitable for treating surveys containing different species of tracers of large-scale structure, such as galaxies, quasars, Lyman- $\alpha$  absorption systems, etc. We find good agreement between our expectations and the numerical results, and a remarkable superiority of our method when compared to pre-existing ones.



# Agradecimentos

É um tanto injusto que, nessa dissertação de várias dezenas de páginas, o espaço para expressar gratidão seja tão diminuto. Minha sensação reflete o extremo oposto: poderia resumir em apenas uma página qualquer contribuição que tenha feito, mas facilmente cobriria dezenas delas em agradecimentos.

Agradeço primeiramente a Raul Abramo, meu orientador durante essa etapa da pós-graduação, pelos inestimáveis conselhos e ensinamentos, pelo companheirismo e por comigo ter compartilhado sua visão aguçada da Cosmologia e do fazer científico. Dentre os mestres, agradeço também a Marcos Lima, Raúl Angulo e Carlos Hernández, pela paciência em ensinar um jovem não particularmente brilhante, e por terem sido amigáveis e solícitos sempre que busquei ajuda. Agradeço profundamente a todo o pessoal da CEFCA, na Espanha, pela imensa hospitalidade durante o período em que lá estive, e especialmente a Mariano Moles pela preocupação de que eu tivesse as melhores condições de trabalho possíveis.

Aos colegas de sala André Alencar, Arthur Loureiro, Carolina Queiroz e Rodrigo Voivodic, agradeço pelas inúmeras discussões proveitosas e pela convivência amigável nos últimos anos. Aos outros colegas da USP, entre eles Michel Aguená, Hugo Camacho, Leila Graef, Leandro Beraldo, Meera Machado, Pramod Padmanabhan, Henrique Xavier, Irène Balmès e Loic Le Tiran, por tantos momentos de descontração. A todo o pessoal do CosmoClub, obrigado por terem sido um grupo animado e tão aberto à troca de ideias. Aos colegas da colaboração do J-PAS, agradeço pela diversão que adicionaram a tantos encontros e conferências.

Agradeço profundamente às funcionárias Amélia, Simone e Cecília, do Departamento de Física Matemática, pela prestatividade e simpatia que beneficiam a todos os cientistas do departamento. Agradeço também às funcionárias Lurdes e Nice pela limpeza primorosa de cada canto do departamento e da nossa sala, e pelos tantos chás e cafés que ajudaram no trabalho de todos. Sou grato também pelo investimento público que tornou possível meu estudo na melhor Universidade da América Latina. Torço para que jamais seja menosprezado o poder do ensino público, acessível e de qualidade, e que, como eu, muitos outros possam dele desfrutar.

Aos amigos de quem estive longe nos últimos anos, agradeço por terem sido presentes

mesmo apesar da distância. Se posso me considerar bem-sucedido, é porque tive pessoas que me apoiaram em tantas quedas e que ofereceram alento em tantas ocasiões de desânimo. Vou cometer o erro de citar os nomes de todos. É um erro porque vou inevitavelmente deixar gente querida de fora. Muito obrigado Luan, Rafael B., Pedro, Marion, Mariana, Selau, Amanda, Matheus, Riana, Vinícius, Magda, Demétrius, Alexandra, Rui, Guilherme, Augusto, Stock, Ramiro, Camile, Porcher, Larissa, Guto, Vicky, Dudu, Moni, Favero, Flávia, Maria, Dede, Ana, Tortelli, Thyago, Paine, Bruna K., Maria Clara, Pati, Rafa, Pedro M., Ike, Marco, Jean, William, Be, Leo, Poletto, João, Jumpa, Sérgio. Sinto-me muito afortunado de ter tantas pessoas a quem possa chamar de amigos.

Deixo por último aqueles por quem “gradidão” não é um termo que expressa suficientemente o que sinto. A meu pai Amauri, minha mãe Eleanir e a minha irmã Luise, por quem tenho amor infindável, muito obrigado por estarem comigo em todos os momentos, bons e ruins, de toda essa vida.

Tenho tido a alegria de, até hoje, lutar contra as mesmas perguntas que me fazia quando criança: como surgiu o Universo? De que é feito? Para onde vai? Se tenho sido feliz, é porque pude aliar a essas indagações pueris a convivência com pessoas realmente maravilhosas. Muito obrigado.

Agradeço à Fundação de Amparo à Pesquisa do Estado de São Paulo pelo financiamento do presente trabalho (Processo 2012/23400-1).

*“Let us fight for a world of reason, a world where science and progress  
will lead to all men’s happiness.”*

Charlie Chaplin (The Great Dictator - 1940)



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Structure in the Universe</b>	<b>7</b>
2.1	The Smooth Universe . . . . .	7
2.2	The Inhomogeneous Universe . . . . .	11
2.2.1	Evolution Equations . . . . .	14
2.2.2	Large Scales . . . . .	17
2.2.3	Small Scales . . . . .	19
2.2.4	The Effect of Baryons . . . . .	22
<b>3</b>	<b>Power Spectrum Estimation</b>	<b>27</b>
3.1	Signal $\times$ Noise . . . . .	27
3.1.1	Poisson Noise . . . . .	28
3.1.2	Cosmic Variance . . . . .	31
3.2	General Formalism . . . . .	32
3.3	The FKP Estimator . . . . .	40
3.3.1	Formalism . . . . .	40
3.3.2	Implementation . . . . .	43
3.4	The PVP Estimator . . . . .	44
3.4.1	Different Galaxies are Different . . . . .	44
3.4.2	Formalism and Implementation . . . . .	46
<b>4</b>	<b>The Multi-Tracer Estimator</b>	<b>51</b>
4.1	Towards an Optimal Estimator . . . . .	51
4.2	Fisher Matrix for Galaxy Surveys . . . . .	54
4.2.1	Single Species . . . . .	54
4.2.2	Multiple Species . . . . .	58
4.3	The Optimal Multi-tracer Quadratic Estimator . . . . .	60
4.4	Testing the Multi-tracer Estimators . . . . .	65
4.4.1	Lognormal Maps and the Algorithm . . . . .	66

---

4.4.2 Results . . . . .	71
<b>5 Conclusions</b>	<b>79</b>
<b>Appendices</b>	<b>85</b>
<b>A Brief Thermal History</b>	<b>85</b>
<b>B The Multi-tracer Algorithm</b>	<b>87</b>
<b>Bibliography</b>	<b>93</b>

# Chapter 1

## Introduction

Modern survey projects and ongoing experiments promise to deliver overwhelming quantities of data with breakneck speed. With different strategies, the current, state-of-the-art astronomical instruments will play a crucial role in unveiling fundamental characteristics of the cosmos and explaining a plethora of astrophysical processes that populate it.

Different observational techniques will be employed by many groups and collaborations across the world, many of them already taking and analyzing data. Even though ingenuity is a common feature among all different instruments, the science that can be done with each of them varies significantly.

The current generation of astronomical instruments consist of broad-band ([1, 2, 3] Pan-STARRS<sup>1</sup>) and narrow-band imaging [4], with spectroscopic follow-ups in a number of them ([5], eBOSS<sup>2</sup>), as well as high-performance space missions (JWST<sup>3</sup>, [6]) which combine a set of instruments for both imaging and spectroscopy. The CCD technology of these experiments's respective cameras has evolved to a point where the typical number of pixels is of around several hundred Megapixels to a few Gigapixels.

The upcoming large-scale structure surveys with such instruments will cover typical areas of around 10,000 square degrees, a significant fraction of the whole sky, and will easily gather tens of Petabytes of data after a few years. The predicted number of observed objects is also enormous: J-PAS alone expects to obtain images and photometric redshifts for 14 million early-type, luminous red galaxies (LRG), as well as several million AGNs and quasars, and many millions of other galaxies.

Not only these promising surveys cover a broad area, they go also very deep in redshift. This combination will enable the exploration of the universe not only in its spatial domain, but also in its history and evolution in time. As another example, J-PAS expects to observe galaxies up to  $z \approx 1.3$ , emission-line galaxies up to  $z \approx 2.5$  and quasars up to  $z \approx 6$ . As

---

<sup>1</sup><http://pan-starrs.ifa.hawaii.edu/public>

<sup>2</sup><https://www.sdss3.org/future/eboss.php>

<sup>3</sup><http://www.stsci.edu/jwst/doc-archive/white-papers>

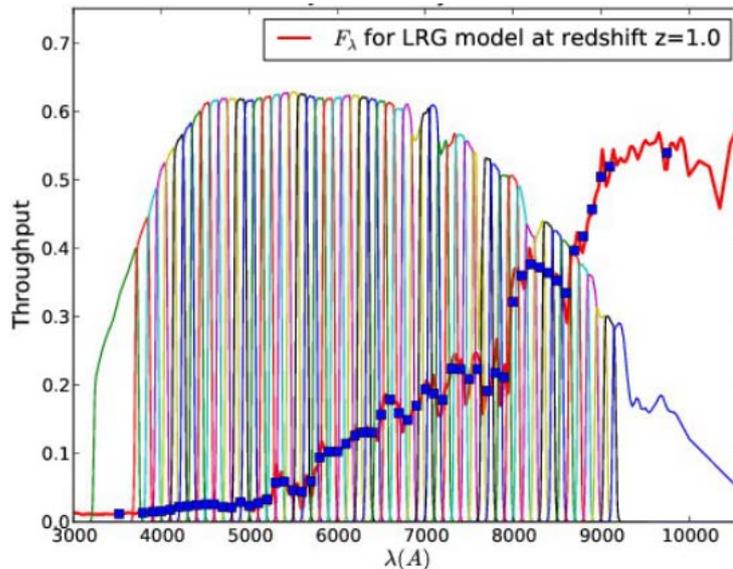
a simple comparison, a redshift of  $z = 1$  corresponds to 6 billion years into the past, and the images of the furthest objects in the Hubble Ultra Deep Field, star-forming galaxies at  $z \approx 8$  [7], correspond to looking 13 billion years back in time – when the universe itself was only around 800 million years old. In other words, we are about to map larger and larger fractions of the observable universe.

The complication introduced by such powerful instruments, however, is the unavoidable need for sophisticated and efficient data analysis tools and, at least ideally, complete control over systematic effects that will likely undermine eventual conclusions. As a convenient example, the statistical method previously used for power spectrum estimation had to change from a direct Fourier sum to modern implementations of Fast Fourier Transforms (FFTs), which will be thoroughly explained in the following sections.

Different observational techniques and astrophysical observables have, each of them, different systematic effects to be taken into account. These characteristics are specially relevant for measurements of the large-scale distribution of matter across the universe. A brief description of these observables is listed in the rest of this section.

Firstly, from an observational point of view, one could say that the more natural tracers of large-scale structure (LSS) are galaxies themselves. Even though their sheer numbers and spatial densities are of major importance for cosmological LSS studies, it is becoming ever more appealing to consider also their diversity of types. The reasons for this will become clear throughout the present work. A precise determination of the position of every galaxy in a survey is perhaps the number one challenge in this context. Broad-band and, to a lesser extent [8], narrow-band imaging (1.1), bear an inherent uncertainty in the determination of redshifts, and, therefore, of radial positions. With only a few known points in the spectral energy distribution (SED), specific emission lines must be inferred and then a photometric redshift (photo- $z$ ) – hopefully an accurate estimate of the true redshift – can be obtained. Different strategies are currently used to infer the photo- $z$  of an object, as extensively studied by [9]. The alternative method, employing spectroscopy in redshift surveys, goes around this problem and provides highly accurate redshift determinations. The drawback is that spectroscopy generally takes much longer to be carried out, which translates into fewer observed galaxies and poorer statistics than photometric surveys.

In sufficiently dense environments, galaxies clump together to form galaxy clusters. Clusters are interesting for cosmologists for a number of reasons. Their hot gas component, for instance, can be observed in X-ray and analyzed along with the Cosmic Microwave Background (CMB) through the kinetic Sunyaev-Zeldovich effect [10]. Also, the most massive ones betray the position of the largest dark matter haloes, which would be otherwise unobservable. Additionally, cluster counts conveniently probe both the growth rate of density perturbations and the geometry and expansion rate of the universe [11], making



**Figure 1.1:** The J-PAS filter system as presented in [4]. Tens of narrow filters (FWHM of  $145\text{\AA}$ ) cover the visible spectrum, while another 2 broad-band filters cover the infrared and ultraviolet.

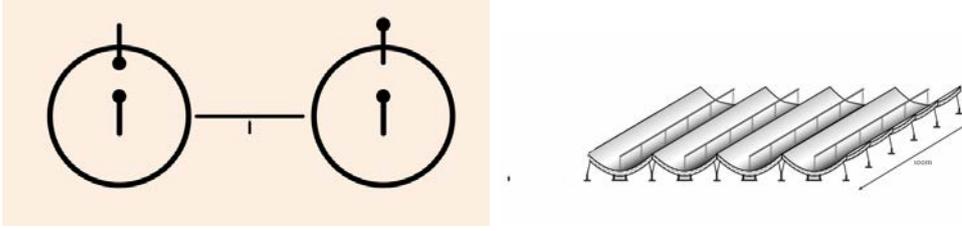
it a powerful observational tool.

The largest baryonic component of the universe, neutral hydrogen, also traces the density field across the universe. Upcoming experiments such as [SKA<sup>4</sup>, CHIME<sup>5</sup>, [12]] will map the position of large hydrogen clouds even in the neutral regime of the primordial universe, before reionization. The rest-frame characteristic wavelength of neutral hydrogen is the 21cm line, emitted from the transition between hyperfine energy levels of the H atom. For being such a ubiquitous component and for having a transition line which is believed to be exactly the same all over the universe (see, for instance, [13] for bounds on the variation of the fine-structure constant), a representation of the hyperfine transition was even inscribed in the Pioneer plaques launched in the 1970's (figure 1.2). The radio-telescopes employed in this field consist of large dishes to collect radiation with a characteristic wavelength of a few meters. The statistical methods of analysis in 21cm radiation science is actually a mixture of galaxy power spectrum techniques for low redshifts, where neutral hydrogen clouds are found mostly within galaxies, therefore sparsely accumulated, and CMB power spectrum techniques for redshifts close to reionization, where hydrogen is spread and broadly distributed.

Lyman- $\alpha$  forests, which are dense sequences of absorption lines in frequencies higher than the Lyman- $\alpha$  emission line of distant galaxies and quasars, are also powerful tracers of neutral hydrogen clouds, and also probe the reionization epoch through the Gunn-Peterson trough.

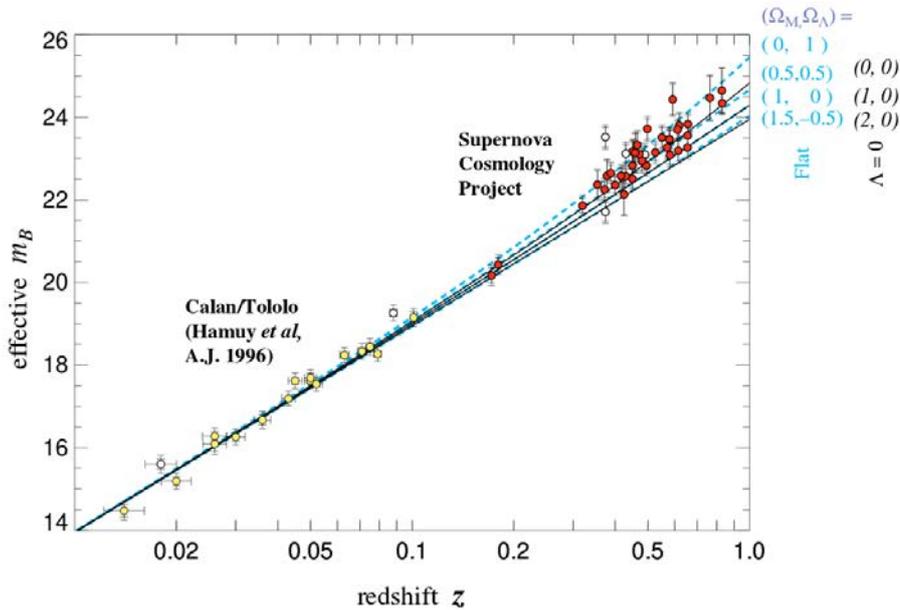
<sup>4</sup><https://www.skatelescope.org/>

<sup>5</sup><http://chime.phas.ubc.ca/>



**Figure 1.2:** Left panel: The representation of the 21cm hiperfine transition as depicted by Carl Sagan and Frank Drake, inscribed in the Pioneer plaques in the 1970's. Right panel: the Canadian Hydrogen Intensity Mapping Experiment (CHIME).

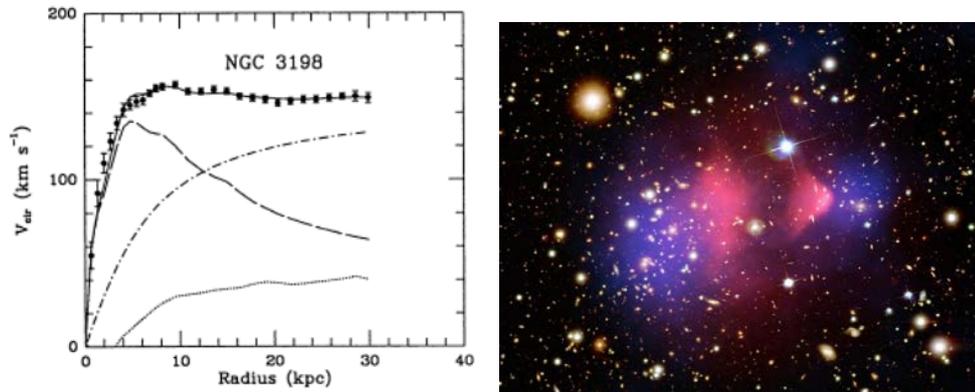
Other astrophysical observables, even though not tracing large-scale structure directly, are extremely useful in cosmology. In the late 1990's, for instance, a great discovery came from the study of type I-a supernovae (SNIa) [14, 15], which was later honored with the Physics Nobel Prize. Analyzing several tens of SNIa and making use of the fact that their light curves (i.e. the evolution of their apparent magnitudes during their lifetime) are standardizable, Schmidt, Perlmutter, Riess and others showed that the universe is in accelerated expansion, and that some strange energy component, later called dark energy, is responsible for this – or, strictly speaking, that the distance moduli of the observed SNIa were dimmer than they would be in an Einstein-de Sitter universe, and that a  $\Omega_m \approx 0.2$ ,  $\Omega_\Lambda \approx 0.8$  universe with an FRLW metric was a best fit to data (figure 1.3).



**Figure 1.3:** The Hubble diagram for the 42 SNIa detected by the Supernova Cosmology Project [15], where the best fit is a universe with a cosmological constant.

Another landmark in the cosmological scenario of the last few decades is the set of compelling evidences for dark matter [16], most likely in the form of a weakly interacting massive particle (WIMP). The excessive velocity in the rotation curves of spiral galaxies

[17] and the observations of the bullet cluster [18] (figure 1.4) are perhaps the most striking evidences for our current paradigm.



**Figure 1.4:** Left panel: the measured rotation curve (black dots) cannot be explained by luminous matter and gas (dotted and dashed curves respectively). An unobservable dark halo (dot-dashed curve) is also necessary. Extracted from [17]. Right panel: hot gas (red) as mapped by its X-ray emission seems disrupted by the collision, while in blue is the dark matter halo, which emerged unscathed from the collision, mapped by gravitational lensing of the background galaxies. Composite image from [18].

One could not forget to mention also the modern and past CMB experiments. Since the accidental discovery in 1964 of the cosmic microwave background by astronomers Penzias and Wilson, later honored with the Physics Nobel Prize in 1978, the experimental methods have grown in complexity, reaching sensitivities of a few  $nK$ . Space missions [COBE, WMAP, Planck] have estimated with great precision and accuracy the background temperature and the CMB angular power spectra, confirming predictions of the standard model of cosmology which quickly became landmark discoveries. In fact, CMB played a crucial role in establishing the standard model as such, being perhaps the most important probe in the history of the field. The geometrical flatness of the universe, the near-perfect blackbody spectrum of the background radiation, the validity of Big Bang Nucleosynthesis (BBN), the oldest snapshot of the universe at  $z \approx 1000$ , just to list a few, are remarkable byproducts of the endeavor of the scientific community to understand and explore CMB.



# Chapter 2

## Structure in the Universe

### 2.1 The Smooth Universe

The standard model of cosmology slowly began taking its current form after Einstein's general theory of relativity was recognized as a useful framework to describe the large-scale universe. In short, the current paradigm has a hot Big Bang as origin, followed by a period of cosmic inflation that leads to a spatially flat, isotropic and homogeneous FLRW universe, composed of cold dark matter, dark energy, baryons, photons and neutrinos, with general relativity connecting the dynamics of these components with the metric.

Even though experimental results in general agree with that scenario, the physics involved is not cast in stone, in the sense that each single part of this “vanilla” description are still active research fields.

The Einstein field equations (EFE) of GR can be given some motivation, but must ultimately be postulated [19]. In units of  $\hbar = c = k_B = 1$ , which will be used throughout this work, the EFE read

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 8\pi GT_{\mu\nu}. \quad (2.1)$$

It is said that the left-hand side of this equation contains information about the geometry of spacetime, with  $G_{\mu\nu}$  being the Einstein tensor,  $R_{\mu\nu}$  the Riemann tensor,  $R$  the Ricci scalar and  $g_{\mu\nu}$  the metric tensor. The right-hand side regards the energy and matter content in it, where  $G$  is Newton's gravitational constant and  $T_{\mu\nu}$  the energy-momentum tensor. Here greek indexes range from 0 to 3, encoding the 3+1 dimensions, and latin indexes, when they appear, refer strictly to the spatial dimensions 1 to 3. The energy-momentum tensor effectively includes an important term proportional to  $\Lambda g_{\mu\nu}$ , which will be made explicit shortly.

To solve the EFE is to obtain a relation between the metric tensor and the energy-momentum content. We make a few assumptions about the underlying characteristics of

such relation and proceed to solve equation (2.1). Indeed, our assumptions should later be tested against observations.

The simplest 4-dimensional, constant and flat metric tensor is diagonal and, to properly reproduce the spacetime invariant of special relativity  $ds^2 = -dt^2 + dx^2$ , should be defined as  $\eta_{\mu\nu} = \text{diag}[-1, 1, 1, 1]$ , also known as Minkowski's metric. If we want the spatial components of this flat metric to represent an expanding space, we can define *comoving coordinates* which, for the sake of simplicity, factor out this expansion from the physical coordinates  $\vec{r}$  and introduce a fixed coordinate system  $\vec{x}$ . In other words, we define the comoving coordinates  $\vec{x}$  such that  $\vec{r} = a(t)\vec{x}$ , where  $a(t)$  is the scale factor. We then rewrite an expanding version of  $\eta_{\mu\nu}$  as

$$g_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & a^2(t) & 0 & 0 \\ 0 & 0 & a^2(t) & 0 \\ 0 & 0 & 0 & a^2(t) \end{pmatrix}. \quad (2.2)$$

Equation (2.2) is known as the Friedmann-Lemaitre-Robertson-Walker metric, and readily encapsulates the very reasonable assumption that the universe is isotropic and homogeneous, i.e., that the universe as observed from our point in space is, on average, no different than what a distant observer in another galaxy would find. The FLRW is indeed more complicated for a spacetime with an arbitrary geometry, – a geometrically open or closed universe – but equation (2.2) suffices for the present work.

Our next assumption is that the energy-momentum content of the universe is similar to that of a perfect fluid. In this case,  $T_{\nu}^{\mu} = \text{diag}[-\rho, p, p, p]$ , where  $\rho$  is the density of the fluid and  $p$  its pressure. Here we are also using Einstein's summation convention and the usual identity between vectors and 1-forms  $A_{\mu} = g_{\mu\nu}A^{\nu}$ .

We now have a physical motivation for both sides of eq. (2.1). Solving the EFE now means solving a set of differential equations for the scale factor  $a(t)$ . Substituting the  $T_{\nu}^{\mu}$  for the perfect fluid on the right-hand side and (2.2) on the other one, and using the definition of  $G_{\mu\nu}$  [20], we find 2 independent differential equations for  $a(t)$ , known as Friedmann Equations:

$$\begin{cases} H^2(t) & \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho \\ \ddot{a}/a & = -\frac{4\pi G}{3}(\rho + 3p), \end{cases} \quad (2.3)$$

where we have defined  $H(t) \equiv \dot{a}/a$  as the Hubble parameter. The first equation can be rewritten for the current time  $t = t_0$  as  $\rho_{\text{crit},0} = 3H_0^2/8\pi G \approx 2.77 \times 10^{11} h^2 \text{M}_{\odot} \text{Mpc}^{-3} \approx$

$10^{-30}\text{g/cm}^3$ , which is called the *critical density*, and where subscripts 0 mean current time ( $t = t_0$ ) quantities. The physical meaning of  $\rho_{\text{crit}}$  is better understood when we consider a universe with an arbitrary geometry, in which case the critical density is the value for which the universe is flat.

Now, considering a universe composed of several different perfect fluids is equivalent to writing  $\rho(t) = \sum_i \rho_i(t)$ . We can then rewrite the first equation of (2.3) as

$$H^2(t) = H_0^2 \sum_i \frac{\rho_i(t)}{\rho_{\text{crit},0}}. \quad (2.4)$$

Right now, we just lack an understanding of how the densities  $\rho_i(t)$  evolve with time. To sort this out, let us consider the expansion of the universe to be approximately adiabatic. This assumption holds quite well during epochs distant enough from the transition between the domination of different components. Following this idea, we have, for an adiabatic process,  $dE = d(\rho V) = -pdV$ . However, since  $V \propto a^3(t)$ , we have  $d(\rho a^3) = -pd(a^3)$  and then  $a^3 d\rho + 3\rho a^2 da = -3a^2 p da$  which finally yields

$$\frac{\partial \rho_i}{\partial t} + 3H(1 + w_i)\rho_i = 0, \quad (2.5)$$

where  $w_i = p_i/\rho_i$  is the equation of state parameter for the fluid  $i$ . It can be readily checked that, with initial conditions  $\rho_i(t = t_0) = \rho_0$ , equation (2.5) is solved by

$$\rho_i(a) = \rho_{i,0} a^{-3(1+w_i)}. \quad (2.6)$$

One can also consider two photons, following null geodesics, i.e. with null relativistic invariants  $ds^2$ , the first emitted at  $t_e$  and observed at  $t_o$  and the second emitted at  $t_e + \delta t_e$  and observed at  $t_o + \delta t_o$ , but traveling the same comoving distance. In this case,  $ds^2 = -dt^2 + a^2 dx^2 = 0$  implies

$$\int_{t_e}^{t_o} \frac{dt}{a(t)} = \int_{t_e + \delta t_e}^{t_o + \delta t_o} \frac{dt}{a(t)}. \quad (2.7)$$

Regarding  $\delta t \ll t$ , the previous equation is equivalent to writing  $\lambda_e/\lambda_o = a(t_e)/a(t_o)$  (since  $c\delta t = \lambda$ , but recall that we define  $c = 1$ ). Finally, defining the *redshift*  $z$  as the relative variation of the photon's wavelength, we find a useful relation between the scale factor, which is an unobservable quantity, and a measurable wavelength shift

$$1 + z = \frac{\lambda_0}{\lambda_e} = \frac{1}{a}, \quad (2.8)$$

where we also used  $a(t_0) = 1$ .

Using equation (2.8), we can rewrite (2.6) as

$$\rho_i(z) = \rho_{i,0}(1+z)^{3(1+w)}. \quad (2.9)$$

The question to be asked now is what are the  $w_i$  for each component of the universe's energy-momentum budget. Matter, which can be treated as dust, is a pressureless fluid, so  $w = 0$ . For radiation, the statistical mechanics of a photon fluid yields  $w = 1/3$ . A third component, of yet unexplained origin, which we currently call *dark energy*, and forces the universe to expand, can also be treated as a perfect fluid, with  $w \simeq -1$  [21].

Going back to equation (2.4) and introducing the *density parameters*

$$\Omega_i(t) \equiv \frac{\rho_i(t)}{\rho_{\text{crit},0}}, \quad (2.10)$$

we can finally write the first Friedmann equation in a very convenient way as

$$E(z) = \frac{H(z)}{H_0} = [\Omega_{r,0}(1+z)^4 + \Omega_{m,0}(1+z)^3 + \Omega_\Lambda]^{1/2}. \quad (2.11)$$

One can also note from eq. (2.4) that, at  $t = t_0$ ,

$$\Omega_r + \Omega_m + \Omega_\Lambda = 1. \quad (2.12)$$

The power of the Friedmann equations relies on the fact that, given a set of reasonable and simplifying assumptions, the expansion history of the universe and the energy density of all its components are uniquely determined. The assumptions employed should, obviously, be minutely tested. As an example, one can carry out a straightforward integration to determine the solution  $a(t)$  for the case of a flat universe with matter as its only component, also known as an Einstein-de Sitter universe:

$$a(t) = \frac{3}{2}H_0\Omega_m t^{2/3}. \quad (2.13)$$

Another important relation we get from the Friedmann equations is the *redshift-distance relation*, which tells us how the comoving distance out to redshift  $z$ , which we shall call  $\chi(z)$ , depends on the cosmology. For an object at redshift  $z$ , this relation is

$$\chi(z) = \frac{1}{H_0} \int_0^z \frac{dz'}{E(z')}, \quad (2.14)$$

where  $H_0 = 100 h \text{ Mpc}^{-1} \text{ km s}^{-1}$  in conventional units. Here  $h$  is a dimensionless factor which represents our ignorance of the correct value. Recalling that light follows the null spacetime trajectory  $ds^2 = 0$ , we can also define the *conformal time* as the total comoving distance light could have traveled out to time  $t$ , which is nothing more than the integrated  $dx = dt/a(t)$  (recall that  $c = 1$ ):

$$\eta(t) = \int_0^t \frac{dt'}{a(t')}. \quad (2.15)$$

The quantity  $\eta$  monotonically increases with time and has a very interesting physical meaning: it determines the size of the comoving horizon. If two objects lie at a distance greater than  $\eta$ , there will be no causal contact between them at instant  $t$ , assuming that the expansion started at  $t = 0$ .

For completeness, the reader can refer to [22] for other important definitions of cosmological distances such as the *angular diameter distance* and the *luminosity distance*.

One final remark is necessary. Equation (2.6) shows that radiation and matter evolve differently, so we expect them to contribute in different proportions to the energy budget of the universe. Since  $\rho_r \propto a^{-4}$  for radiation while  $\rho_m \propto a^{-3}$  for matter, the further we look into the past, the more radiation will dominate over matter<sup>1</sup>. We call this regime *radiation domination era*. Conversely, when (mostly dark) matter dominates, we say it is the *matter domination era*. As will become clear shortly, this has a great impact on how structures evolve in the universe.

Of course, there should be an  $a_{\text{eq}} \equiv a(t = t_{\text{eq}})$  such that the energy of radiation and matter are equal. Taking into account the effective number of neutrino species and the current cosmic background temperature, we can equate  $\Omega_m = \Omega_r$  to get

$$a_{\text{eq}} \approx \frac{1}{4000}, \quad (2.16)$$

which also yields, for the useful case of a flat universe containing only dark matter and radiation,

$$k_{\text{eq}} = H_0 \sqrt{2 \frac{\Omega_m a_0}{a_{\text{eq}}}} = 0.072 \Omega_m h^2 \text{ Mpc}^{-1}. \quad (2.17)$$

## 2.2 The Inhomogeneous Universe

Until now, we've been describing the physics of a smooth universe. It is smooth in the sense that all of its components are perfect fluids without any clumpiness. The density

---

<sup>1</sup>For a brief discussion on the kind of physics that will emerge as a consequence, see Appendix A.

of each fluid in any point of space is precisely the average density over all space.

This smooth FLRW universe, however, won't generate the kind of structures we observe in nature. Matter, for instance, clumps in scales that span many orders of magnitude, from a few atoms/ $m^3$  in the intergalactic medium to  $10^{17} \text{ kg m}^{-3}$  in neutron stars. The unperturbed universe is clearly not rich enough to describe this variety of environments.

In this section, we will consider a flat universe which contains only dark matter and radiation. The contribution of baryons will be neglected. We will introduce perturbations in the radiation and matter densities and see how they evolve in time. The matter and radiation domination epochs described in the previous section will be useful here: we can mostly neglect radiation in the matter domination period and vice-versa.

We should keep two things in mind: first, gravity is a “greedy” force. Without an external agent, even small overdensities will tend to collapse and attract matter towards them. The second thing is that radiation has pressure, while matter hasn't. This means that when the photon bath in the universe is too hot, it can erase and smooth out overdense regions. These forces are thus competing against each other in the scenario of structure formation.

We can already guess that  $a_{\text{eq}}$  will be an important quantity in what follows: the domination of matter over radiation or vice-versa will explicitly tell whether gravity or pressure wins the fight. Also, since we're dealing with an expanding universe, the characteristic length of perturbations will be important to determine how they evolve. Suppose a perturbation with a very long wavelength (we are implicitly describing these perturbations in Fourier space) emerges when the universe itself is much smaller than it. In this case, we say that this mode has not yet *entered the horizon*,<sup>2</sup> or that it's *super-horizon*. Accordingly, short wavelength modes will enter the horizon much earlier in the history of the universe. It is then called *sub-horizon*.

An important aspect of this section is that we will assume the primordial universe is already, in fact, slightly inhomogeneous. There is a large, rich branch of cosmology that describes how this is possible, and *inflation* [23] is perhaps the most successful of these theories of the primordial universe. Further detailing is beyond the scope of this work.

The gravitational potential  $\Phi$  will respond to the distribution of matter and radiation. Since in general the potential will also affect them, the evolution of all these quantities will be coupled. To step away from the smooth universe, we must define perturbation quantities. The gravitational potential  $\Phi$  itself will be one of them. We will also have the radiation temperature contrast field  $\Theta = \Delta T/\bar{T}$ , the matter density contrast  $\delta = \Delta\rho/\bar{\rho}$

---

<sup>2</sup>For the sake of all the beginners who, like the author, struggled to make sense of this abstruse jargon: “entering the horizon” means that the perturbation as a whole is in causal contact. It will then evolve according to the state in which the universe is. Before that, when this same perturbation was “super-horizon”, the lack of causal contact would keep it from “feeling” the epoch of the universe and evolving accordingly.

and the peculiar velocity of matter  $v$ . We perturb the smooth metric tensor (2.2) by introducing two potentials:

$$g_{\mu\nu} = \begin{pmatrix} -1 - 2\Psi(\vec{x}, t) & 0 & 0 & 0 \\ 0 & a^2(1 + 2\Phi(\vec{x}, t)) & 0 & 0 \\ 0 & 0 & a^2(1 + 2\Phi(\vec{x}, t)) & 0 \\ 0 & 0 & 0 & a^2(1 + 2\Phi(\vec{x}, t)) \end{pmatrix}. \quad (2.18)$$

Although this could have been done in a number of ways, we chose this specific one for the simplicity of keeping the metric tensor diagonal. Also, for now we will differentiate between  $\Psi$  and  $\Phi$ , but later on we will take them to be the opposite of each other, based on physical arguments.

We will write the perturbation solutions in term of  $\Phi$ . Since the present-day matter distribution traces it very accurately, we can (and certainly will) use redshift surveys to compare the real world with our theoretical expectations. The link between theory and observations is the purpose of this whole section.

Schematically, we can write the potential as

$$\Phi(\vec{k}, a) = \Phi_p(\vec{k}) [\text{Transfer function}(k)] [\text{Growth function}(a)] = \Phi_p(\vec{k}) T(k) \frac{D_1(a)}{a}. \quad (2.19)$$

Here,  $\Phi_p(\vec{k})$  is the primordial inhomogeneity, which we are taking for granted. The *transfer function* is the part that carries information about when different modes will enter the horizon – before, approximately on or after  $a_{\text{eq}}$ . For convenience, we define the *growth function* as  $D_1(a)/a$  because, as it will become clear, the potential is a constant for  $a \ll a_{\text{eq}}$  and  $a \gg a_{\text{eq}}$  and  $D_1$ , which is actually the factor by which the matter overdensities grow, is roughly proportional to  $a$ . In other words,  $D_1(a)$  is the part that carries our notion that matter will tend to accumulate due to gravity. Formally, we define

$$T(k) = \frac{\Phi(k, a_{\text{late}})}{\Phi_{LS}(k, a_{\text{late}})} \quad (2.20)$$

$$\frac{D_1(a)}{a} = \frac{\Phi(a)}{\Phi(a_{\text{late}})}, \quad (2.21)$$

where  $a_{\text{late}}$  is much later than the regime during which the transfer function acts, – recall that it will “kick in” only during horizon crossing – and  $\Phi_{LS}$  is the primordial potential decreased by a small amount, a factor of 9/10 as we shall soon show.

Now,  $\Phi_p(\vec{k})$  is essentially stochastic, which is perfectly reasonable. If it was not this way, the initial potential would be deterministically defined over all space (and therefore  $k$ -space), and we would be able to predict the position of every single overdensity in

the universe. In other words, we would have an infinitely fine-tuned – and infinitely unappealing – theory of the primordial universe. The goal of such theories is exactly the opposite: to simplify the initial conditions.

What we must actually do, then, is predict observables in terms of statistical moments of the potential. We define the *power spectrum* of a stochastic variable in Fourier space, in this case  $\Phi(\vec{k}, a)$ , as the second moment of its distribution:

$$P_{\Phi} = \langle |\Phi^2| \rangle. \quad (2.22)$$

The power spectrum of the matter density contrast  $\delta$  is then

$$\langle \delta(\vec{k}, a) \delta^*(\vec{k}', a) \rangle = (2\pi)^3 \delta(\vec{k} - \vec{k}') P(k, a), \quad (2.23)$$

where we explicitly dropped the directional dependence of  $P$  on  $\vec{k}$  on the basis of isotropy. Finally,  $\delta$  relates to eq. (2.22) through the Poisson equation in Fourier space (for late time, large- $k$  and no radiation limit):

$$\Phi = \frac{4\pi G \rho_m a^2 \delta}{k^2}, \quad (2.24)$$

which, together with eqs. (2.10), the first (2.3) and (2.19), reads

$$\delta(\vec{k}, a) = \frac{3}{5} \frac{k^2}{\Omega_m H_0^2} \Phi_p(\vec{k}) T(k) D_1(a). \quad (2.25)$$

The bottom line here is: if we have a good description for what the transfer and growth functions are, we can use eqs. (2.25) and (2.23) to compare theory with observations. The ultimate goal of the present work is to obtain  $P(k)$  directly from observations. First, however, we need to know what to expect from  $T(k)$ . The rest of this chapter is dedicated to that.

## 2.2.1 Evolution Equations

We shall begin by deriving the general evolution equations for cold dark matter (CDM) and radiation. For CDM, we will use the fluid mechanics approach. This can be made easy by the fact that dark matter is pressureless and cold (non-relativistic). For the (Newtonian) fluid mechanics description, we will need a generalized version of the Euler equation for the conservation of energy and momentum, and another for continuity.

The generalized continuity equation is

$$\frac{\partial \rho}{\partial t} + \frac{1}{a} \vec{\nabla} \cdot (\rho \vec{v}) + 3 \left( H + \frac{\partial \Phi}{\partial t} \right) \rho = 0, \quad (2.26)$$

which properly reduces to eq. (2.5) when the perturbations, given by  $\vec{v}$  and  $\Phi$ , are neglected. Here  $\rho$  represents the density of dark matter alone, as previously explained. To introduce the density perturbation, let

$$\rho \equiv \bar{\rho}(1 + \delta),$$

where  $\bar{\rho}$  is the unperturbed matter density, a zero-order quantity in terms of the perturbation. Since we want to deal only with first-order perturbations, we replace by  $\bar{\rho}$  the  $\rho$  factors in equation (2.26) that are already multiplying the first order quantities  $\vec{v}$  and  $\Phi$ , and by  $\bar{\rho}(1 + \delta)$  the other ones. The result is

$$\frac{\partial \bar{\rho}(1 + \delta)}{\partial t} + \frac{1}{a} \vec{\nabla} \cdot (\bar{\rho} \vec{v}) + 3H\bar{\rho}(1 + \delta) + 3\frac{\partial \Phi}{\partial t} \bar{\rho} = 0.$$

To obtain the purely first-order terms, we subtract the zero-order continuity equation from the perturbed one and, dividing by  $\bar{\rho}$ , we finally have

$$\frac{\partial \delta}{\partial t} + \frac{1}{a} \vec{\nabla} \cdot \vec{v} + \frac{\partial \Phi}{\partial t} = 0. \quad (2.27)$$

In the same spirit, the generalized version of the Euler equation is

$$\frac{\partial(\rho v)}{\partial t} + 4H\rho v + \frac{\rho}{a} \vec{\nabla} \Psi = 0. \quad (2.28)$$

Since none of the terms in the above equation is of order zero in the perturbations, we can simply replace  $\rho \rightarrow \bar{\rho}$ , use the usual zero-order relation (2.5) and then divide by  $\bar{\rho}$  to arrive at

$$\frac{\partial v}{\partial t} + Hv + \frac{1}{a} \vec{\nabla} \Psi = 0. \quad (2.29)$$

Equations (2.27) and (2.29) describe the evolution of dark matter. We now want a set of equations to describe the evolution of radiation. The newtonian approach would only be useful in this case if we tackled the relativistic formulation of fluid dynamics. Another way to proceed, however, is to try and describe the evolution of the distribution function for radiation. Appendix A briefly describes various phenomena in the thermal history of the universe which cannot be explained with equilibrium thermodynamics. What we want, then, is to use the Boltzmann equation

$$\frac{df}{dt} = C[f]. \quad (2.30)$$

The left-hand side of eq. (2.30) expresses the variation in time of the distribution function  $f = f(\vec{x}, p, t)$ , where  $p$  is the momentum of the photons. The perturbed distribution function is given by

$$f(\vec{x}, p, t) = \frac{1}{\exp [p/\bar{T}(1 + \Theta)] - 1}, \quad (2.31)$$

where  $\Theta = \Delta T/\bar{T}$ , in the same way we did with the matter density perturbations. The right-hand side of eq. (2.30) is a functional of  $f$  which comprises all possible interactions (collision) terms. Generally, we would introduce the Compton scattering terms in this side. However, since we intend to deal with radiation in a broad sense, which includes neutrinos too, and we want both photons and neutrinos to follow the same evolution equations, we will disregard these scattering terms. In other words, we will neglect variations in *optical depth*.

The left-hand side can be rewritten as

$$\frac{\partial f}{\partial t} + \frac{\partial f}{\partial x^i} \frac{dx^i}{dt} + \frac{\partial f}{\partial p} \frac{dp}{dt}. \quad (2.32)$$

From the geodesics of the photon we can determine  $dx^i/dt$ , and from the dynamics in the perturbed metric we determine  $dp/dt$ . We get

$$\frac{dx^i}{dt} = \frac{\hat{p}^i}{a}$$

$$\frac{dp}{dt} = -p \left[ H + \frac{\partial \Phi}{\partial t} + \frac{\hat{p}^i}{a} \frac{\partial \Psi}{\partial x^i} \right].$$

The physical meaning of  $dp/dt$  is that the photon will gain energy when falling into potential wells and lose it once it emerges. Using the functional form (2.31) we can derive the other terms in (2.32). The first-order part is then given by

$$\frac{df}{dt} = -p \frac{\partial f}{\partial p} \left[ \frac{\partial \Theta}{\partial t} + \frac{\hat{p}^i}{a} \frac{\partial \Theta}{\partial x^i} + \frac{\partial \Phi}{\partial t} + \frac{\hat{p}^i}{a} \frac{\partial \Psi}{\partial x^i} \right]. \quad (2.33)$$

Since we are neglecting the collision term, it suffices to equate (2.33) to zero. By doing this, we get the evolution equation for radiation in first-order perturbation theory:

$$\frac{\partial \Theta}{\partial t} + \frac{\hat{p}^i}{a} \frac{\partial \Theta}{\partial x^i} = -\frac{\partial \Phi}{\partial t} - \frac{\hat{p}^i}{a} \frac{\partial \Psi}{\partial x^i}. \quad (2.34)$$

We just need one more equation to close the set of perturbation equations we need to describe the universe, along with (2.27), (2.29) and (2.34). That would be a generalized version of the Poisson equation, which we can get from the Einstein equations (2.1) in the perturbed metric. After calculating the connections, the Riemann tensor and the

Ricci scalar for the metric (2.18)<sup>3</sup> and the energy-momentum tensor, the “00” Einstein equation yields

$$k^2\Phi + 3\frac{\dot{a}}{a}\left(\dot{\Phi} - \frac{\dot{a}}{a}\Psi\right) = 4\pi Ga^2[\rho_{dm}\delta + 4\rho_r\Theta_0], \quad (2.35)$$

or, in a useful algebraic form, without derivatives but carrying the same information,

$$k^2\Phi = 4\pi Ga^2\left[\rho_{dm}\delta + 4\rho_r\Theta_0 + \frac{3aH}{k}(i\rho_{dm}v + 4\rho_r\Theta_1)\right]. \quad (2.36)$$

In both (2.35) and (2.36), for convenience, we are using dots to denote derivatives with respect to the conformal time defined in (2.15). The final step is to express (2.27), (2.29) and (2.34) also in conformal time and Fourier space. We can separate (2.34) in its monopole and dipole terms (respectively  $\Theta_0$  and  $\Theta_1$ ), which is the same we already did in eqs. (2.35) and (2.36), through multiplication by Legendre polynomials and integration. One more thing is that we will set  $\Psi = -\Phi$ , which means we will neglect higher moments of the photon distribution<sup>4</sup>. The final set of equations is:

$$\dot{\Theta}_0 + k\Theta_1 = -\dot{\Phi} \quad (2.37)$$

$$\dot{\Theta}_1 - \frac{k}{3}\Theta_0 = -\frac{k}{3}\Phi \quad (2.38)$$

$$\dot{\delta} + ikv = -3\dot{\Phi} \quad (2.39)$$

$$\dot{v} + \dot{a}v = ik\Phi, \quad (2.40)$$

and recall that we have 2 choices for the generalized Poisson equation that closes our set of 5 perturbed quantities and 5 equations: (2.36) or

$$k^2\Phi + 3\frac{\dot{a}}{a}\left(\dot{\Phi} + \frac{\dot{a}}{a}\Phi\right) = 4\pi Ga^2[\rho_{dm}\delta + 4\rho_r\Theta_0], \quad (2.41)$$

which is exactly (2.35) with the  $\Psi = -\Phi$  approximation. We can finally start to understand how different scales evolve and find the transfer function.

### 2.2.2 Large Scales

As the universe expands, so does its comoving horizon  $\eta$ . The expansion also slowly changes which component dominates the energy budget of the universe. As we were previously discussing, a perturbation mode starts entering the horizon when its characteristic length is close to the horizon size. Since large scales correspond to very long wavelength

<sup>3</sup>Which is a once-in-a-lifetime effort, then one is allowed to cheat and copy the final results for the rest of his/her life.

<sup>4</sup>In other words, we are neglecting any kind of anisotropic stress in the distribution of the components.

modes (therefore small  $k$ ), we expect these modes to enter the horizon relatively late in the history of expansion. The meaning of “late” here is  $a(t_{\text{crossing}}) \gg a_{\text{eq}}$ . In other words, we will start dealing with modes that enter the horizon during the matter domination epoch.

During the radiation and equality epochs, then, these modes are super-horizon. We will track the evolution of the potential before the perturbations enter the horizon, and then through horizon crossing.

Much before entering the horizon (for very large scales), since  $\eta$  is the size of the size of the horizon itself, it is valid to assume that  $k\eta \ll 1$  for these modes, so we can start by neglecting all terms with  $k$  in the equations (2.37)-(2.40). Since (2.36) has terms which are inversely proportional to  $k$ , we will use equation (2.41). The relevant equations are (2.41) and

$$\dot{\Theta}_0 = -\dot{\Phi} \quad (2.42)$$

$$\dot{\delta} = -3\dot{\Phi}. \quad (2.43)$$

Eqs. (2.42) and (2.43) tell us that  $\delta - 3\Theta_0$  is a constant, which we know is zero from initial conditions. Setting  $\delta/3 = \Theta_0$  and defining a new evolution variable  $y \equiv a/a_{\text{eq}} = \rho_{dm}/\rho_r$ , such that

$$\frac{d}{d\eta} = \frac{dy}{d\eta} \frac{d}{dy} = aHy \frac{d}{dy},$$

the Einstein equation becomes

$$3\frac{\dot{a}}{a} \left( \dot{\Phi} + \frac{\dot{a}}{a} \Phi \right) = 4\pi G a^2 \rho_{dm} \delta \left[ 1 + \frac{4}{3y} \right]. \quad (2.44)$$

We want a differential equation that decouples the system. We can do that by differentiating one of the two first-order equations, then plugging it into the other will yield a second-order equation. Doing this and changing derivatives with respect to  $\eta$ , denoted by dots, by derivatives with respect to  $y$ , denoted by primes, we have the second-order equation for the potential given by

$$\Phi'' + \frac{21y^2 + 54y + 32}{2y(y+1)(3y+4)} \Phi' - \frac{\Phi}{y(y+1)(3y+4)} = 0. \quad (2.45)$$

The relevant solution to this equation was found by [24]:

$$\Phi = \frac{\Phi(0)}{10} \frac{1}{y^3} \left[ 16\sqrt{1+y} + 9y^3 + 2y^2 - 8y - 16 \right] \quad (2.46)$$

Taking the limits correctly, we see that the  $\Phi \rightarrow \Phi(0)$  when  $y$  is small, and  $\Phi \rightarrow 9/10 \Phi(0)$  when  $y \gg 1$ , after equality. This is the origin of the factor of 9/10 explained in the beginning of the section.

Solution (2.46) is valid only for super-horizon modes. For late times, we just showed that the potential remains constant. In these late times, radiation can safely be neglected since we are in the matter domination epoch. To solve our set of equations (2.37)-(2.40) for this different regime, we will disregard both radiation equations and use the algebraic Einstein equation (2.36). We then have

$$\dot{\delta} + ikv = 0 \quad (2.47)$$

$$\dot{v} + aHv = ik\Phi \quad (2.48)$$

$$k^2\Phi = \frac{3}{2}a^2H^2 \left[ \delta + \frac{3aHiv}{k} \right], \quad (2.49)$$

and from the previous case for a super-horizon mode, we expect that a solution for the potential will be a constant. Indeed the above set can be rearranged in order to get a differential equation of the form  $\alpha\ddot{\Phi} + \beta\dot{\Phi} = 0$ , for which a constant  $\Phi$  is a solution. The conclusion is that  $\Phi = 9/10$  remains constant during the matter domination for the large modes. We could then say that the power spectrum is roughly *scale-independent* for this regime.

Conceptually, we had already realized that the scale up to which this conclusion holds is approximately given by eq. (2.17). If the potential is constant, it means that  $P(k) \propto k$  with the initial conditions given by inflation, which is called the Harrison-Zel'dovich spectrum. A constant potential also means that

$$\delta \propto a,$$

and the consequence for  $T(k)$  is that it will be constant for the large scales we are dealing with. We will now investigate the small scales ( $k > k_{\text{eq}}$ ) of the spectrum.

### 2.2.3 Small Scales

The fundamental difference now is that the modes we will deal with are small enough so that they enter the horizon during the radiation epoch, i.e. when  $a < a_{\text{eq}}$ . Since in this case super-horizon modes will not have much time to evolve, we will study the cases of modes entering the horizon during the radiation domination, and modes already inside the horizon when  $a = a_{\text{eq}}$ .

Analogously to what we did for the matter domination epoch, here we will disregard the contribution of  $\delta$  to the potential. Even though matter will be influenced by the potential, it won't influence it in this epoch.

We start by taking equation (2.36) without the matter source terms and use it to

eliminate  $\Theta_0$  from (2.37) and (2.38). We are left with

$$-\frac{3}{k\eta}\dot{\Theta}_1 + k\Theta_1 \left[1 + \frac{3}{k^2\eta^2}\right] = -\dot{\Phi} \left[1 + \frac{k^2\eta^2}{6}\right] - \Phi \frac{k^2\eta}{3} \quad (2.50)$$

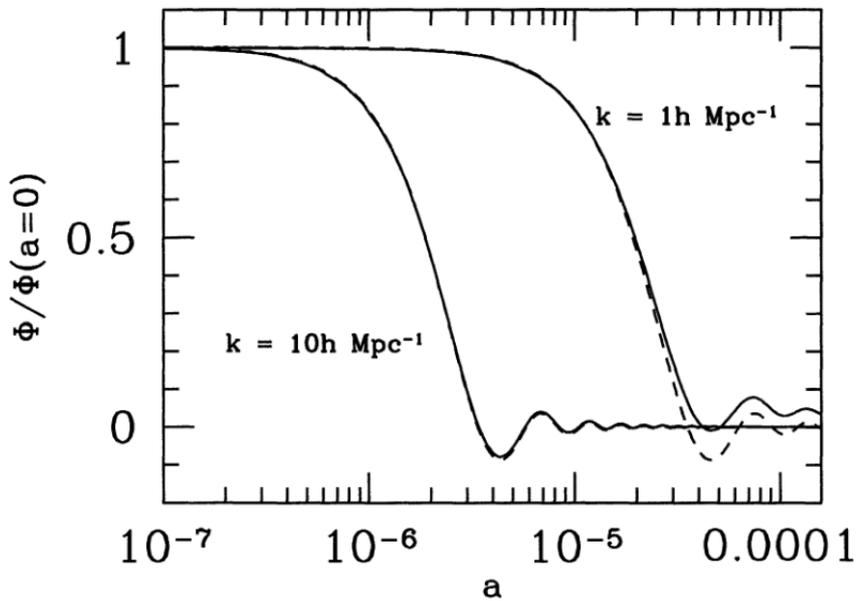
$$\dot{\Phi} + \frac{1}{\eta}\Phi = -\frac{6}{k\eta^2}\Theta_1. \quad (2.51)$$

Again, we differentiate to find a single second-order equation for  $\Phi$ , given by

$$\ddot{\Phi} + \frac{4}{\eta}\dot{\Phi} + \frac{k^2}{3}\Phi = 0 \quad (2.52)$$

With a clever change of variables, we can see that this equation is a spherical Bessel equation of order 1, with solutions given by a Neumann function (which blows up for  $\eta \ll 1$ , therefore will be disregarded) and a Bessel function. We can express the Bessel function of order 1 in terms of trigonometric functions, and the solution is finally

$$\Phi = 3\Phi_p \left[ \frac{\sin(k\eta/\sqrt{3}) - (k\eta/\sqrt{3}) \cos(k\eta/\sqrt{3})}{(k\eta/\sqrt{3})^3} \right]. \quad (2.53)$$



**Figure 2.1:** The evolution of the potential during the epoch of radiation domination. The dashed curve is the approximate solution (2.53), while the solid lines are the actual numerical evolution of the perturbation equations. Before oscillating, the potential drops significantly as the perturbation is suppressed. Extracted from [20].

In the limits of large and small  $\eta$ , the potential will tend to a constant. Right after crossing the horizon, however, the potential decays and starts to oscillate with a decreasing amplitude (see figure 2.1). With the solution for the potential we can now track what

happens to the matter perturbations we had disregarded before. Differentiating eq. (2.40) and eliminating  $v$  using (2.40) again, the evolution for  $\delta$  becomes

$$\ddot{\delta} + \frac{1}{\eta}\dot{\delta} = -3\ddot{\Phi} + k^2\Phi - \frac{3}{\eta}\dot{\Phi}. \quad (2.54)$$

The dominating part of the solution to the above equation is that

$$\delta \propto \ln a,$$

which means that the matter perturbations will still grow in the radiation epoch, but otherwise slower than the  $\delta \propto a$  behavior in the matter epoch, due to the radiation pressure.

Now, what about the modes that are already inside the horizon (sub-horizon) when the transition from radiation to matter domination occurs? When this happens, we can approximately disregard the contribution of radiation to the potential – dark matter will begin affecting it more than radiation could. We will look for a description in terms of our usual set of equations (2.39), (2.40) and (2.36). Changing variables to the convenient  $y = a/a_{\text{eq}}$ , we have

$$\delta' + \frac{ikv}{aHy} = -3\Phi' \quad (2.55)$$

$$v' + \frac{v}{y} = \frac{ik\Phi}{aHy} \quad (2.56)$$

$$k^2\Phi = \frac{3y}{2(y+1)}a^2H^2\delta. \quad (2.57)$$

Differentiating and eliminating variables other than  $\delta$ , we arrive at the Meszaros equation for the evolution of sub-horizon modes of cold, dark matter perturbations when  $a = a_{\text{eq}}$  and radiation starts to become negligible:

$$\delta'' + \frac{2+3y}{2y(y+1)}\delta' - \frac{3}{2y(y+1)}\delta = 0. \quad (2.58)$$

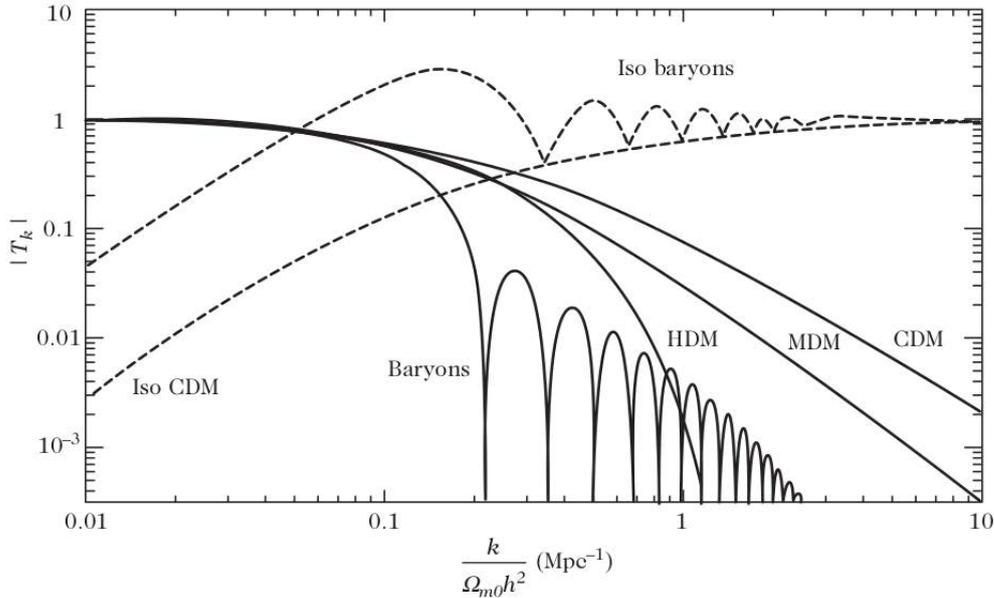
The solution is of the form

$$\delta(k, y) = C_1D_1(y) + C_2D_2(y),$$

where  $D_1(y) = y + 2/3$  is the growing solution, which will dominate at late times. The transfer function for late times ( $k \gg k_{\text{eq}}$ ) is such that [25]

$$T(k) \propto \frac{\ln k}{k^2},$$

which is a decreasing function.



**Figure 2.2:** Transfer function for many different scenarios. The one we’ve been treating is “CDM”, which is constant for large scales and is suppressed for small scales. Extracted from [26].

We now understand the asymptotic behavior of  $T(k)$  for large and small scales. The exact form of the transfer function is only available to us using numerical techniques, but there are known fits for it. One of them was given by [25], and appropriately recovers the asymptotic limits:

$$T(x) = \frac{\ln(1 + 0.171x)}{0.171x} [1 + 0.284x + (1.18x)^2 + (0.399x)^3 + (0.49x)^4]^{-1/4}$$

where  $x \equiv k/k_{\text{eq}}$ . The derivation carried out in this and the previous sections is restricted to a flat universe in which  $\Omega_{\text{cdm}} = 1$ . The  $y \gg 1$  limit in the Meszaros equation, however, is not valid if another component dominates over dark matter at late times. Currently, from the SNIa research and many other probes, we know that this is precisely the case: a form of dark energy  $\Omega_{\Lambda}$  currently dominates the energy budget. For a general cosmology, the shape of the transfer function can change dramatically (see figure 2.2), and the appropriate expression for the growing solution is

$$D_1(a) \propto H(a) \int^a \frac{da'}{[a'H(a')]^3}.$$

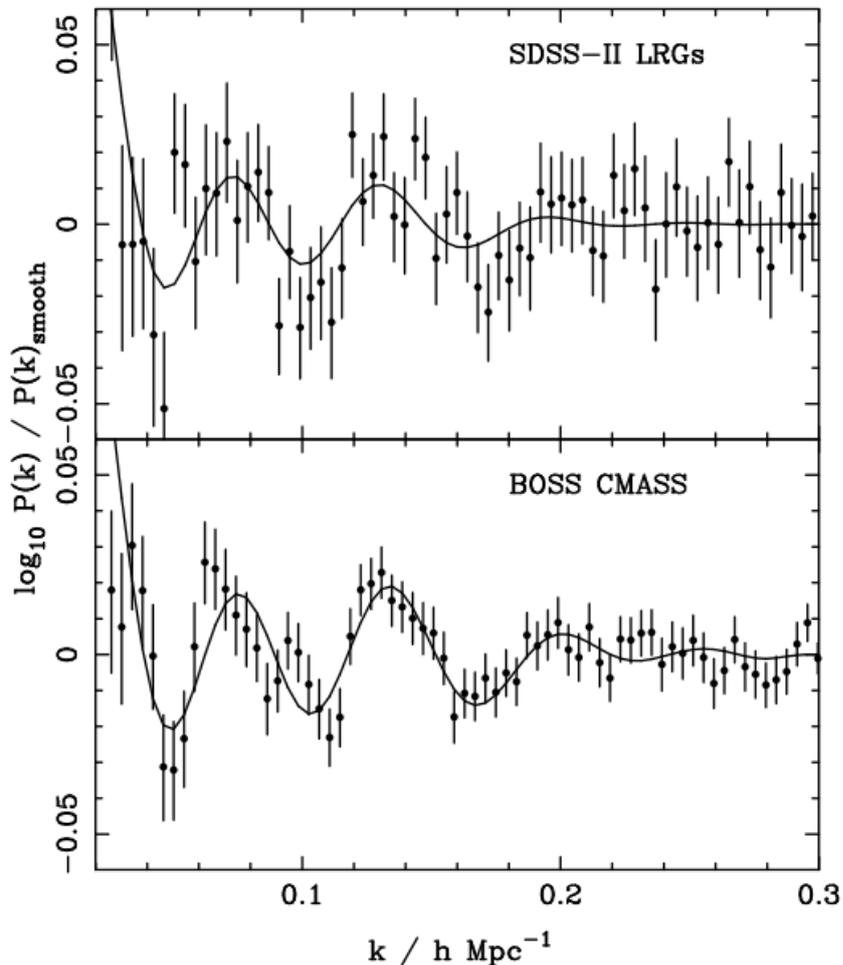
## 2.2.4 The Effect of Baryons

Interactions between cold dark matter and radiation occur only indirectly: both of them influence the gravitational potential  $\Phi$  in different epochs, and the potential in turn in-

fluences the other component. The interaction between them, therefore, relies on their coupling with the gravitational potential. This should not come as a surprise, since eqs. (2.37)-(2.35) clearly couple matter and radiation perturbations to the potential.

Recall also that we have neglected Compton scattering terms to derive eqs. (2.37) and (2.38). Scattering, after all, is not an issue between radiation and a dark matter fluid that by definition does not feel electromagnetic forces.

As mentioned in the first section, the universe in which we actually live does contain baryonic matter, and it is significantly less abundant than its dark counterpart. The perturbation theory treatment we used is therefore incomplete.



**Figure 2.3:** The ratio between the experimentally obtained  $P(k)$  from the SDSS-II and BOSS surveys and a smooth, featureless  $P_{\text{CDM}}(k)$ . The oscillation pattern is very clear in the best-fit curve (solid line). Extracted from [27].

At very early times, the universe was hotter<sup>5</sup> and the baryon-photon plasma was tightly coupled via Compton scattering. Not too long after the matter-radiation equality, approximately at  $z = 1100$ , baryons recombine to form neutral hydrogen and the universe

<sup>5</sup>See Appendix A.

suddenly becomes transparent to radiation. In other words, the Compton cross-section is not large enough to keep photons and baryons coupled anymore.

Although we have not paid too much attention to the evolution of the radiation perturbation  $\Theta$ , directing our focus to  $\delta$ , we surely expect the oscillation in the potential represented in figure 2.1 to be followed by the baryon-photon plasma. A physical interpretation to this oscillation lies in the competition between gravity and pressure in the primordial universe.

As gravity tends to accumulate the perturbation and make it denser, radiation pressure will dominate at some point and pull the fluid apart. Schematically, we can write this simple harmonic oscillator as

$$\ddot{\delta} - [\text{Pressure} - \text{Gravity}] \delta = 0.$$

We can then think of it as a sequence that starts with the gravitational collapse of an overdensity, followed by the response of the enhanced radiation pressure, followed by an intense rarefaction of the fluid, when gravity starts dominating again and the cycle restarts with a recollapse. For radiation, a similar qualitative picture can be drawn, which roughly captures behavior of the plasma:

$$\ddot{\Theta}_0 + k^2 c_s^2 \Theta_0 = f. \quad (2.59)$$

Equation (2.59) also represents a forced harmonic oscillator, with  $f$  being a driving force due to gravity. Here,  $c_s$  is the sound speed of the combined baryon-photon fluid, and it is given by

$$c_s = \sqrt{\frac{1}{3 \left(1 + \frac{3\rho_b}{4\rho_\gamma}\right)}}. \quad (2.60)$$

Equation (2.59) qualitatively translates the oscillations on the matter distribution, shown on figure 2.3, to the anisotropies found in the CMB angular power spectrum (figure 2.4). Looking at eq. (2.60), if the amount of baryons is enhanced, the sound speed will go down and, keeping the forced oscillator in mind, the frequency will decrease too. We can think of the quantity of baryons as a component of inertia, changing the frequencies at which we see the oscillation peaks.

The tight coupling we have been describing, however, does not account for the damping that increases towards the small scales. To take it into account, we must remember that, not only photons will follow the exact movement of baryons, but also they will suffer diffusion.

The path a photon follows in a sea of electrons is well described by a random walk. The mean free path between two scatterings is roughly  $\lambda_{\text{MFP}} \sim 1/n_e \sigma_T$ , so that if the

scattering rate  $\sigma_T$  or the density of electrons  $n_e$  becomes too large, the free path will become too short. Now, the number of scatterings suffered by a photon in a characteristic time scale  $H^{-1}$  is the product between the rate and the time ( $n_e\sigma_T H^{-1}$ ). Since the total distance traveled in a random walk is given by the mean free path times the square root of the number of steps, the mean distance a photon moves in a Hubble time is

$$\lambda_D \sim \lambda_{\text{MFP}} \sqrt{n_e \sigma_T H^{-1}} = \frac{1}{\sqrt{n_e \sigma_T H^{-1}}}. \quad (2.61)$$

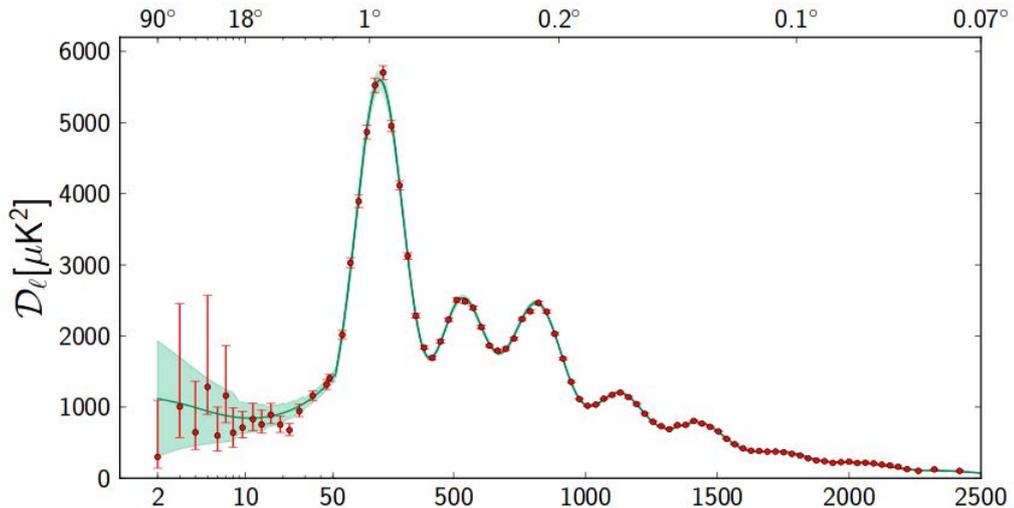
Perturbations smaller than  $\lambda_D$ , therefore, will be suppressed. This corresponds to high  $k$ -modes in Fourier space, as we already expected.

Once photons and baryons decouple, however, the oscillation will become “frozen” and leave an imprint on both the matter and radiation distributions. We can think of it as the *sound horizon* at the time of last scattering ( $z \approx 1100$ ). This scale is a measure of how much a sound wave has traveled by the time  $\eta$  if the expansion starts at  $\eta = 0$ . The comoving sound horizon is given then by

$$r_s(\eta) \equiv \int_0^\eta d\eta' c_s(\eta'). \quad (2.62)$$

Using the sound speed (2.60) in the above equation and knowing the conformal time at the last scattering from eq. (2.15), we find that

$$r_s(\eta = \text{last scattering}) \approx 100 h^{-1} \text{ Mpc}.$$



**Figure 2.4:** The CMB temperature (TT) angular power spectrum obtained from the Planck satellite [28]. The horizontal axis is the multipole  $l$ , and the plot also shows the cosmic variance errors as green shaded regions. The positions of the peaks and small-scale damping are strongly affected by the baryon density  $\rho_b$ .

This scale can be understood as a *standard ruler* for, after decoupling, matter and radiation will evolve independently, and that distance scale will remain pristine<sup>6</sup>.

The resulting effect on the galaxy power spectrum  $P(k)$  will be a damped oscillatory pattern corresponding to the Bessel function in eq. (2.53), on scales smaller than the turnover mode  $k_{\text{eq}}$  from eq. (2.17).

Since luminous matter, and not its dark counterpart, is the component most influenced by these *baryon acoustic oscillations* (BAOs), we also verify that the oscillatory pattern is damped by a factor  $\Omega_b = 0.05$ . This feature has already been found in actual galaxy surveys of luminous red galaxies (SDSS-II) and LRG+quasars (BOSS) [27], as shown in figure 2.3.

---

<sup>6</sup>This statement is only exact within the linear perturbation theory framework we have been employing. The velocity equations will play a role in adding noise to this standard ruler, blurring the distance scale. For more on how to overcome this issue, see [29].

# Chapter 3

## Power Spectrum Estimation

The goal of the previous chapter was to build up a consistent framework to link cosmological perturbations (our theoretical knowledge about the inhomogeneous universe) and observations. The transfer function  $T(k)$  and the growth function  $D_1(a)$  contain relevant information about the evolution of those perturbations, and the bottom line is that we can analyze the statistics of matter distribution to obtain that information.

The rest of this work focuses on the other part of the game: how to estimate the matter power spectrum  $P(k)$  directly from redshift surveys? In fact, our goal here goes beyond estimating a spectrum, we want to do it in an, at least in principle, *optimal* way. The powerful modern surveys will deliver data more accurate than ever, and it would be a waste of time (and money) not to extract as much information as possible from it. We should start by understanding how statistical *noise* can influence our analysis.

### 3.1 Signal $\times$ Noise

The fundamental quantity we will use in our statistical study of galaxy surveys will again be the density contrast  $\delta$ , to wit

$$\delta(\vec{x}) = \frac{\rho(\vec{x}) - \langle \rho \rangle}{\langle \rho \rangle} \quad (3.1)$$

where now  $\langle \rho \rangle$ , the average density over all the universe, is the average density within the whole survey area. From this definition, it is easy to check that  $\langle \delta \rangle = 0$ .

We will deal with statistical properties, and since we have a definition for the second moment in Fourier space (2.23), it would be useful to have an analog for the real space.

We define the *correlation function*<sup>1</sup> as

$$\langle \delta(\vec{x})\delta(\vec{x} + \vec{r}) \rangle = \xi(\vec{r}) = \xi(r), \quad (3.2)$$

where we again assume isotropy to disregard the direction of  $\vec{r}$ . We can recover our previous definition, from eq. (2.23), by taking the Fourier transform of (3.2):

$$\begin{aligned} \langle \delta(\vec{k})\delta^*(\vec{k}') \rangle &= \left\langle \int d^3r e^{i\vec{k}\cdot\vec{r}}\delta(\vec{r}) \int d^3x e^{-i\vec{k}'\cdot(\vec{r}+\vec{x})}\delta(\vec{r} + \vec{x}) \right\rangle \\ &= \int d^3r e^{i(\vec{k}-\vec{k}')\cdot\vec{r}} \int d^3x e^{-i\vec{k}\cdot\vec{x}} \underbrace{\langle \delta(\vec{r})\delta(\vec{r} + \vec{x}) \rangle}_{\xi(x)} \\ &= (2\pi)^3 \delta_D(\vec{k} - \vec{k}') P(k), \end{aligned}$$

which clearly recovers our previous definition. This calculation also yields an important connection between the correlation function and the power spectrum:

$$P(k) = \int d^3x e^{-i\vec{k}\cdot\vec{x}}\xi(x) = 4\pi \int dx x^2 \sin(kx)\xi(x).$$

This means that, by definition, the correlation function and the power spectrum are a Fourier pair. Notice that we started using  $D$  as a subscript on  $\delta$  to differentiate between the density contrast and the Dirac delta-function. In principle,  $\xi$  and  $P$  contain exactly the same information. This is only true, however, for the analytical form of these quantities, and will not necessarily hold when they are estimated directly from data. We will comment on this issue later. Notice also that the previous calculation defines our Fourier transform convention.

The power spectrum and correlation function are the signal we try to measure at the end of the day. We now turn to the types of noise that can hinder our analysis. Problems arise when the volume we are dealing with is finite (and possibly irregularly shaped) and when the density field  $\delta(\vec{x})$  is not continuous across the universe, making itself available only through some other stochastic point process. The first problem we call *cosmic variance* and the second will add what we call *Poisson shot noise* (or simply *shot noise*) to our estimates.

### 3.1.1 Poisson Noise

We begin by describing the Poisson noise. For what follows, we will assume that galaxies form approximately a Poisson sampling of the density field  $\delta$ . To be more specific, this

---

<sup>1</sup>In the context of galaxy surveys, the correlation function (or 2-point auto-correlation function) can be thought of as the excess in probability of finding a galaxy in the position  $\vec{x} + \vec{r}$ , given that one was observed in  $\vec{x}$ .

means that in every fraction  $i$  of the whole survey volume (which we shall call a *cell*), the number of observed galaxies represents a Poisson point-process with a mean  $\bar{n}_i$ , with units of  $[\text{volume}]^{-1}$ .

We need to review the definition (3.1) to properly account for a discrete field. We can describe such a distribution by taking  $n(\vec{x}_i)$  as the position of  $i$ th galaxy, so that

$$n(\vec{x}) = \sum_i \delta_D(\vec{x} - \vec{x}_i).$$

Equation (3.1) then becomes

$$\delta(\vec{x}) = \frac{n(\vec{x}) - \bar{n}(\vec{x})}{\bar{n}(\vec{x})}. \quad (3.3)$$

The average number of galaxies  $\bar{n}(\vec{x})$ , which is precisely the mean number density that characterizes the Poisson distribution, can now be a complicated function of space. It usually varies significantly with the radial position, and can also vary angularly. We include in it, therefore, every kind of systematic uncertainty and geometric issue of the survey<sup>2</sup>. For our description of shot noise, however, we will take  $\bar{n}$  to be uniform and constant throughout the survey volume – which is a strong assumption, but nevertheless necessary in order to study the simpler aspects of statistical noise. We will drop this simplifying assumption soon, when dealing with the *window* of the survey.

The main idea here is that, since we are dealing with a Poisson process for the  $i$ th fraction of the survey volume ( $V_i$ ),  $\bar{n}(\vec{x}_i)V_i$  represents the *expected* number of galaxies in that volume, not the real, observed number, which will be given by  $\mathbb{P}\{\bar{n}(\vec{x}_i)V_i\}$ , where  $\mathbb{P}\{\lambda\}$  is a Poisson distribution with mean  $\lambda$ .

Taking the Fourier transform  $\delta(\vec{k})$  in the discrete definition (3.3), we have

$$\delta(\vec{k}) = \int d^3r \delta(\vec{x}) e^{i\vec{k}\cdot\vec{r}} \rightarrow \frac{1}{\bar{n}} \sum_i n_i e^{i\vec{k}\cdot\vec{r}_i} - \delta_{\vec{k},0}^K,$$

where  $\delta^K$  is the Kronecker delta and the discrete limit comes from slicing the volume  $V$  into cells  $dV_i$  that are small enough so that each cell contains either 0 or 1 galaxy,  $n_i = 0, 1$ , then  $\langle n_i \rangle = \langle n_i^2 \rangle$ , which is a signature of the Poisson distribution (i.e. it

---

<sup>2</sup>All sorts of real-world effects are encapsulated in  $\bar{n}(\vec{x})$ . Radial variations can be attributed not only to the limitation in magnitude of our instruments, but also to the evolution of the *luminosity function*, which describes the number counts of objects in redshift bins. The phases of the moon will also change the minimum magnitude achievable, and the observations will be shallower or deeper according to the time of the month. Other effects might influence the angular part of the selection. Galactic extinction, for instance, will redden objects observed through the dust of our galaxy. The anisotropic distribution of nearby globular clusters will unevenly contaminate galaxy samples with stars. Atmospheric conditions such as clouds or humidity will change the *seeing* during the observing night, and even non-uniform sensitivity across the surface of the CCD will play a role. The function  $\bar{n}(\vec{x})$ , thus, is a very compact way to introduce highly non trivial complications.

naturally appeared when we discretized the distribution). Now, knowing that

$$\langle n_i \rangle = \bar{n} dV_i \quad (3.4)$$

$$\langle n_i n_j \rangle_{i \neq j} = \bar{n}^2 dV_i dV_j [1 + \langle \delta(\vec{x}_i) \delta(\vec{x}_j) \rangle], \quad (3.5)$$

we can calculate the ensemble average  $\langle \delta_d(\vec{k}_1) \delta_d^*(\vec{k}_2) \rangle^3$ , which yields:

$$\begin{aligned} \langle \delta_d(\vec{k}_1) \delta_d^*(\vec{k}_2) \rangle &= \left\langle \frac{1}{\bar{n}^2} \sum_{ij} n_i n_j e^{i\vec{k}_1 \cdot \vec{r}_i} e^{-i\vec{k}_2 \cdot \vec{r}_j} - \frac{1}{\bar{n}} \sum_i n_i e^{i\vec{k}_1 \cdot \vec{r}_i} \delta_{\vec{k}_2,0}^K \right. \\ &\quad \left. - \frac{1}{\bar{n}} \sum_j n_j e^{i\vec{k}_2 \cdot \vec{r}_j} \delta_{\vec{k}_1,0}^K + \delta_{\vec{k}_1,0}^K \delta_{\vec{k}_2,0}^K \right\rangle. \end{aligned}$$

The cross terms reduce to Kronecker deltas and then we have

$$\begin{aligned} \langle \delta_d(\vec{k}_1) \delta_d^*(\vec{k}_2) \rangle &= \frac{1}{\bar{n}^2} \sum_{ij} \langle n_i n_j \rangle e^{i\vec{k}_1 \cdot \vec{r}_i} e^{-i\vec{k}_2 \cdot \vec{r}_j} - \delta_{\vec{k}_1,0}^K \delta_{\vec{k}_2,0}^K \\ &= \frac{1}{\bar{n}^2} \sum_{i \neq j} \langle n_i n_j \rangle e^{i\vec{k}_1 \cdot \vec{r}_i} e^{-i\vec{k}_2 \cdot \vec{r}_j} + \frac{1}{\bar{n}^2} \sum_i \langle n_i \rangle e^{i\vec{r}_i \cdot (\vec{k}_1 - \vec{k}_2)} - \delta_{\vec{k}_1,0}^K \delta_{\vec{k}_2,0}^K \\ &= \langle \delta(\vec{k}_1) \delta^*(\vec{k}_2) \rangle + \frac{\delta_{\vec{k}_1, \vec{k}_2}^K}{\bar{n}} \end{aligned}$$

where the last step comes from taking the continuum limit to retrieve the initial expected value, and assumes  $\vec{k}_1$  or  $\vec{k}_2 \neq 0$ . Schematically, what we just found is that

$$P_{\text{discrete}}(k) = P(k) + \frac{\delta_{\vec{k}_1, \vec{k}_2}^K}{\bar{n}}. \quad (3.6)$$

Equation (3.6) is telling us two things. First, for a given distribution of galaxies, the power spectrum will have an added noise which is correspondent to the inverse global number density of those galaxies. This should not come as a surprise, for it is natural to think that, the higher the density of some kind of discrete tracer of a continuous field, the more accurate our information of that field is. In the limit  $\bar{n} \rightarrow \infty$ , we have perfectly recovered the continuous field and therefore Poisson shot noise is null.

Secondly, shot noise is *diagonal* in  $k$ , as the Kronecker delta in eq. (3.6) shows. This is a fortunate fact, because we know that Poisson fluctuations will *correlate* the estimate of different  $k$ 's. As a final remark, one can think of shot noise as the spurious correlation of a galaxy with itself, which does not represent true information about the density field and must therefore be subtracted.

---

<sup>3</sup>Subscripts “ $d$ ”, only for the present calculation, stand for *discrete*.

### 3.1.2 Cosmic Variance

The range of wave numbers  $k \sim 1/\lambda$  that can be assessed by a survey obviously depends on the characteristic size of the survey itself. We will not be able to measure and make statistics with, for instance, a scale  $k = 0.01 \text{ Mpc}^{-1}$  ( $\lambda \sim 10^2 \text{ Mpc}$ ) with a survey that does not cover that (comoving) volume. We also will not be able to extract useful information about this survey in its extremely small scales, because we often reach a limit of computational resources in terms of how fine a Fast Fourier Transform grid can be.

In other words, at least for the largest scales, the limitation in  $k$  has a physical and statistical meaning. Not only some modes are completely inaccessible, but also the larger modes that the survey does contain are very poor in statistics, which will affect the precision (the size of the error bars) of our estimates. This limitation is called *cosmic variance*, (whose definition is close to what is called *sample variance*) and, along with shot noise, is one of the first problems encountered in power spectrum estimation.

For a more quantitative notion on how cosmic variance acts, consider a survey of characteristic volume of order  $V$ . The number of subvolumes of characteristic length  $L$  that fit inside  $V$  is of order  $V/L^3$ . Consider now that some quantity that we want to evaluate follows an underlying probability distribution that is approximately Gaussian, within each subvolume. We know that the uncertainty in a gaussian distribution falls as  $\sim \sqrt{N}$  where  $N$  is number of events (or samples) drawn from that distribution. This means that, for a characteristic length which is around half the survey size  $\sigma_{\text{rel}} \frac{\sqrt{2}}{2} \sim 0.7$ .

That means that there is an intrinsic limitation on how much an error bar can be shrunk which corresponds to around 70% of the signal, for modes close to the length of the survey. These scales naturally correspond to the largest scales of the power spectrum, i.e. those that contain information about the early universe. This problem would be easily circumvented if we had more *realizations* of the density field on these large scales – more realizations of the universe itself – which is clearly not possible.

This problem is not limited to galaxy surveys, it also affects CMB analysis. For CMB, the important observable is the spherical harmonic decomposition  $Y_m^l(\theta, \phi)$  of the temperature (or polarization) field. For a mode  $l$ , there are  $2l+1$  “independent subvolumes” which correspond to number of values of the azimuthal number  $m$ . The intrinsic uncertainty for the CMB spectrum is then roughly

$$\frac{\Delta C_l}{C_l} \approx \sqrt{\frac{2}{2l+1}}.$$

## 3.2 General Formalism

The least we expect from an estimator, in a qualitative way, is for it to be both *accurate* and *precise*. For a method to be accurate, it means that, given that the act of measuring a quantity has an intrinsic probability distribution of output values<sup>4</sup>, then the ensemble average of all measurements coincides with the true value of that quantity. If an offset exists between the average of a sufficiently large ensemble and the true value, this offset is called a *bias*. In a more quantitative form, suppose that an estimator  $\hat{E}$  is derived for a physical quantity whose true value is  $E_0$ , then  $\hat{E}$  is unbiased if

$$\langle \hat{E} \rangle = E_0.$$

Precision, however, does not reflect how close to the true value the measurement gets. It quantifies how narrow the distribution function is around its ensemble average: the standard deviation  $\sigma$ . In the approximation that the distribution is approximately gaussian around its peak, we can think of  $\sigma$  as defining a  $\sim 68\%$  confidence level region surrounding it. This section follows closely the work of Tegmark *et al* (1998)[30], and a series of related papers.

These two features are the least we expect. Apart from that, there are still the real-world necessities of *feasibility* and *efficiency*. When considering parameter estimation, for instance, the usual Bayesian statistical analysis for a data set  $\mathbf{x} = (x_1, \dots, x_N)$  (with ensemble averages  $\mu_i = \langle x_i \rangle$ ) to constrain a parameter vector  $\Theta$  is to write down the probability distribution  $f(\mathbf{x}, \Theta)$  (called *likelihood*) and reinterpret it to find the probability of the parameters given the data, using Bayes's theorem<sup>5</sup>. If  $\mathbf{x}$  is the “raw” data set containing the positions of  $N$  observed galaxies, the likelihood is impossible to be computed, with the CPU time growing faster than exponentially with  $N$  for the correlations between the  $N(N+1)/2$  galaxy pairs.

Following the general formalism introduced in [31, 30], we take  $\mathbf{x}$  to be a compressed data set instead, and subsequently make sure that such compression does not destroy data. The convenient choice in power spectrum estimation is to consider *band-power estimates* in shells that exhaust  $k$ -space. Even for the compressed data set, within the (sometimes inaccurate) simplification that a multivariate Gaussian (3.7) for the likelihood

---

<sup>4</sup>We are considering a quantity that is deterministically defined, so that there is indeed a true value for it. In this case, we are simply introducing a probability distribution to quantify the ignorance in our instruments/methods.

<sup>5</sup>Bayes's theorem can be neatly expressed as  $P(A|B) = P(B|A)P(A)/P(B)$  provided that  $P(B) \neq 0$ , where  $P(B|A)$  is the likelihood function itself,  $P(A)$  is the *prior*, which encapsulates our prejudices regarding the parameter  $A$ ,  $P(B)$  is a normalization called *evidence*, and  $P(A|B)$  is the *posterior* probability. In practical terms, we use this theorem to invert conditional probabilities and find the probability of a set of parameters given the data. For this, we must make assumptions on how we obtain data (the likelihood) and on the values we expect for the inferred parameters (the prior). The explicit quantification of our assumptions is in stark contrast with the *frequentist* approach in probability.

is applicable

$$G(\mathbf{x} = x_1, x_2, \dots, x_N) = \frac{1}{(2\pi)^{N/2} |C|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^t C^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}, \quad (3.7)$$

where  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_N)$ , there is still a time-consuming step of calculating the determinant and inverse of the *covariance matrix*  $C$ . For an  $N \times N$  matrix, this step alone scales as  $N^3$  in CPU time. Another problem is that perhaps none but the most powerful computers can even handle, in terms of RAM memory, matrices such that  $N \sim 10^6$ . A method that yields an analytic expression for the covariance, then, is highly desirable. We will see that this is indeed the case for the estimators in the literature and for those derived in the present work.

A complete description of the errors of an estimator is given by the covariance matrix  $C$ , introduced in eq. (3.7). In a sense, it quantifies the notion of precision, extending it also to the case when different band-power estimates (or bins) are not completely independent, in which case they are said to be *correlated*. When the  $x_i$  of  $\mathbf{x}$  are random variables with finite variance, the covariance matrix is defined as

$$C_{ij} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{N1} & \sigma_{N2} & \cdots & \sigma_N^2 \end{pmatrix} \quad (3.8)$$

where  $\sigma_{ij} = \langle (x_i - \mu_i)(x_j - \mu_j) \rangle$ ,  $\sigma_i^2 \equiv \sigma_{ii}$ .

The off-diagonal terms in the covariance matrix express the correlation between quantities, or, in our context, between different bins of  $k$  in the estimated  $P(k)$ . The diagonal terms  $\sigma_i^2$  represent the square of what we quote as *error bars*, the squared standard deviation (or *variance*). A method that yields perfectly uncorrelated errors between different bins will have a diagonal covariance. That is the best we can hope for, since it trivially simplifies the matrix operations required by eq. (3.7).

A related quantity is the *correlation matrix*, defined as

$$Corr_{ij} = \frac{C_{ij}}{\sqrt{\sigma_{ii} \sigma_{jj}}}. \quad (3.9)$$

By this definition, the diagonal of  $Corr_{ij}$  is identically unitary. It is therefore a version of the covariance matrix normalized by its diagonal, its error bars. The effect of this normalization is to subtract the structure of the diagonal terms and enhance the cross-covariances in the off-diagonals. The cross-correlations then become more apparent.

If the covariance matrix represents the error bars in our estimates and how correlated

different band-powers are, and if in principle we can find an analytical covariance, then we should try to make it as small as possible. The smaller the error bars, the better we constrain parameters of interest. It turns out there is a well-established framework to do so. It relies heavily on the *Fisher information matrix*, which we define as

$$F_{ij} \equiv \left\langle \frac{\partial^2 \mathcal{L}}{\partial \theta_i \partial \theta_j} \right\rangle, \quad (3.10)$$

where  $\mathcal{L} = -\ln f$ , the logarithmic of the likelihood.

To better understand what the Fisher matrix means, let us Taylor expand  $\mathcal{L}$  around its maximum value, the vector  $\Theta_0$  of true parameters

$$\mathcal{L}(\mathbf{x}, \Theta) = \mathcal{L}(\mathbf{x}, \Theta_0) + (\mathbf{x} - \boldsymbol{\mu}) \left. \frac{\partial \mathcal{L}}{\partial \theta_i} \right|_{\Theta=\Theta_0} + \frac{(\mathbf{x} - \boldsymbol{\mu})^2}{2} \left. \frac{\partial^2 \mathcal{L}}{\partial \theta_i \partial \theta_j} \right|_{\Theta=\Theta_0} + \mathcal{O}\{3\} \quad (3.11)$$

By definition, the first derivative will vanish at the maximum point, since the likelihood has zero slope there. Also, we will consider the third-order terms to be subdominant, then the relevant term will be the second-order one. Since  $f = \exp(-\mathcal{L})$ , we see that the likelihood close to the maximum is Gaussian to a very approximation.

We can even rewrite  $f$  explicitly as

$$f \propto \exp \left[ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^t \frac{\partial^2 \mathcal{L}}{\partial \theta_i \partial \theta_j} (\mathbf{x} - \boldsymbol{\mu}) \right] \quad (3.12)$$

and immediately associate the inverse covariance matrix to the second derivative of  $\mathcal{L}$ , by comparing it with (3.7). The Fisher information matrix is then tightly related to the expectation value of the covariance at the maximum likelihood point. The second derivative, by its turn, is also a measure of the curvature of this likelihood peak. As we could expect, then, a sharp peak in  $f$  would have a small covariance around it, and a large information content  $F$ . In other words, the inverse Fisher matrix roughly estimates the covariance  $C$ .

Nevertheless, we can refine this statement a bit more. Still in the Gaussian case, from the definition (3.7) we have, after dropping an irrelevant constant factor,

$$2\mathcal{L} = \ln \det C + (\mathbf{x} - \boldsymbol{\mu})^t C^{-1} (\mathbf{x} - \boldsymbol{\mu}). \quad (3.13)$$

We can also recast  $C_{ij}$  from (3.8) in a more neat way as

$$C = \langle (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^t \rangle \quad (3.14)$$

and define a data matrix as  $D \equiv (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^t$ . Then, using the matrix identity

$\ln \det C = \text{Tr} \ln C$  and the associativity of the trace, we rewrite (3.13) as

$$2\mathcal{L} = \text{Tr}[\ln C + C^{-1}D]. \quad (3.15)$$

We now want to differentiate  $\mathcal{L}$  with respect to its parameters  $\theta$ . For this we will use the usual comma notation  $C_{,i} \equiv \partial C / \partial \theta_i$ . Using the chain rule, we have

$$(C^{-1})_{,i} = -C^{-1}C_{,i}C^{-1} \quad \text{and} \quad (\ln C^{-1})_{,i} = C^{-1}C_{,i}, \quad (3.16)$$

which we use to obtain

$$2\mathcal{L}_{,i} = \text{Tr} [C^{-1}C_{,i} - C^{-1}C_{,i}C^{-1}D + C^{-1}D_{,i}]. \quad (3.17)$$

Since we are evaluating the likelihood at its maximum point (the true parameters  $\Theta$ ), the expectation values of the derivatives of  $D$  can be written as

$$\langle D \rangle = C \quad (3.18)$$

$$\langle D_{,i} \rangle = 0 \quad (3.19)$$

$$\langle D_{,ij} \rangle = \boldsymbol{\mu}_{,i} \boldsymbol{\mu}_{,j}^t + \boldsymbol{\mu}_{,j} \boldsymbol{\mu}_{,i}^t \quad (3.20)$$

From eqs. (3.17) and (3.18)-(3.20), we can check that, as expected,  $\langle \mathcal{L} \rangle = 0$ . Applying the chain rule to (3.17), we get

$$\begin{aligned} 2\mathcal{L}_{,ij} = & \text{Tr}[-C^{-1}C_{,i} - C^{-1}C_{,j} + C^{-1}C_{,ij} \\ & + C^{-1}(C_{,i}C^{-1}C_{,j} + C_{,j}C^{-1}C_{,i})C^{-1}D \\ & - C^{-1}(C_{,i}C^{-1}D_{,j} + C_{,j}C^{-1}D_{,i}) \\ & - C^{-1}(C_{,ij}C^{-1}D - C^{-1}D_{,ij})], \end{aligned}$$

which, going back to the definition (3.10), neatly reduces to a short and useful expression:

$$F_{ij} = \langle \mathcal{L}_{,ij} \rangle = \frac{1}{2} \text{Tr} [C^{-1}C_{,i}C^{-1}C_{,j} + C^{-1}\langle D_{,ij} \rangle]. \quad (3.21)$$

In the case of null expectation values for  $\boldsymbol{\mu}$  and going back to the usual partial derivative notation, we finally arrive at

$$F_{ij} = \frac{1}{2} \text{Tr} \left[ C^{-1} \frac{\partial C}{\partial \theta_i} C^{-1} \frac{\partial C}{\partial \theta_j} \right], \quad (3.22)$$

which is a very well-known result, obtained also by [32, 33].

Not only we have found a connection between  $F$  and  $C$ , but also we have a pow-

erful theorem relating these 2 quantities that will be very useful. This theorem is the *Cramér-Rao inequality* [34], which asserts that, for an unbiased estimator, when all the parameters are being estimated from data, we are assured that

$$\sigma(\theta_i) \geq \sqrt{(F^{-1})_{ii}} \quad (3.23)$$

where  $\sigma(\theta_i)$  is the uncertainty for the parameter  $\theta_i$ .

All the statistical subtleties discussed above were employed by [30, 31, 32, 33] and others to maximize the gains from analyzing galaxy surveys. They reinterpreted pre-existing useful theorems in statistics in light of cosmological data analysis. The goal was to develop *maximum likelihood* (ML) estimators. In fact, another important result obtained in those works was the connection between the statistical tools of CMB and those of galaxy surveys. In the CMB formalism, the definition of a *pixel* that contains information comes almost naturally, for it is a 2D discretized problem.

It turns out that power spectrum estimates can be treated along the same lines, regarding subvolumes of the galaxy distribution as pixels containing noisy information about the actual distribution. With a sensible definition of pixels and how they carry information, we can reinterpret eq. (3.22) as the Fisher matrix for (galaxy) *counts-in-cells* and eq. (3.8) as the *pixel covariance*. But how do we accurately pixelize this 3D problem?

We should make an interlude now and expose a clear distinction between the methods under analysis. The maximum likelihood methods explained above culminated in what is known as the *Karhunen-Loève* (also called *linear*) method. The KL method employs linear operations to the data to extract the relevant information, more specifically in terms of Singular Value Decompositions. Even though it is widely applicable to CMB problems (which are only 2-dimensional in essence), the amount of data in a 3-dimensional survey makes its application much more cumbersome. The alternatives are known as the *traditional* method (the FKP and PVP methods <sup>6</sup> - see sections 3.3 and 3.4) and *quadratic* method (the multi-tracer estimator - see 4.1). In the end, the gains of the KL method over the others is not so expressive. As an example, while the SDSS-II  $P(k)$  estimation used a version of the KL method [35], the SDSS-III (BOSS) returned to the usual FKP recipe [27]. We will, therefore, focus on the traditional and quadratic methods for what follows.

The first step is to quantify our band power estimates  $q_i$ , which are the estimated amplitudes of the power spectrum in a given band  $i$ . A very general way to do this is

---

<sup>6</sup>We are making reference to two methods that have not been explained yet, but this interlude serves only as a contextualization for the derivations that will be studied soon.

defining it as

$$q_i = \int \int E_i(\vec{r}, \vec{r}') \frac{n(\vec{r})}{\bar{n}(\vec{r})} \frac{n(\vec{r}')}{\bar{n}(\vec{r}')} d^3r d^3r', \quad (3.24)$$

where the  $E_i$  are pair-weighting functions. Different methods basically consist in finding different  $E_i$  functions. Ultimately, the weighting scheme is what we need to optimize in order to obtain an estimator that is, at least under approximations, *optimal*, in the sense that it has the minimum possible variance. The expectation value of the band power  $q_i$  is then

$$\langle q_i \rangle = W_i(0) + \int \frac{E_i(\vec{r}, \vec{r}')}{\bar{n}(\vec{r})} d^3r + \int W_i(\vec{k}) P(k) \frac{d^3k}{(2\pi)^3}. \quad (3.25)$$

We will make this step more explicit when analyzing the estimators in the literature, in the next sections. The first term on the right-hand side is the 3-dimensional *window function* of the survey, evaluated at  $\vec{k} = 0$ . This window is defined as  $W(\vec{k}) = E_i(\vec{k}, \vec{k})$ , where

$$E_i(\vec{k}, \vec{k}') \equiv \int E_i(\vec{r}, \vec{r}') e^{-i\vec{k}\cdot\vec{r}} e^{i\vec{k}'\cdot\vec{r}'} d^3r d^3r' \quad (3.26)$$

is the Fourier transform of  $E_i$ . The second term in (3.25) is precisely the shot noise bias (which one could immediately guess from the  $\propto \bar{n}^{-1}$  dependence) defined as

$$b_i \equiv \int \frac{E_i(\vec{r}, \vec{r}')}{\bar{n}(\vec{r})} d^3r. \quad (3.27)$$

The last term on the right-hand side of (3.24) is the desired power spectrum. It is convolved, however, with the window function.

A very important case is when  $E_i$  is a separable function of the galaxy pair for some set of functions  $\psi_i$ :

$$E_i(\vec{r}, \vec{r}') = \psi_i(\vec{r}) \psi_i(\vec{r}')^*,$$

in which case the window can be neatly written as

$$W_i(\vec{k}) = \left| \psi_i(\vec{k}) \right|^2. \quad (3.28)$$

Two consistency relations of great importance should hold for the weighting  $E_i(\vec{r}, \vec{r}')$  and for  $\psi_i(\vec{r})$ . The first is a normalization condition for the weights and for the window function, given by

$$\int E_i(\vec{r}, \vec{r}) d^3r = \int W(\vec{k}) \frac{d^3k}{(2\pi)^3} = 1, \quad (3.29)$$

where the first equality follows from Parseval's theorem. The other condition is such that

$$\int \psi_i(\vec{r}) d^3r = 0, \quad (3.30)$$

which basically means that the average density of the survey will be correctly subtracted. The importance of this condition and its direct application will become clear in the next sections.

With these definitions, we can finally define the overdensity in the  $N$  pixels  $x_1, x_2, \dots, x_N$  that exhaust the survey volume as

$$x_i \equiv \int \left[ \frac{n(\vec{r})}{\bar{n}(\vec{r})} - 1 \right] \psi_i(\vec{r}) d^3r. \quad (3.31)$$

The functions  $\psi_i$  must be carefully chosen. As long as the condition (3.30) holds, we are free to choose  $\psi_i$  so that the pixels are well localized in real space or in Fourier space (in which case they will be fuzzy in real space, perhaps even distributed across the whole survey volume). The importance of this choice is that it will basically filter the wave numbers  $k$  of the band powers, so that they are, hopefully, uncorrelated. Since we want to have control over the correlation between different bins in  $k$ , the popular choice is for compact pixels in Fourier space.

In order to make pixels reasonably well localized in Fourier space, (which will not resemble “pixels” in real space, but we will keep using the same word for it nevertheless) we define the set of modes  $\psi_i$  as plane-waves multiplied by some weight

$$\psi_i(\vec{r}, k) = \phi(\vec{r}) e^{i\vec{k}_i \cdot \vec{r}}. \quad (3.32)$$

Two popular choices in the literature are

$$\phi(\vec{r}) \propto \begin{cases} 1 & \text{inside survey volume} \\ 0 & \text{outside survey volume} \end{cases} \quad (3.33)$$

and

$$\phi(\vec{r}) \propto \frac{\bar{n}(\vec{r})}{1 + \bar{n}(\vec{r})P_0(k)}, \quad (3.34)$$

where (3.33), when applied to the pixelization (3.31), becomes nothing more than the Fourier transform of the overdensity field, without actually weighting different cells. Equation (3.34), on the other hand, proved its usefulness as an optimal estimator for a single-species survey.

The descriptions above define what was called by [30] a *traditional* approach: a separable  $E_i(\vec{r}, \vec{r}')$  with well localized pixels in Fourier space. The power is then simply obtained by squaring the pixels

$$q_i = |x_i|^2$$

and eq. (3.25) becomes

$$\langle q_i \rangle = \phi(0)^2 + \int \frac{\phi(\vec{r})^2}{\bar{n}(\vec{r})} d^3r + \int \left| \phi(\vec{k} - \vec{k}') \right|^2 P(\vec{k}) \frac{d^3k}{(2\pi)^3}. \quad (3.35)$$

The above procedure is useful, but does not necessarily fulfill the condition of minimum variance. Even though in principle many functions  $\phi(\vec{r})$  could be used, it is not clear (in fact, very unlikely) that any choice would yield a minimum variance weighting scheme. In order to do that, one still needs to verify that the weighting is indeed optimal. As we shall soon show, eq. (3.34) has been proven to be optimal by [36] (hereafter FKP).

What if we tried to derive weights which are already optimal from the very beginning? Since we already have a useful statistical method based on the Fisher matrix and, if we can derive an unbiased estimator, the Cramér-Rao theorem will assure us that the corresponding covariance is minimal. We would then think about using some type of *inverse covariance weighting*.

This was explored by [37], and explicitly uses the (assumed known) Fisher matrix for a galaxy survey. We start by defining the estimators  $q_i$  as generalized quadratic forms of the data vectors  $\mathbf{d}$

$$q_i \equiv \mathbf{d}^t E_i \mathbf{d} \quad (3.36)$$

where

$$E_i = C^{-1} C_{,i} C^{-1}. \quad (3.37)$$

We now parameterize the power spectrum by its band estimates in such a way that  $P(k) = p_i$  for  $k_i \leq k < k_{i+1}$ , and group them into a vector  $\mathbf{p}$ . This means that the corresponding Fisher matrix we are looking for is not a function of some general parameters  $\theta_i$ , which in principle could be anything, even cosmological parameters. We are explicitly parameterizing  $P(k)$  so that the convenient choice is  $F_{ij} = F(p_i, p_j)$ .

For the Gaussian case, the Fisher matrix can be written as (3.22) and both the mean and covariance of  $q_i$  are given in terms of  $F$ :

$$\langle \mathbf{q} \rangle = F \mathbf{p} \quad (3.38)$$

$$\langle \mathbf{q} \mathbf{q}^t \rangle - \langle \mathbf{q} \rangle \langle \mathbf{q} \rangle^t = F. \quad (3.39)$$

This guarantees that  $F^{-1} \mathbf{q}$  is an optimal estimator for  $\mathbf{p}$ , since its covariance matrix is the inverse of the Fisher matrix, saturating the Cramer-Rao bound. For sufficiently narrow  $k$  bands, this procedure is *lossless* in the sense that the Fisher matrix for  $\mathbf{q}$ , a compressed (quadratic) data set, contains as much information as the Fisher matrix for the original, uncompressed data set [31]. This characteristic guarantees that, when constraining cosmological parameters, the reduced data set is as powerful as the original

data set, which is surely a desirable feature.

### 3.3 The FKP Estimator

#### 3.3.1 Formalism

A very successful power spectrum estimator, both unbiased and with minimal variance, was obtained in 1994 by Feldman, Kaiser & Peacock [36] in the context of large-scale structure analysis of the one-in-six IRAS-QDOT survey (see, for instance [38]). It consisted in a random selection of 1 out of every 6 galaxies from the IRAS Point Source Catalog, covering  $\sim 74\%$  of the whole sky, with redshifts corresponding to objects at a radial distance between  $20 h^{-1}$  Mpc and  $500 h^{-1}$  Mpc. They analyzed 1824 galaxies in that work. Notice that the number of objects is many orders of magnitude smaller than what we are currently observing, as commented on the Introduction.

FKP start the analysis by considering a function  $F(\vec{r})$  (not to be mistaken with the Fisher matrix  $F_{ij}$ , which will always have a pair of sub-indices):

$$F(\vec{r}) = \frac{w(\vec{r}) [n_g(\vec{r}) - \alpha n_r(\vec{r})]}{[\int d^3r \bar{n}^2(\vec{r}) w^2(\vec{r})]^{1/2}}. \quad (3.40)$$

This expression is nothing more than the pixelization scheme (3.31) with a weight given by  $w(\vec{r})$ . The denominator is a convenient normalization, and follows the condition (3.29) as it should:

$$\int d^3r \bar{n}^2(\vec{r}) w^2(\vec{r}) = 1.$$

There is a subtlety regarding the  $\alpha n_r(\vec{r})$  being subtracted from the discrete galaxy distribution  $n_g(\vec{r})$ . Here,  $n_r(\vec{r})$  is a *random catalog*, artificially created with exactly the same radial and angular selection function  $\bar{n}(\vec{r})$ , but otherwise without structure. It is generated by taking a Poisson process in each cell of the survey, but with a larger number of galaxies than the original catalog. The factor  $\alpha$  matches the number density of this new catalog with the original one's. Quantitatively,

$$\bar{n}_g(\vec{r}) = \alpha \bar{n}_r(\vec{r}). \quad (3.41)$$

FKP explicitly used  $\alpha = 0.1$ , which means that their random catalog had 10 times more galaxies than  $n_g$ . Translating back to [30], we see that this is no more than condition (3.30) being fulfilled so that the discretization (3.31) won't have an extra spurious term. We will comment about the role of random catalogs in greater depth shortly.

The next steps in the “mantra” of the traditional methods are: taking the Fourier transform of  $F(\vec{r})$ , squaring it and taking the expectation value. In other words, what we

want is FKP's version of eq. (3.35). Carrying out that procedure on  $F(\vec{r})$ , what we have is

$$\left\langle \left| F(\vec{k}) \right|^2 \right\rangle = \frac{\int \int w(\vec{r}) w(\vec{r}') \langle [n_g(\vec{r}) - \alpha n_r(\vec{r})] [n_g(\vec{r}') - \alpha n_r(\vec{r}')] \rangle e^{i\vec{k} \cdot (\vec{r} - \vec{r}')} d^3 r d^3 r'}{\int d^3 r \bar{n}^2(\vec{r}) w^2(\vec{r})}. \quad (3.42)$$

Three distinct 2-point functions appear in the above expression, and we must calculate them. The first is  $\langle n_g(\vec{r}) n_g(\vec{r}') \rangle$ , but we already know the result for this one: it is nothing but the real-space version of eq. (3.6), derived in the presence of shot noise. The next one, the cross-correlation  $\langle n_g(\vec{r}) n_r(\vec{r}') \rangle$ , is quite simple: since there will be no shot noise and correlation in this cross term (recall that the physical meaning of shot noise is the counting of an object with itself, which won't happen here), it is just the average of  $n_g$  and  $n_r$ . The remaining one is the correlation of the random catalog with itself,  $\langle n_r(\vec{r}) n_r(\vec{r}') \rangle$ . By definition, there will be no correlations in the random catalog, so the only term that survives besides the ensemble average is shot noise. Collecting these results, we have:

$$\langle n_g(\vec{r}) n_g(\vec{r}') \rangle = \bar{n}(\vec{r}) \bar{n}(\vec{r}') [1 + \xi(\vec{r} - \vec{r}')] + \bar{n}(\vec{r}) \delta_D(\vec{r} - \vec{r}') \quad (3.43)$$

$$\langle n_r(\vec{r}) n_r(\vec{r}') \rangle = \alpha^{-2} \bar{n}(\vec{r}) \bar{n}(\vec{r}') + \alpha^{-1} \bar{n}(\vec{r}) \delta_D(\vec{r} - \vec{r}') \quad (3.44)$$

$$\langle n_g(\vec{r}) n_r(\vec{r}') \rangle = \alpha^{-1} \bar{n}(\vec{r}) \bar{n}(\vec{r}'). \quad (3.45)$$

Now, translating the FKP window function  $W_{FKP}(\vec{k})$  from (3.26) and plugging the 2-point functions into (3.42), we get

$$\left\langle \left| F(\vec{k}) \right|^2 \right\rangle = \int \frac{d^3 k'}{(2\pi)^3} P(\vec{k}') W(\vec{k} - \vec{k}') + P_{\text{shot}}, \quad (3.46)$$

where, with  $W_{FKP}(\vec{k} - \vec{k}') = \left| G(\vec{k} - \vec{k}') \right|^2$  as defined in eq. (3.28),

$$G(\vec{k}) \equiv \frac{\int d^3 r \bar{n}(\vec{r}) w(\vec{r}) e^{i\vec{k} \cdot \vec{r}}}{\left[ \int d^3 r \bar{n}^2(\vec{r}) w^2(\vec{r}) \right]^{1/2}} \quad (3.47)$$

and with the shot noise defined as

$$P_{\text{shot}} = (1 + \alpha) \frac{\int d^3 r \bar{n}(\vec{r}) w^2(\vec{r})}{\int d^3 r \bar{n}^2(\vec{r}) w^2(\vec{r})}. \quad (3.48)$$

Notice that (3.48) already takes into account that the random catalog  $n_r(\vec{r})$  might add shot noise if  $\alpha \gtrsim 1$ . Equation (3.46) is the formal FKP result. To get another insight about the estimator, we can assume the window is compact in  $k$ -space, and sufficiently sharp on the scales in which  $P(\vec{k})$  is sufficiently smooth, so that  $W(\vec{k} - \vec{k}') \rightarrow \delta_D(\vec{k} - \vec{k}')$ ,

in which case our final estimator  $\hat{P}(\vec{k})$  for FKP is just

$$\hat{P}(\vec{k}) = \left| F(\vec{k}) \right|^2 - P_{\text{shot}}. \quad (3.49)$$

The final estimator for the absolute value  $k = \left| \vec{k} \right|$  is an average of  $\hat{P}(\vec{k})$  over shells of volume  $V_k$  in Fourier space:

$$\hat{P}(k) \equiv \frac{1}{V_k} \int_{V_k} d^3 k' \hat{P}(\vec{k}'). \quad (3.50)$$

Until now, there is no reason to believe that this method minimizes variance. The only way to do that in this framework is by specifying the function  $w(\vec{r})$  in such a way that it explicitly does so. Since FKP aims at minimizing variance, the fluctuation in  $\hat{P}(\vec{k})$  must first be analyzed.

The variance in the estimated power of bin  $k$ ,  $\hat{P}(k)$  is given by

$$\sigma_P^2(k) \equiv \left\langle \left[ \hat{P}(k) - P(k) \right]^2 \right\rangle = \frac{1}{V_k^2} \int_{V_k} \int_{V_{k'}} \left\langle \delta \hat{P}(\vec{k}) \delta \hat{P}(\vec{k}') \right\rangle d^3 k d^3 k', \quad (3.51)$$

where  $\delta$  stands for a small variation. If the underlying fluctuation is Gaussian, i.e., if the density field follows a Gaussian distribution, we can write  $\left\langle \delta \hat{P}(\vec{k}) \delta \hat{P}(\vec{k}') \right\rangle = \left| \left\langle F(\vec{k}) F^*(\vec{k}') \right\rangle \right|^2$ . This result is obtained in the Appendix B of FKP, and relies on the same basic assumption that we used for the Fisher matrix and will continue using: the underlying random process is essentially Gaussian. To obtain  $\left\langle F(\vec{k}) F^*(\vec{k}') \right\rangle$  we retrace the steps that led to (3.46) and make the same assumption on the sharpness of the window function to get

$$\left\langle F(\vec{k}) F^*(\vec{k}') \right\rangle \simeq P(\vec{k}) \underbrace{\frac{\int d^3 r \bar{n}^2(\vec{r}) w^2(\vec{r}) e^{i\vec{k} \cdot \vec{r}}}{\int d^3 r \bar{n}^2(\vec{r}) w^2(\vec{r})}}_{\equiv Q(\vec{k})} + (1 + \alpha) \underbrace{\frac{\int d^3 r \bar{n}^2(\vec{r}) w^2(\vec{r}) e^{i\vec{k} \cdot \vec{r}}}{\int d^3 r \bar{n}^2(\vec{r}) w^2(\vec{r})}}_{\equiv S(\vec{k})}, \quad (3.52)$$

which means that

$$\left\langle \delta \hat{P}(\vec{k}) \delta \hat{P}(\vec{k}') \right\rangle = \left| P(\vec{k}) Q(\delta \vec{k}) + S(\delta \vec{k}) \right|^2. \quad (3.53)$$

Notice that in the previous expressions for the variance, there are factors of  $P(k)$ . If we use  $P(k) = \hat{P}(k)$ , i.e. this input power spectrum happens to be equal to the estimated power spectrum, then the error bars are self-consistent in the sense that they were also obtained from the data. If, on the other hand,  $P(k)$  is the fiducial spectrum  $P_0(k)$  for some cosmology, then we obtain the uncertainties for this cosmology.

With this expression for the uncertainties, we can write the relative variance for FKP as

$$\frac{\sigma_P^2(k)}{P^2(k)} = \frac{(2\pi)^3 \int d^3 r \bar{n}^4 w^4 [1 + 1/\bar{n} P(k)]^2}{V_k \left[ \int d^3 r \bar{n}^2 w^2 \right]^2}. \quad (3.54)$$

The last step is to minimize the previous expressions by varying  $w(\vec{r})$ . Requiring that (3.54) be stationary with respect to variations  $w = w_0 + \delta w$ , FKP found the optimal weight to be

$$w(\vec{r}) = \frac{1}{1 + \bar{n}(\vec{r})P(k)}, \quad (3.55)$$

where  $P(k)$  is the fiducial spectrum as explained above. Notice that  $w(\vec{r})$  is always multiplied by  $\bar{n}(\vec{r})$  so that, again drawing a parallel with [30] in the previous section,  $\phi_{FKP}(\vec{r}) = \bar{n}(\vec{r})w(\vec{r})$  in a direct comparison with eq. (3.34).

### 3.3.2 Implementation

With the machinery fully developed and having guaranteed that the covariance is minimal for the weighting scheme given by (3.55), FKP applied their estimator for the IRAS-QDOT survey. The original paper contains a general description on how to actually implement the method, which basically consists in discretizing the integrals by defining, for instance:

$$F(\vec{k}) = \int w(\vec{r}) [n_g(\vec{r}) - \alpha n_r(\vec{r})] e^{i\vec{k}\cdot\vec{r}} d^3r \rightarrow \sum_g w(\vec{r}_g) e^{i\vec{k}\cdot\vec{r}_g} - \alpha \sum_r w(\vec{r}_r) e^{i\vec{k}\cdot\vec{r}_r}. \quad (3.56)$$

The previous equation is a discrete Fourier transform. For the  $\sim 2,000$  galaxies in their original catalog, plus  $\sim 20,000$  galaxies in the random catalog, they explicitly performed this Fourier sum. We can call this a “brute force” Fourier Transform. Of course the computing time for such a method would, roughly speaking, scale linearly with the number of galaxies, and maybe even faster than that when increasing the range of  $k$  bins. In a current catalog, we deal with  $\sim 10^2$ ,  $10^3$  times as many galaxies, plus a random catalog  $n_r$  which, for low signal-to-noise ( $\bar{n}_g \sim 10^{-5}$ ), should have around  $10^5$  as many galaxies as  $n_g$ . A direct Fourier summation is extremely inefficient and, therefore, clearly not a choice.

The most convenient way to proceed is by using some Fast Fourier Transform (FFT) implementation. There are two choices in order to do so. The first one is to turn the arbitrary distribution of galaxies into a square grid using some mass-assignment scheme<sup>7</sup> that creates a smooth field, sampled on the grid positions. The second option is to come up with a method to generate mock galaxy distributions that are already in the form of a square grid. The latter method was extensively used in this work, and will be described soon, in Section 4.4.1. Both methods are equivalent in the sense that they smear the structure on scales smaller than the size of a cell.

In defining the weights  $w$  from eq. (3.55), there is an important caveat: we have to set the fiducial  $P_0(k)$  (we will introduce the subscript 0 from now on, for the fiducial

---

<sup>7</sup>These methods introduce an additional convolution to the power spectrum, with the new window given by the order of the mass-assignment scheme. The final spectrum should obviously be deconvolved.

spectrum). In principle, the dependence  $w = w(\vec{r}, k)$  implies that a single weighting function cannot be optimal for every scale  $k$ . The weights should, therefore, be chosen to have a fiducial  $P_0(k)$  matching the power of the scale which is being estimated. Of course this would require knowing the estimated  $\hat{P}(k)$  before setting the weights and actually estimating the spectrum, which is not possible. One sensible way around this is performing an *iterative* process: an initial guess  $P_0(k_i)$  is taken for each scale  $k_i$ , then the obtained  $\hat{P}(k)$  feeds back a new fiducial guess,  $\hat{P}(k_i) \rightarrow P_0(k_i)$ . The procedure goes on until convergence, or can be truncated if there is danger of underestimating uncertainties with this type of “bootstrapping”.

The iterative process described above, however, is very cumbersome for two reasons. The first is that, for a large number of bins of  $k$ , and several iterations for each bin, the algorithm would take a long time to run, which would render it inefficient when using along with an MCMC method, for instance. Second, it could somehow fine-tune the obtained spectrum according to our theoretical prejudice, implicit in  $P_0(k)$ .

What was done in actual surveys [39, 36] is to choose a single fiducial  $P_0(k) = P_0$  that represents roughly an average of the expected spectrum of the survey. In our applications, we extensively tested the weighting scheme and checked that different choices in  $P_0$  changed the recovered power by a factor much smaller than 1%. Again, the minute details of the application to our mock catalogs will be postponed to the last chapter.

We make one final remark. According to eq. (3.50), the isotropic power spectrum is an average over shells in Fourier space. The explicit calculation of this average in the discretized grid will not really take into account the analytical volume in  $k$ -space, which would be simply  $4\pi k^2 \Delta k$ , where  $\Delta k$  is the width of the bin. Since we are dealing with a discretized quantity, it makes more sense to consider the volume of the shell with width  $\Delta k$  as the *number of modes* that fall within that shell. Indeed, equation (2.4.5) of FKP, reproduced below as (3.57), regards exactly that:

$$\hat{P}(k) = \frac{1}{N_k} \sum_{k < |\vec{k}| < k + \delta k} \left[ |F(\vec{k})|^2 - P_{\text{shot}} \right], \quad (3.57)$$

where  $N_k$  is number of modes between  $k$  and  $k + \delta k$  the discrete version of the volume for the shell.

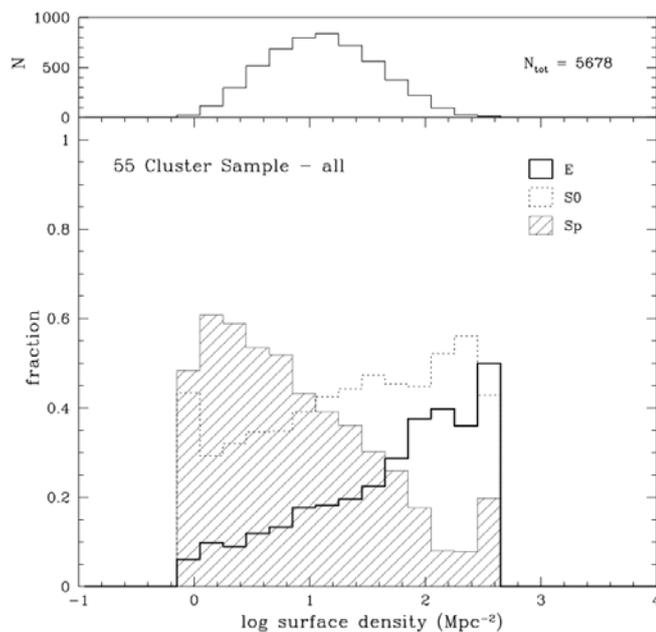
## 3.4 The PVP Estimator

### 3.4.1 Different Galaxies are Different

Until now, we have been considering an abstract, ideal galaxy field  $n_g(\vec{r})$  that happens to follow the underlying dark matter field, and we translate it into  $\delta(\vec{r})$ . No matter the

characteristics of this galaxy field, we say that it is a *tracer* of the dark matter field. Hence, red galaxies, blue galaxies, quasars, etc, are all tracing the same underlying density field, but in different ways. Within the previous assumption, we should expect that such a galaxy field  $n_g$  would follow exactly the same statistics of dark matter.

It has been known for a long time, however, that in *galaxy clusters*, the densest objects in the universe, galaxies of different morphological types inhabit *environments* with different densities, as shown by [40] using 55 low-redshift rich clusters (figure 3.1). While elliptical galaxies (E) and lenticular galaxies (S0) inhabit the denser regions of rich clusters, towards their center, spiral galaxies (S) inhabit the lightest, more external parts. Early-types are therefore more clustered than late-types. It was also shown that this effect could not be explained by dynamical phenomena within the clusters, such as the stripping of gas from spirals to form S0's. This *morphological segregation* is then interpreted as an effect of the environment on the morphology of galaxies.



**Figure 3.1:** The study of 55 rich clusters in the local universe by [40] showed that galactic morphology is correlated with local density. While spirals prefer the outmost regions, early-types prefer the denser, central regions. Extracted from [41], in a reanalysis of the study conducted in [40].

It was later shown by [41] using 10 medium-redshift clusters ( $z \sim 0.5$ ) that the fraction of S0 in these clusters is significantly different from the low-redshift clusters, and the spirals do not have a distribution as steep as in the low-redshift case. This indicates that galaxy evolution must also play a role in determining the *morphology-density* relation.

More recently, [42] analyzed clusters even further ( $z \sim 1$ ) and found segregations which resembled more the case of  $z \sim 0.5$  than the local universe ( $z \sim 0$ ), possibly indicating that segregation starts dominating around  $z \sim 0.5$ .

With these evidences, we can safely say that galaxies are *biased* tracers of the underlying density field. Indeed, it has been shown that morphology will play a role in large-scale structure studies [43]. Our estimators, therefore, should somehow take into account the fact the statistics of different species of tracers will deviate in distinct ways from that of dark matter. One of the first direct approaches was the one by [44] (hereafter PVP), which we describe in this section.

The biasing of peaks in the framework of a Gaussian distribution was studied by [25] and provides hints on how to describe a biased galaxy overdensity field  $\delta_g$ . In the crude approximation of linear, deterministic biasing, we can recast the correlation function of a galaxy field as

$$\xi_g(r) = b^2 \xi(r), \quad (3.58)$$

where  $\xi(r)$  is the correlation function for the underlying dark matter field and  $b = b(\vec{k}, z, \dots)$  is the bias of the corresponding galaxy, which in principle can be a complicated function with non-trivial spatial dependences. Still within the linear approximation, we can translate this relation for the correlation function into the biasing of the overdensity field itself:

$$\delta_g(\vec{r}) = b \delta(\vec{r}). \quad (3.59)$$

Even though eq. (3.58) follows from (3.59), the reverse is not true. The deterministic, linear biasing characterized by eq. (3.59) is specially controversial when dealing with voids, in which case we might well have  $\delta_g < -1$  if  $b \gtrsim 1$ , an unphysical result if one considers the very definition of  $\delta$ , eq. (3.1). Furthermore, within the context of the evolution of gravitational instabilities, we would not expect galaxies to be linearly biased anyway.

The linear, deterministic bias, thus, is not a very realistic option. A more complete picture should include stochastic and nonlinear effects [45]. Nevertheless, it is a starting point, a simplification of an overly complicated problem that potentially mixes nonlinear evolution of perturbations, stochasticity and the dynamical details of galaxy formation. Perhaps a more compelling motivation is that the second moment of the distribution of some types of galaxies, LRGs for instance, does behave, to a good approximation, like it is linearly biased [46].

### 3.4.2 Formalism and Implementation

The power spectrum version of (3.58) is straightforward:

$$P_g(k) = b^2 P_m(k), \quad (3.60)$$

where  $P_m(k)$  is the actual power spectrum of dark matter (differently from the previously used  $P_0(k)$ , which was a “first guess” spectrum, although from now on there will be little or no difference between these two quantities). PVP assumes linear bias, since this is a good approximation. Given the previous motivation with the connection between morphology and large-scale clustering, we can take  $b$  to have a dependence on, for instance, luminosity. For the case studied in PVP,  $b = b(L)$ , and in their study it could, in principle, be any function of the luminosity, as long as the bias of each set of galaxies can be entirely determined. We will see soon in which context this assumption may be dropped.

In a magnitude-limited survey, galaxies more luminous than the average will be relatively more numerous than others at high redshifts. At low-redshifts, conversely, low luminosity galaxies will have a larger number density. The power spectrum at large-scales, therefore, will be preferentially measured by the very luminous ones, which can potentially distort the recovered spectrum.

The usual assumption that galaxies form, in each subvolume  $dV$ , a Poisson sampling of the density field with mean  $\bar{n}$  now becomes

$$\text{Prob}(dV, L) = \bar{n}(\vec{r}, L) dV dL [1 + b(\vec{r}, L)\delta(\vec{r})], \quad (3.61)$$

where  $\text{Prob}(dV, L)$  is the probability of finding a galaxy of luminosity  $L$  in the cell  $dV$ . The luminosity dependence is acquired not only by  $b = b(\vec{r}, L)$  but also by  $\bar{n} = \bar{n}(\vec{r}, L)$ .

The weighted galaxy fluctuation field is now defined as

$$F(\vec{r}) \equiv \frac{1}{N} \int dL \frac{w(\vec{r}, L)}{b(\vec{r}, L)} [n_g(\vec{r}, L) - \alpha n_r(\vec{r}, L)], \quad (3.62)$$

where the normalization  $N$  is now given by

$$N \equiv \left[ \int d^3r \left( \int dL \bar{n}(\vec{r}, L) w(\vec{r}, L) \right)^2 \right]^{1/2}. \quad (3.63)$$

Equations (3.62) and (3.63) can be directly compared with (3.40). The integral over the luminosity represents a sum over different galaxy sets, amounting to a single overdensity field  $F$ . Such a distribution as a function of luminosity certainly wouldn't be continuous. In practice, the classification of galaxies into different species would be a discrete one. The integral, therefore, is just a formality. PVP does present, however, a discretized version of (3.62), which is a straightforward replacement of the integral sign by a sum.

Two things must be noticed about equation (3.62). The first is the natural redefinition

$$n_g(\vec{r}) = \sum_i \delta_D(\vec{r} - \vec{r}_i) \rightarrow n_g(\vec{r}, L) = \sum_i \delta_D(\vec{r} - \vec{r}_i) \delta_D(L - L_i).$$

The second is that there is a bias term in the denominator, already factored out, which guarantees a bias correction.

The same 2-point functions (3.43)-(3.45) apply to the PVP formalism, with additional Dirac deltas and bias terms, although the window function now changes accordingly:

$$G(\vec{r}) \equiv \frac{1}{N} \int dL \bar{n}(\vec{r}, L) w(\vec{r}, L). \quad (3.64)$$

The final PVP estimator is then very similar to the FKP eq. (3.52):

$$\left\langle \left| F(\vec{k}) \right|^2 \right\rangle = \int \frac{d^3 k'}{(2\pi)^3} P(\vec{k}) \left| G(\vec{k} - \vec{k}') \right|^2 + P_{\text{shot}}, \quad (3.65)$$

where

$$P_{\text{shot}} \equiv \frac{1 + \alpha}{N^2} \int d^3 r \int dL \bar{n}(\vec{r}, L) \frac{w^2(\vec{r}, L)}{b^2(\vec{r}, L)}. \quad (3.66)$$

Just like in the FKP case, the connection between PVP and the general formalism introduced in section 3.2 is straightforward.

Again, the last step is to carefully determine the weights in order to make the estimator optimal. Under exactly the same approximations employed in FKP, the relative error for PVP is given by

$$\frac{\sigma_P^2(k)}{P^2(k)} = \frac{(2\pi)^3}{V_k N^4} \int d^3 r \left\{ \left[ \int dL \bar{n}(\vec{r}, L) w(\vec{r}, L) \right]^2 + \left[ \int dL \bar{n}(\vec{r}, L) \frac{w^2(\vec{r}, L)}{b^2(\vec{r}, L)} \right] \frac{1}{P_0(k)} \right\}^2. \quad (3.67)$$

The same considerations regarding the fiducial spectrum  $P_0(k)$  in FKP apply to this case (section 3.3.1).

Following the same procedure of variation calculus employed in the previous case, we require that  $\sigma_P^2(k)$  be stationary with respect to small variations  $\delta w$ . The solution, representing the optimal PVP weights, is given by

$$w(\vec{r}, L') = \frac{b^2(\vec{r}, L') P_0(k)}{1 + \int dL \bar{n}(\vec{r}, L) b^2(\vec{r}, L) P_0(k)}. \quad (3.68)$$

Equation (3.68) has a subtlety: the weights depend not only on the bias and selection function of one type of tracer in a location  $\vec{r}$ , but also on the bias of every other species of tracer in the same location. Furthermore, even though PVP defined the overdensity field (3.62) with a factor of  $b^{-1}$ , the weights become proportional to  $b$ , which means that the clustering strength is related to the signal-to-noise.

We implemented the PVP estimator for the case of only 1 tracer, in which case it is completely analogous to the FKP implementation discussed in section 3.3.2. PVP tested their estimator with a useful set of galaxy mocks generated from lognormal realizations

that matched the selection effects of the 2dFGRS survey. We will discuss how to efficiently generate them and what are their characteristics shortly.



# Chapter 4

## The Multi-Tracer Estimator

### 4.1 Towards an Optimal Estimator

An observed galaxy field includes a combination of at least two stochastic processes. On the one hand, galaxies and even larger structures in the universe are sampling an underlying density field which is (or was, at early times) fundamentally the result of a Gaussian process in each point. On the other hand, galaxies themselves form an approximately Poisson distribution over this Gaussian field. We can relate the discussion of section 3.1) to this issue by realizing that shot noise is a byproduct of the Poisson process, while cosmic variance comes from the Gaussian realization.

However, the fact that, on the largest scales, we cannot reduce the error bars below order unity for some quantities that are proportional to the power spectrum can be circumvented. This can be achieved if one considers that multiple species trace the same field  $\delta$  [47]. On the linear bias approximation  $\delta_i = b_i\delta$  for the tracer  $i$ , the ratio between the density fields, corresponding to  $b_1/b_2$ , will not be subject to the same statistical fluctuations that affect  $\delta_i$ :

$$\frac{P_1(k)}{P_2(k)} = \frac{b_1^2 P_0(k)}{b_2^2 P_0(k)}.$$

In other words, the Gaussian fluctuations in  $P_0(k)$  are exactly canceled when taking the ratio of the tracers.

It has been found that “beating” cosmic variance this way can reduce the uncertainty in some key cosmological parameters by around an order of magnitude, specially the  $f_{NL}$  factor of non-gaussianity and the redshift-space distortion parameter  $\beta$  [48, 49].

Diminishing the effects of cosmic variance is of special importance currently, since we are aiming to build large and dense data sets in which shot noise will not dominate over signal. The *multi-tracer* approach can also be studied in terms of the Fisher matrix formalism, which yields a clear understanding of the auto and cross-covariances, which did not exist in FKP and were not fully developed in PVP [50, 51].

Suppressing cosmic variance in some cosmological observables is indeed a very good perspective for an estimator. We must not forget, however, that to efficiently exploit a data set and extract all the information available in it, we should make use of an optimal estimator. Combining the multitracer methods with the statistical tools of section 3.2 is, thus, a promising approach. We will follow along the lines of [52] to derive optimal multi-tracer estimators and test them. This reference is actually the byproduct of the present work, developed by the author of this work and collaborators.

In a general way, we can regard the power spectra of multiple galaxy species as

$$P_\alpha \equiv B_\alpha^2 P_m, \quad (4.1)$$

where  $P_m$  is the underlying power spectrum they are all supposed to follow, and  $B_\alpha = B_\alpha(\vec{k}, z, \mu_k)$  is a generalized bias. This bias can have dependences on the cosine of angle between the Fourier mode and the line of sight ( $\mu_{\vec{k}}$ ) as well as on redshift, scales, etc. In practice, we are introducing in this multiplicative factor any physical phenomenon that might distort or shift the power spectrum. Note that we will denote different tracers with greek indices –  $\alpha = 0, 1, \dots, N$  for  $N$  different sets of objects tracing the same underlying structure. In this notation, cross-spectra are defined as

$$P_{\alpha\beta} = B_\alpha B_\beta P_m. \quad (4.2)$$

To properly exploit the Fisher matrix, we rely on the Cramér-Rao bound and the formalism from section 3.2. Let us consider our measured quantities, the data, as vectors  $d_i$  that obey  $\langle d_i \rangle = 0$ . Let us also regard the Gaussian assumption commonly used in the literature: the parameters  $p_\mu$  we want to extract from the data are well described by a multivariate Gaussian likelihood. According to (3.22), we can then write

$$F_{\mu\nu} = \frac{1}{2} \text{Tr} \left[ C^{-1} \frac{\partial C}{\partial p_\mu} C^{-1} \frac{\partial C}{\partial p_\nu} \right] = \frac{1}{2} \sum_{ijkl} C_{ij}^{-1} \frac{\partial C_{jk}}{\partial p_\mu} C_{kl}^{-1} \frac{\partial C_{li}}{\partial p_\nu}, \quad (4.3)$$

where the second equality is just an explicit way to represent the trace, with a convenient notation.

The next step, in direct analogy with eqs. (3.36) and (3.37), is to define a quadratic form  $\hat{q}_\mu$  with generalized weights given by  $E_\mu^{ij}$ :

$$\hat{q}_\mu \equiv \sum_{ij} d_i E_\mu^{ij} d_j - \Delta_\mu, \quad (4.4)$$

where

$$E_\mu^{ij} \equiv \frac{1}{2} \sum_{kl} C_{ik}^{-1} \frac{\partial C_{kl}}{\partial p_\mu} C_{lj}^{-1}. \quad (4.5)$$

For convenience, in eq. (4.4) we have explicitly subtracted the bias<sup>1</sup> of the estimator,  $\Delta_\mu$  and in eq. (4.5) we have factored out a  $1/2$ .

With the quadratic form, the final goal is to build an estimator that obeys

$$\hat{p}_\mu = \sum_\alpha F_{\mu\alpha}^{-1} \hat{q}_\alpha \quad (4.6)$$

and whose covariance is the inverse of the Fisher matrix,  $Cov(\hat{p}_\mu, \hat{p}_\nu) = F_{\mu\nu}^{-1}$ . This can be achieved, with the definition (4.6), if

$$Cov(\hat{q}_\alpha, \hat{q}_\beta) = F_{\alpha\beta}. \quad (4.7)$$

We must then show that the covariance of the quadratic form  $\hat{q}_\mu$  is the Fisher matrix.

Disregarding the bias term, which will have no correlation, we can write the covariance for  $\hat{q}$  as

$$Cov(\hat{q}_\mu, \hat{q}_\nu) = \langle (d_i E_\mu^{ij} d_j) (d_k E_\nu^{kl} d_l) \rangle - \langle d_i E_\mu^{ij} d_j \rangle \langle d_k E_\nu^{kl} d_l \rangle \quad (4.8)$$

$$= E_\mu^{ij} E_\nu^{kl} (\langle d_i d_j d_k d_l \rangle - \langle d_i d_j \rangle \langle d_k d_l \rangle). \quad (4.9)$$

The Gaussian distribution can be completely described by its 2-point function, which means that N-point functions can be written as permutations of the second moment. Exploiting this approximation again, we can write the 4-point function  $\langle d_i d_j d_k d_l \rangle$  in the previous equation as

$$\langle d_i d_j d_k d_l \rangle = \langle d_i d_j \rangle \langle d_k d_l \rangle + \langle d_j d_k \rangle \langle d_l d_i \rangle + \langle d_k d_i \rangle \langle d_j d_l \rangle \quad (4.10)$$

$$= C_{ij} C_{kl} + C_{jk} C_{li} + C_{ki} C_{jl}. \quad (4.11)$$

Using this property, we finally get

$$Cov(\hat{q}_\mu, \hat{q}_\nu) = \sum_{ijkl} E_\mu^{ij} E_\nu^{kl} (C_{jk} C_{li} + C_{ki} C_{jl}) \quad (4.12)$$

$$= 2 \text{Tr} (E_\mu^{ij} C_{jk} E_\nu^{kl} C_{li}) \quad (4.13)$$

$$= \frac{1}{2} \text{Tr} \left( C^{-1} \frac{\partial C}{\partial p_\mu} C^{-1} \frac{\partial C^{-1}}{\partial p_\nu} \right) = F_{\mu\nu}, \quad (4.14)$$

where the third equality follows directly from the definition (4.5). Having confirmed that

---

<sup>1</sup>We will use the same word for the offset between the ensemble average of the estimator and its true value, – a measure of its accuracy – and for the multiplicative term  $B_\alpha$  – a quantity degenerated with the amplitude of the spectrum. Confusion is easily avoidable if one keeps in mind that  $B_\alpha$  cannot ever be “subtracted” and its value has nothing to do with an estimator being optimal or not, it just reflects the combined effect of many physical phenomena that modify  $P_m$ .

$Cov(\hat{q}_\alpha, \hat{q}_\beta) = F_{\mu\nu}$ , it follows that

$$Cov(\hat{p}_\mu, \hat{p}_\nu) = \sum_{\alpha\beta} F_{\mu\alpha}^{-1} F_{\nu\beta}^{-1} \underbrace{\langle \hat{q}_\alpha \hat{q}_\beta \rangle}_{F_{\alpha\beta}} = F_{\mu\nu}^{-1}, \quad (4.15)$$

exactly as we expect from an optimal estimator in the sense of the Cramér-Rao bound. Also, when requiring that the estimator be accurate, the bias  $\Delta_\mu$  to be subtracted can be directly found by taking  $\langle \hat{p}_\mu \rangle = \bar{p}_\mu$  in (4.6), which yields

$$\begin{aligned} \Delta_\mu &= \sum_{ij} E_\mu^{ij} C_{ij} - \sum_{\beta} F_{\alpha\beta} \bar{p}_\beta \\ &= \frac{1}{2} \sum_{ki} C_{ik}^{-1} \frac{\partial C_{ki}}{\partial p_\mu} - \sum_{\beta} F_{\alpha\beta} \bar{p}_\beta \end{aligned} \quad (4.16)$$

With these definitions, we can start building the actual multi-tracer estimator. We still lack, however, an expression for the covariance and the Fisher matrix of the multi-tracer survey. Not only will we derive these quantities, but also we shall see how they reduce to the FKP and PVP cases in the appropriate limits.

## 4.2 Fisher Matrix for Galaxy Surveys

### 4.2.1 Single Species

The limit in which the quadratic method of section 3.2 reduces to the FKP weighting scheme (which is also a way to derive the FKP Fisher matrix) was obtained in [30]. We retrace these steps below, and the result will ultimately be the optimized FKP weights, but obtained by taking a different route from the one adopted in the FKP paper.

To obtain the pixel covariance in real space, we take the inverse Fourier transform of eq. (3.6) (notice that eq. (3.6) is the pixelized second moment of our data):

$$C(\vec{x}, \vec{y}) = \int e^{-i\vec{k}\cdot(\vec{x}-\vec{y})} P(\vec{k}) \frac{d^3k}{(2\pi)^3} + \frac{\delta_D(\vec{x}-\vec{y})}{\bar{n}(\vec{x})}, \quad (4.17)$$

where  $\vec{x}$  and  $\vec{y}$  are the center of real-space bins<sup>2</sup>. The derivative of  $C(\vec{x}, \vec{y})$  with respect to the parameters  $\theta_i = P(\vec{k}_i)$  is then

$$C_{,i}(\vec{x}, \vec{y}) = \frac{\partial C(\vec{x}, \vec{y})}{\partial P(\vec{k}_i)} = \frac{1}{(2\pi)^3} e^{-i\vec{k}_i\cdot(\vec{x}-\vec{y})}. \quad (4.18)$$

---

<sup>2</sup>Just as a remark, we have been using bins in real space *and*  $k$ -space for a while now, but this interpretation will become more crucial as we move on and define a Fisher matrix for information *density*.

If  $P(\vec{k})$  varies slowly within the  $k$  bin of interest, we can approximate the covariance by<sup>3</sup>

$$C(\vec{x}, \vec{y}) \approx \left[ P + \frac{1}{\bar{n}(\vec{x})} \right] \delta_D(\vec{x} - \vec{y}), \quad (4.19)$$

which implies that the inverse covariance, since  $\int d^3x' \delta_D(\vec{x} - \vec{x}') \delta_D(\vec{x}' - \vec{y}) = \delta_D(\vec{x} - \vec{y})$ , is given by

$$C^{-1}(\vec{x}, \vec{y}) \approx \left[ P + \frac{1}{\bar{n}(\vec{x})} \right]^{-1} \delta_D(\vec{x} - \vec{y}) \quad (4.20)$$

$$= \phi_{FKP}(\vec{x}) \delta_D(\vec{x} - \vec{y}), \quad (4.21)$$

where  $\phi_{FKP}(\vec{x})$  is precisely the FKP weight with an additional normalization factor  $\bar{n}$  as defined in (3.34).

Using the expression for the Fisher matrix under the Gaussian approximation, we have (where latin indices stand for actual matrix bins, since greek indices are clearly unnecessary in this single species case)

$$F_{ij} = \frac{1}{2} \text{Tr} [C^{-1} C_{,i} C^{-1} C_{,j}] \quad (4.22)$$

$$\approx \frac{1}{2} \int \phi_{FKP}(\vec{x}) e^{-i\vec{k}_i \cdot (\vec{x} - \vec{y})} \phi_{FKP}(\vec{y}) e^{-i\vec{k}_j \cdot (\vec{y} - \vec{x})} d^3x d^3y \quad (4.23)$$

$$= \frac{1}{2} \left| \phi_{FKP}(\vec{k}_i - \vec{k}_j) \right|^2. \quad (4.24)$$

In a direct comparison with (3.35), we see that the Fisher matrix is acting as a window:

$$\langle q_i^{FKP} \rangle = \int F_{ij} P(\vec{k}_j) \frac{d^3k_j}{(2\pi)^3},$$

in which case we have to normalize it to unity, according to (3.29), by dividing it by the integral of  $F_{ij}$  over  $k$ -space

$$\int F_{ij} \frac{d^3k_j}{(2\pi)^3} = \frac{1}{2} \int \left( \frac{\bar{n}}{1 + \bar{n}P} \right)^2 e^{i(\vec{k}_i - \vec{k}_j) \cdot \vec{x}} d^3x \frac{d^3k_j}{(2\pi)^3} = \frac{V_{\text{eff}}^{ij}}{2P^2}, \quad (4.25)$$

where we have used Parseval's theorem and the definition for  $\phi_{FKP}(\vec{x})$ .

In the previous equation, we have introduced an important quantity  $V_{\text{eff}}$  defined as

$$V_{\text{eff}}(k) = \int \left( \frac{\bar{n}(\vec{x}) P(k)}{1 + \bar{n}(\vec{x}) P(k)} \right)^2 d^3x. \quad (4.26)$$

The effective volume represents the information contained in the spatial volume probed.

---

<sup>3</sup>This is also commonly known as the *stationary phase* approximation.

We can regard  $\bar{n}P = P/(1/\bar{n})$  as the signal-to-noise ratio in the approximation that the nature of the noise is completely Poissonian. When  $\bar{n}P \gg 1$ , i.e. signal dominates,  $V_{\text{eff}} \rightarrow V$ , which means we are collecting the full information contained in the volume probed. If, however,  $\bar{n}P \ll 1$ ,  $V_{\text{eff}} \ll V$  which means that we are not efficiently exploiting the survey volume. Another way to realize the importance of the effective volume is considering that, since it is directly connected to the error bars in FKP – as we shall soon show –, shrinking the error bars by a factor 2 is equivalent to probing a galaxy number density and survey volume larger by a factor of 4.

For the final connection between the Fisher matrix approach and the FKP formalism, consider equation (3.39), which shows that

$$\frac{\sigma(q_i^{FKP})}{\langle q_i^{FKP} \rangle} \approx 2 \frac{\int \phi_{FKP} d^3r}{\int \phi_{FKP}^2 d^3r} = \sqrt{2} \left( 1 + \frac{1}{\bar{n}P} \right), \quad (4.27)$$

where the last equality holds if  $\bar{n}$  is constant over a volume-limited survey. Since we are guaranteed that the covariance is the inverse Fisher matrix, and since eq. (4.27) is nothing but the square root of the covariance, we immediately find the FKP Fisher matrix:

$$F_{FKP} = \frac{1}{2} \left( \frac{\bar{n}P}{1 + \bar{n}P} \right)^2. \quad (4.28)$$

Furthermore, since averaging equation (4.27) over shells in  $k$ -space yields exactly the FKP error estimate (3.54), we have also shown that the FKP method is indeed optimal in the quadratic estimator framework.

We have just derived an important result: both the method developed in FKP (weighting the density contrast and then forcing the weighting scheme to be optimal) and the Fisher matrix formalism (defining a quadratic form whose covariance is the inverse of the Fisher matrix) reached an optimal estimator. Both routes were completely different, but reached the same result.

A few final remarks are necessary. Firstly, note that equation (4.28) encapsulates both signal-to-noise limits which we described in section 3.1. For  $\bar{n}P \ll 1$ , i.e., when shot noise is much larger than the power in a given bin,  $F_{FKP} \rightarrow 0$  like a power-law. This means that the information contained in the survey volume decreases rapidly as the galaxy field becomes sparser. The other limit is  $\bar{n}P \gg 1$ , when signal dominates and shot noise is negligible. In this case,  $F_{FKP} \rightarrow 1/2$ , which means that, no matter how finely the underlying density field is sampled, there is a maximum amount of information that can be extracted. This is nothing but a re-statement of the cosmic variance effect. We will see soon that for the multi-tracer estimator this is not true.

Secondly, since  $V_{\text{eff}}$  is actually a *spatial* integral of  $F_{FKP}$ , and since the actual FKP error bars are a *k-space* integral (over shells) of the same quantity, we should regard

$F_{FKP}$  as a *density of information over phase space*. In other words, we can rewrite the uncertainty in the power estimate of bin  $k_i$  as

$$\left(\frac{\sigma_P}{P}\right)^2 \equiv \left(\frac{\sigma(q)}{\langle q \rangle}\right)^2 \approx \frac{2}{V_{k_i} V_{\text{eff}}},$$

where  $V_{k_i} = 4\pi k_i^2 \Delta k$ , with  $\Delta k$  defined as the thickness of the  $i$ th shell. This distinction appeared naturally when comparing the Fisher matrix to the FKP prediction, and shall continue to play an important role.

Also, from our derivation, the covariance is necessarily diagonal in  $k$ . Looking at the Fisher matrix, it really could not be otherwise. We implicitly imposed this in the approximation that led to eq. (4.19). It remains true as long as we consider bins in  $k$  which are narrow enough when compared to the characteristic scale of the survey, which ultimately constrains the range of wavelengths that  $P(k)$  can cover. In other words, bins separated by much more than  $\Delta k$  will be uncorrelated. In practice, we can easily *define* the bins in  $k$  taking this constrain into account, so that our estimates retain as much (uncorrelated) information as possible. Note also that bins narrower than  $\Delta k$  will be correlated with the surrounding ones, while those much wider than  $\Delta k$  will be averaging out the intrinsic fluctuations of the spectrum. Both limits are undesirable.

Let us reconsider the definition (4.3) in the continuum limit, now being careful about the distinction between the integrated Fisher matrix and the one that represents its density over phase-space. The FKP Fisher matrix can be rewritten as

$$F(\theta_i, \theta_j) = \int \frac{d^3x d^3k}{(2\pi)^3} \frac{\partial \log P_\alpha}{\partial \theta_i} \mathcal{F}_\alpha \frac{\partial \log P_\alpha}{\partial \theta_j}, \quad (4.29)$$

where the Fisher matrix density  $\mathcal{F}_\alpha$  for species  $\alpha$  is nothing but our previous result of eq. (4.28), but including the bias  $B_\alpha$ , as defined in (4.1):

$$\mathcal{F}_\alpha(\vec{x}, \vec{k}) = \frac{1}{2} \left( \frac{\mathcal{P}_\alpha}{1 + \mathcal{P}_\alpha} \right)^2, \quad (4.30)$$

and

$$\mathcal{P}_\alpha(\vec{x}, \vec{k}) \equiv \bar{n}_\alpha(\vec{x}) P_\alpha(\vec{x}, \vec{k}) \quad (4.31)$$

is the signal-to-noise of tracer  $\alpha$ . Note also that we have chosen to express (4.29) in terms of  $\log P_\alpha$ , which implies a simple projection of the Fisher matrix with a Jacobian that is trivially given by the chain rule in this case. The expression (4.30) is the more general form of writing the information density, because it expresses the total information contained in the clustering of a species  $\alpha$  as  $\mathcal{P}_\alpha$ . To obtain the Fisher matrix for other variables, we must know how the new set of parameters relate to the old ones, so that we

can project the Fisher matrix. In other words, we need to know what is the Jacobian of the transformation from a set of variables to another,  $\Theta \rightarrow \tilde{\Theta}$ . We write the Jacobian matrices as

$$J_{ij} = \frac{\partial \theta_i}{\partial \tilde{\theta}_j},$$

which means that the projection given by

$$\mathcal{F}'_{\mu'\nu'} = J_{\mu'\mu} \mathcal{F}_{\mu\nu} J_{\nu\nu'} \quad (4.32)$$

can be written more explicitly as

$$F(\theta_i, \theta_j) \rightarrow F(\tilde{\theta}_i, \tilde{\theta}_j) = \sum_{lk} \frac{\partial \theta_l}{\partial \tilde{\theta}_i} F(\theta_l, \theta_k) \frac{\partial \theta_k}{\partial \tilde{\theta}_j}. \quad (4.33)$$

## 4.2.2 Multiple Species

We finally get to the point when we can start a studying of the covariance and Fisher matrix of multi-tracer surveys. Alternative routes for the derivation have been proposed [49, 53], and we will employ a generalization of the already known covariance for the single species case [52, 50].

The generalization of (4.3) for the multi-tracer estimator is

$$F_{\mu\nu;ij} = \sum_{\alpha\beta\gamma\sigma} \int d^3x d^3x' d^3x'' d^3x''' C_{\alpha\beta}^{-1}(\vec{x}, \vec{x}') \frac{\partial C_{\beta\gamma}(\vec{x}', \vec{x}'')}{\partial P_{\mu,i}} C_{\gamma\sigma}^{-1}(\vec{x}'', \vec{x}''') \frac{\partial C_{\sigma\alpha}(\vec{x}''', \vec{x})}{\partial P_{\nu,j}}, \quad (4.34)$$

where the pixel covariance also takes a generalized form, given by

$$\begin{aligned} C_{\alpha\beta}(\vec{x}, \vec{x}') &= \xi_{\alpha\beta}(\vec{x}, \vec{x}') + \frac{\delta_{\alpha\beta}}{\bar{n}_\alpha(\vec{x})} \delta_D(\vec{x} - \vec{x}') \\ &= \int \frac{d^3k}{(2\pi)^3} e^{i\vec{k}\cdot(\vec{x}-\vec{x}')} \left[ B_\alpha(\vec{x}, \vec{k}) P_m(\vec{x}, \vec{k}) B_\beta(\vec{x}', \vec{k}) + \frac{\delta_{\alpha\beta}}{\bar{n}_\alpha(\vec{x})} \right]. \end{aligned} \quad (4.35)$$

Under the same approximation that led to equation (4.19), we can rewrite this general covariance in a way that can be inverted analytically:

$$C_{\alpha\beta}(\vec{x}, \vec{x}') \approx \delta_D(\vec{x} - \vec{x}') \left[ \frac{\delta_{\alpha\beta}}{\bar{n}_\alpha} + B_\alpha P_m B_\beta \right]. \quad (4.36)$$

One can check that the inverse of eq. (4.36) is given by

$$C_{\alpha\beta}^{-1}(\vec{x}, \vec{x}') \approx \delta_D(\vec{x} - \vec{x}') \left[ \delta_{\alpha\beta} \bar{n}_\alpha - \bar{n}_\alpha \frac{B_\alpha P_m B_\beta}{1 + \mathcal{P}} \bar{n}_\beta \right], \quad (4.37)$$

where we define the *total clustering strength* as the sum of all signal-to-noise ratios for

each species (or the individual *clustering strengths*) as

$$\mathcal{P}(\vec{x}, \vec{k}) \equiv \sum_{\mu} \bar{n}_{\mu}(\vec{x}) B_{\mu}^2(\vec{x}, \vec{k}) P_{\mu}(\vec{x}, \vec{k}) = \sum_{\mu} \mathcal{P}_{\mu}. \quad (4.38)$$

Note that the covariance matrix and its inverse follow the property

$$\sum_{\beta} \int d^3 x' C_{\alpha\beta}^{-1}(\vec{x}, \vec{x}') C_{\beta\gamma}(\vec{x}', \vec{x}'') = \delta_{\alpha\gamma} \delta_D(\vec{x} - \vec{x}''). \quad (4.39)$$

We already know the covariance and its inverse. The last piece needed to assemble the right-hand side of eq. (4.34) is the derivative of  $C$  with respect to  $P$ . Using the fact that<sup>4</sup>

$$\frac{\partial f(\vec{x}, \vec{k})}{\partial f(\vec{x}', \vec{k}')} = (2\pi)^3 \delta_D(\vec{x} - \vec{x}') \delta_D(\vec{k} - \vec{k}'), \quad (4.40)$$

we can write  $\partial P_{\alpha}(\vec{x}, \vec{k}) / \partial P_{\mu, i} = \delta_{\vec{x}, \vec{k}}^i$ . Using this in eq. (4.35), we get

$$\frac{\partial C_{\alpha\beta}(\vec{x}, \vec{x}')}{\partial P_{\mu, i}} = \int \frac{d^3 k}{(2\pi)^3} e^{i\vec{k} \cdot (\vec{x} - \vec{x}')} \left( \delta_{\alpha\mu} \delta_{\vec{x}, \vec{k}}^i + \delta_{\beta\mu} \delta_{\vec{x}', \vec{k}}^i \right) \frac{B_{\alpha}(\vec{x}, \vec{k}) B_{\beta}(\vec{x}', \vec{k})}{2B_{\mu}^2(\vec{x}_i, \vec{k}_i)}, \quad (4.41)$$

where we have introduced the notation  $\delta_{\vec{x}, \vec{k}}^i$  to represent the limiting of the integral to a bin  $i$ . More explicitly:

$$\int \frac{d^3 x d^3 k}{(2\pi)^3} [\dots] \times \delta_{\vec{x}, \vec{k}}^i = \int_{V_i} \frac{d^3 x d^3 k}{(2\pi)^3} [\dots]. \quad (4.42)$$

Plugging eqs. (4.36), (4.37) and (4.41) into (4.34) and noticing that the Kronecker deltas are always accompanied by their restriction over phase-space volume, we get, after a lengthy but straightforward calculation, a neat expression for the Fisher matrix:

$$F_{\mu\nu; ij} = \frac{\delta_{ij}}{P_{\mu, i} P_{\nu, j}} \int_{V_i} \frac{d^3 x d^3 k}{(2\pi)^3} \mathcal{F}_{\mu\nu}, \quad (4.43)$$

where the tracer  $\mu$  corresponds to the phase-space bin  $i$  and  $\nu$  to the bin  $j$ , and the density of information in phase-space, for  $\log \mathcal{P}_{\mu}$  [50], is

$$\mathcal{F}_{\mu\nu}(\vec{x}, \vec{k}) = \frac{1}{4} \frac{\delta_{\mu\nu} \mathcal{P}_{\mu} \mathcal{P} (1 + \mathcal{P}) + \mathcal{P}_{\mu} \mathcal{P}_{\nu} (1 - \mathcal{P})}{(1 + \mathcal{P})^2}. \quad (4.44)$$

It is straightforward to check that eq. (4.44) reduces to FKP (4.28) in the limit of 1 species, in which case  $\mathcal{P}_{\mu} = \mathcal{P} = \bar{n}P$ ,  $\mu = \nu$ .

<sup>4</sup>This concept is borrowed from functional analysis.

### 4.3 The Optimal Multi-tracer Quadratic Estimator

We can finally build the multi-tracer power spectrum estimators that will eventually be numerically implemented and tested. We have developed all the necessary tools in this chapter – the covariances and the Fisher matrix – and now all we need to do is retrace the steps of section 3.2, explored in greater detail in section 4.1.

We begin by making an explicit choice of what to consider “data”: we shall use the density contrasts  $\delta_\mu$  for each species  $\mu$ . This means that we will need to subtract the mean density of the survey volume. We shall follow the approach of PVP and FKP, and do so using random maps, which we discuss in greater detail in the next section.

The quadratic form of equation (4.4), if we define the bias of the estimator as  $\delta Q_{\mu,i} \equiv \Delta_\mu$ , becomes

$$\hat{Q}_{\mu,i} = \sum_{\alpha\beta} \int d^3x d^3x' E_{\alpha\beta}^{\mu,i}(\vec{x}, \vec{x}'; \vec{x}_i, \vec{k}_i) \delta_\alpha(\vec{x}) \delta_\beta(\vec{x}') - \delta Q_{\mu,i}. \quad (4.45)$$

The generalized weighting scheme  $E_{\alpha\beta}^{\mu,i}$  is defined analogously to eq. (4.5):

$$E_{\alpha\beta}^{\mu,i} = \frac{1}{2} \sum_{\sigma\gamma} \int d^3y d^3y' C_{\alpha\sigma}^{-1}(\vec{x}, \vec{y}) \frac{\partial C_{\sigma\gamma}(\vec{y}, \vec{y}')}{\partial P_{\mu,i}} C_{\gamma\beta}^{-1}(\vec{y}', \vec{x}'). \quad (4.46)$$

Continuing our analogy with the optimal quadratic scheme, we plug the multi-tracer version of eq. (4.41), the derivative of the covariance, and (4.37), the inverse covariance, into the previous expression for  $E_{\alpha\beta}^{\mu,i}$ . Equation (4.45) then leads to

$$\hat{Q}_{\mu,i} = \frac{1}{4B_{\mu,i}^2} \sum_{\sigma\gamma} \int d^3x d^3x' \int \frac{d^3k}{(2\pi)^3} e^{i\vec{k}\cdot(\vec{x}-\vec{x}')} \left[ f_\sigma(\vec{x}, \vec{k}) \left( \delta_{\sigma\mu} \delta_{\vec{x}, \vec{k}}^i + \delta_{\gamma\mu} \delta_{\vec{x}', \vec{k}}^i \right) f_\gamma(\vec{x}', \vec{k}) \right] - \delta Q_{\mu,i}, \quad (4.47)$$

where

$$f_\sigma(\vec{x}, \vec{k}) = \sum_\alpha w_{\sigma\alpha}(\vec{x}, \vec{k}) \delta_\alpha(\vec{x}) \quad (4.48)$$

is our weighted data, with the multi-tracer weights given by

$$w_{\sigma\alpha}(\vec{x}, \vec{k}) = \left[ \delta_{\sigma\alpha} - \frac{\mathcal{P}_\sigma(\vec{x}, \vec{k})}{1 + \mathcal{P}(\vec{x}, \vec{k})} \right] \bar{n}_\alpha B_\alpha(\vec{x}, \vec{k}), \quad (4.49)$$

which properly reduce to the FKP case if only one tracer is considered.

Limiting the integrals to specific bins using the  $\delta_{\vec{x}, \vec{k}}^i$  and applying the Kronecker deltas,

we can rearrange eq. (4.47) and write

$$\begin{aligned}\hat{Q}_{\mu,i} &= \frac{1}{4B_{\mu,i}^2} \int_{V_{\vec{k}_i}} \frac{d^3k}{(2\pi)^3} \int d^3x e^{i\vec{k}_i \cdot \vec{x}} f_{\mu}(\vec{x}) \int d^3x' e^{-i\vec{k}_i \cdot \vec{x}'} f(\vec{x}') + \text{c.c.} - \delta Q_{\mu,i} \\ &= \frac{1}{4B_{\mu,i}^2} \int_{V_{\vec{k}_i}} \frac{d^3k}{(2\pi)^3} \left[ \tilde{f}_{\mu}(\vec{k}) \tilde{f}^*(\vec{k}) + \text{c.c.} \right] - \delta Q_{\mu,i}\end{aligned}\quad (4.50)$$

where we define the total density field as

$$f = \sum_{\sigma} f_{\sigma} = \frac{1}{1 + \mathcal{P}} \sum_{\sigma} \bar{n}_{\sigma} B_{\sigma} \delta_{\sigma}.\quad (4.51)$$

Equation (4.50) is, up to a volume factor, the average over shells in  $k$ -space. We can then rewrite it as

$$\hat{Q}_{\mu,i} = \frac{V_{\vec{k}_i}}{4B_{\mu,i}^2} \left\langle \tilde{f}_{\mu} \tilde{f}^* + \text{c.c.} \right\rangle_{\vec{k}_i} - \delta Q_{\mu,i}.\quad (4.52)$$

The last step is to make the correspondence between the quadratic form  $\hat{Q}$  and the final estimator, as eq. (4.6) suggests. Following this procedure, we have

$$\hat{P}_{\mu,i} = \sum_{\nu j} [F_{\mu\nu;ij}]^{-1} \hat{Q}_{\nu,j}\quad (4.53)$$

$$= \sum_{\nu} [F_{\mu\nu;ii}]^{-1} \hat{Q}_{\nu,i}.\quad (4.54)$$

We now need an expression for the inverse of the Fisher matrix. This is rather simple, since we already know expression (4.43). The result is that

$$F_{\mu\nu}^{-1} = P_{\mu} P_{\nu} \mathcal{F}_{\mu\nu}^{-1},\quad (4.55)$$

where the inverse of the information density is

$$\mathcal{F}_{\mu\nu}^{-1} = \frac{4(1 + \mathcal{P})}{\mathcal{P}} \left( \frac{\delta_{\mu\nu}}{\mathcal{P}_{\mu}} + \frac{\mathcal{P} - 1}{2\mathcal{P}} \right).\quad (4.56)$$

We can see that eqs. (4.53) and (4.52) include both fiducial, theoretical values (or at least initial guesses) –  $B_{\mu}$  and  $P_m$  – as well as the data itself. This mix was already present in FKP and PVP, and it reflects the fact that the estimators do contain our theoretical prejudices in the form of an initial guess, which is reflected on the weights. One can check that the covariance of (4.52) is the Fisher matrix, which means that the covariance of (4.53) is the inverse Fisher matrix, saturating the Cramer-Rao bound. The estimator was indeed constructed to follow this property.

We are certainly guaranteed to have found a minimum covariance estimator, but we

still need to make sure it is unbiased by fixing the correct  $\delta Q_{\mu,i}$ . Equation (4.16) gives us the definition, we just have to plug the already known relations for the covariance and the Fisher matrix in it. We can then rewrite

$$\delta Q_{\mu,i} = \frac{1}{2} \sum_{\alpha\beta} \int d^3x d^3x' \frac{\partial C_{\alpha\beta}(\vec{x}, \vec{x}')}{\partial P_{\mu,i}} C_{\beta\alpha}^{-1}(\vec{x}', \vec{x}) - \sum_{\nu;j} F_{\mu\nu;ij} P_{\nu,j} \quad (4.57)$$

as

$$\delta Q_{\mu,i} = \frac{1}{2B_{\mu,i}^2} \int_{V_i} \frac{d^3x d^3k}{(2\pi)^3} \frac{\bar{n}_\mu B_\mu^2}{1 + \mathcal{P}} - \sum_{\nu} F_{\mu\nu;ii} P_{\nu,i}. \quad (4.58)$$

We will refine this expression for the bias of the estimator soon, when we consider the random maps in greater detail. For now, we can have an insight what it means by taking the 1-tracer (FKP) limit of eq. (4.58), in which case it becomes

$$\delta P_{\mu,i} \equiv \sum_{\nu} [F_{\mu\nu;ii}]^{-1} \delta Q_{\nu,i} \rightarrow \left\langle \frac{P_\mu}{\mathcal{P}} \right\rangle = \frac{1}{\bar{n}}, \quad (4.59)$$

which is exactly the Poisson shot noise, as we should expect.

One major difference between our estimators and FKP or PVP is that we are explicitly including information about the cross-correlations, as can be seen in eqs. (4.47)-(4.50). Even though what we want to measure are the individual spectra  $P_\mu$ , there is a gain in information when there are 2 galaxy fields tracing the same underlying dark matter field.

We can realize that this should make a difference when measuring the spectra of 2 tracers, one with high signal-to-noise and other much sparser. While FKP estimates the spectrum of the sparser sample as if it were the only information available at all, the multi-tracer estimator will take into account that another, more numerous tracer is already sampling the same field, therefore the covariance of the sparser sample can be improved. We actually tested this case and confirmed this scenario. We might wonder if the noise in one sparse tracer might propagate to the others, rendering the covariance of the densely distributed tracers larger than it would if they were analyzed alone. We are protected against this problem: because the noise only appears in the diagonal of the covariance matrix – eqs. (4.55) and (4.56) –, if one tracer has low signal, the noise will only affect that same tracer. This can also be seen in the same test that shows a reduction of covariance for the sparser tracer.

For high signal-to-noises ratios, we can see that eqs. (4.55) and (4.56) will yield a positive cross-covariance. This means that, as expected, both tracers will clump in the same dense regions and become sparser in the same voids. In other words, a high number density for both tracers will yield distributions for them that are positively (and tightly) correlated with each other. If, however, both tracers have very poor signal, the cross-covariances become negative. We can interpret this by considering that tracing the

overdensities is unlikely enough so that, if a galaxy of type  $\alpha$  is at a given position, we won't expect the species  $\beta$  to be found along with it.

Now, recall that estimating the power spectrum from a discrete, potentially irregular galaxy field will yield a convolution of the underlying spectrum with a window function. The ensemble average of the quadratic form (4.47), taking that into account and neglecting the bias of the estimator, is given by

$$\langle \hat{Q}_{\mu,i} \rangle = \frac{1}{4B_{\mu,i}^2} \int_{V_i} \frac{d^3x d^3k}{(2\pi)^3} \int \frac{d^3x' d^3k'}{(2\pi)^3} P_m(\vec{k}') \times e^{i(\vec{k}-\vec{k}')\cdot\vec{x}} G_\mu(\vec{x}, \vec{k}) e^{-i(\vec{k}-\vec{k}')\cdot\vec{x}'} G(\vec{x}', \vec{k}) + \text{c.c.} \quad (4.60)$$

where

$$G_\mu(\vec{x}, \vec{k}) = \frac{1}{P_m} \frac{\mathcal{P}_\mu}{1 + \mathcal{P}} \quad (4.61)$$

and also  $G = \sum_\mu G_\mu$ . We then define the window function  $W_{\mu,i}$  for the quadratic form as

$$W_{\mu,i} = \frac{1}{4B_{\mu,i}^2} \int_{V_i} \frac{d^3k}{(2\pi)^3} \left[ \tilde{G}_\mu \tilde{G}^* + \text{c.c.} \right]. \quad (4.62)$$

Finally, completing the connection with the previous estimators, we rewrite the quadratic form as a convolution given by

$$\langle \hat{Q}_{\mu,i} \rangle \rightarrow \int \frac{d^3k'}{(2\pi)^3} P_m(\vec{k}) W_{\mu,i}(\vec{k}_i, \vec{k}'). \quad (4.63)$$

For a general case, the final spectrum should obviously be deconvolved for a complete analysis. In the tests carried out in the next chapter, however, in order to make a clean analysis of the covariance of our estimators, we explicitly choose a trivially defined window, as will become clear.

Another important remark is that, fixing the effective bias  $B_\mu$ , our results reduce to the PVP estimator. This can be seen by taking eq. (4.43) and setting  $\partial P_{\mu,k} / \partial P_{m,i} = B_{\mu,i}^2 \delta_{k,i}$  in the projection

$$\begin{aligned} F(P_{m,i}, P_{m,j}) &= \sum_{\mu\nu} \sum_{kl} \frac{\partial P_{\mu,k}}{\partial P_{m,i}} F_{\mu\nu;ij} \frac{\partial P_{\mu,l}}{\partial P_{m,j}} \\ &= \frac{\delta_{ij}}{P_{m,i}^2} \sum_{\mu\nu} \mathcal{F}_{\mu\nu,ii} \\ &= \frac{\delta_{ij}}{P_{m,i}^2} \int_{V_i} \frac{d^3x d^3k}{(2\pi)^3} \frac{1}{2} \left( \frac{\mathcal{P}}{1 + \mathcal{P}} \right)^2, \end{aligned}$$

so that the optimal estimator becomes

$$\hat{P}_{m,i}^{PVP} = \frac{1}{N_i} \left\langle \left| \tilde{f} \right|^2 \right\rangle_{\vec{k}_i}, \quad (4.64)$$

where the normalization  $N_i$  is given by

$$N_i = \frac{1}{V_{k_i} P_{m,i}^2} \int_{V_i} \frac{d^3x d^3k}{(2\pi)^3} \left( \frac{\mathcal{P}}{1 + \mathcal{P}} \right)^2. \quad (4.65)$$

Since  $\mathcal{P}$  is really the sum of all clustering strengths and  $f$  is the sum of all the overdensities, we do recover the PVP estimator from section 3.4.

We now turn our attention to the random catalogs broadly used in Fourier analysis across the literature – and even in correlation function estimation. The objective of the synthetic catalog is to properly subtract the mean density of the survey, and reduce the noise caused by local variations in the Poisson sampling.

For each species with number density  $\bar{n}_\mu$ , we define a random map with number density  $\bar{n}_\mu^r \alpha_\mu = \bar{n}_\mu$ . The selection function of the random catalog must have, therefore, exactly the same angular and radial dependences of the actual selection function of the survey, but without structure. This means that they are created as a Poisson sample of the original selection function. In order to reduce the fluctuations in the random map, however, we might want to generate it with a much larger number of objects, which means that the constant  $\alpha_\mu$  will be much smaller than 1, matching the number densities of the real and the synthetic catalog.

Subtracting the average density introduces a problem known as *integral constraint*: to find the density contrast, we would like to subtract the average density over the whole universe, but instead we subtract the average density within the survey volume itself. Suppose that the survey area under study is some relatively large overdensity, with a mean density significantly larger than the universe's. This means that the average density of the survey will be an overestimate of the true  $\bar{\rho}$ . Even though this offset generates noise in correlation function estimates, the power spectrum estimation is mostly unaffected by complete ignorance on the average density  $\bar{n}$ , since it will only affect the mode  $k = 0$  [36]. Subtracting the average of the survey itself, however, forces  $\hat{P}(k = 0) = 0$ . It has been shown that this might also lead to some systematic damping in the largest scales [54], but otherwise leaves the shape of  $\hat{P}(k \gg 0)$  unaltered. We circumvent this problem ignoring the very largest scales of our estimates (corresponding up to the first or second bins at most).

The power spectrum estimation, therefore, is robust against problems of this nature. A quote from [30] summarizes the problem: “*we have a systematic problem with a certain mode and can immunize our results from this problem by making them independent of this*”

*untrustworthy mode*". This is true once the condition (3.30) is fulfilled and we require that the window is such that  $W_i(0) = 0$ , so that the band expectation value is truly unbiased, as can be seen on eq. (3.25). Power spectrum estimation, therefore, is robust against the integral constraint. We have also checked that our estimator has the same property.

We can make these statements more quantitative defining the density contrasts as

$$\begin{aligned} f_\mu(\vec{x}, \vec{k}) &= \sum_\nu w_{\mu\nu}(\vec{x}, \vec{k}) \frac{n_\nu(\vec{x}) - A_\nu n_\nu^r(\vec{x})}{\bar{n}_\nu} \\ &= \sum_\nu w_{\mu\nu}(\vec{x}, \vec{k}) \left[ \delta_\nu(\vec{x}) - \frac{A_\nu}{\alpha_\nu} \delta_\nu^r(\vec{x}) + 1 - \frac{A_\nu}{\alpha_\nu} \right], \end{aligned} \quad (4.66)$$

where eq. (4.66) is what we actually use in the implementation, and  $A_\nu$  is such that  $\int d^3x f_\mu = 0$ , given by

$$A_\mu = \sum_\nu R_{\mu\nu}^{-1} D_\nu, \quad (4.67)$$

where

$$D_\nu = \int d^3x \sum_\sigma w_{\nu\sigma}(\vec{x}, \vec{k}) [1 + \delta_\sigma(\vec{x})] \quad , \quad R_{\mu\nu} = \frac{1}{\alpha_\nu} \int d^3x w_{\mu\nu}(\vec{x}, \vec{k}) [1 + \delta_\nu^r(\vec{x})].$$

Taking the expectation value of eq. (4.58) with these additional terms and setting  $\alpha_\mu = \alpha$ , which is what we actually do in the implementation for simplicity, without losing accuracy, we find an expression for the bias of the estimator which actually takes into account any level of shot noise generated by the random catalog itself:

$$\delta Q_{\mu,i} = \frac{1 + \alpha}{2B_{\mu,i}^2} \int_{V_i} \frac{d^3x d^3k}{(2\pi)^3} \frac{\bar{n}_\mu B_\mu^2}{(1 + \mathcal{P})^2}. \quad (4.68)$$

In practice this won't matter too much, since we make  $A_\nu \approx \alpha \ll 1$  in order to have a random catalog much denser than the original one.

## 4.4 Testing the Multi-tracer Estimators

In the last section, we derived the main properties and subtleties of the multi-tracer estimators. We did it by following the general formalism described in sections 3.2 and 4.1, and constructed the Fisher matrix and covariances that would guarantee an optimal behavior of the estimator.

In the present section, we apply the formalism to simple mock catalogs and test our predictions. The implementation is straightforward and should be quite easy for those already familiar with the FKP/PVP recipes. We test the method in real space, but the

generalization to redshift-space is also quite simple.

For the generation of the galaxy maps we chose a simple method that is both efficient and computationally cheap enough that hundreds of realizations of a single fiducial matter power spectrum can be analyzed. We implemented the multi-tracer estimators in a cubic grid with constant, uniform mean number density (or selection function), for the case of two different species of tracers, with biases  $b_1 = 1.0$  and  $b_2 = 1.2$ . It can be shown that the convolution with the corresponding window function for this geometry is the identity, so we don't have to worry about deconvolving the spectrum before analyzing it. Nevertheless, we have checked that the estimators are as robust as the FKP or PVP methods against variations in the survey geometry.

In order to test the performance of the estimators in situations of high and low signal-to-noise, we consider three different cases, as shown in table 4.1. In each case we generate 1000 galaxy maps (each consisting of two catalogs, one for each tracer), and estimate the spectra using the methods described in this chapter.

Case	$\bar{n}_1(h^3 \text{ Mpc}^{-3})$	$b_1$	$\bar{n}_2(h^3 \text{ Mpc}^{-3})$	$b_2$
A	$10^{-2}$	1.0	$10^{-2}$	1.2
B	$10^{-2}$	1.0	$10^{-5}$	1.2
C	$10^{-5}$	1.0	$10^{-5}$	1.2

**Table 4.1:** Three cases which we use to test the robustness of the multi-tracer estimators. In all cases, the two tracers used have biases  $b_1 = 1.0$  and  $b_2 = 1.2$ . Case A has overall high signal-to-noise, while case C is the opposite. Case B is a mixed case in which tracer 1 is very dense and tracer 2 is sparse over the grid.

A physical analogy could be drawn for each of these cases. Case A could represent a low-redshift survey with high completeness, so both tracers are densely distributed over the whole volume. Case B represents, on the one hand, a highly biased, luminous tracer, reaching out to long radial distances and therefore becoming very sparsely distributed – quasars, for instance. On the other hand, it also contains a denser tracer – LRGs, for instance. Case C represents a high redshift survey, with both species being sparse.

#### 4.4.1 Lognormal Maps and the Algorithm

Our mocks follow the same procedure used in, e.g., PVP. A detailed description of the generation of lognormal maps can be found in [55]. The basic idea is that a Gaussian density contrast  $\delta^{(G)}(\vec{x})$  is not bounded from below, which implies that negative values for the density are possible in any finite-volume realization of such a Gaussian field. Lognormal fields, on the other hand, are positive-definite, so we map the Gaussian field into a lognormal field. By this argument alone, however, any distribution which respects

the bound  $\delta_{\text{distr.}} \geq -1$  could be a candidate. The choice for a lognormal distribution is more subtle: the current density field actually approximately resembles a lognormal one. This can be understood qualitatively if we regard the fact that gravitational instabilities to grow close to exponentially in the nonlinear regimes. Gravity thus acts as a map from the primordial, Gaussian density field, to another one which underwent exponential clustering. This turns out to be the very definition of a lognormal distribution: one whose logarithm is a normal distribution.

Going back to our starting point, a lognormal field obeys the condition  $\delta^{(L)}(\vec{x}) \geq -1$ . We can obtain a lognormal density field in terms of a Gaussian density field through the definition

$$1 + \delta^{(L)}(\vec{x}) = \exp[\delta^{(G)}(\vec{x}) - \sigma_G^2/2], \quad (4.69)$$

where  $\sigma_G^2$  is the variance of the Gaussian field inside a cell. Furthermore, one can check that

$$\langle \delta^{(L)} \rangle = 0, \quad \langle \delta^{(L)} \delta^{(L)} \rangle = e^{\sigma_G^2} - 1 \approx \sigma_G^2 + \mathcal{O}(\sigma_G^4), \quad (4.70)$$

where we used the property  $\langle \exp[\delta_g] \rangle = \exp[\langle \delta_g \rangle + \sigma_G^2/2] = \exp[\sigma_G^2/2]$ , for the exponential of a Gaussian random variable. What eq. (4.70) is telling us is that the lognormal density field not only fulfills the requirement that it be greater than -1, but also that it has the correct first moment and approximately reproduces the correct second moment.

Now, the Gaussian correlation function is related to the physical (assumed lognormal) correlation function by

$$\xi^{(G)}(x) = \ln[1 + \xi^{(ph)}(x)]. \quad (4.71)$$

We may want, however, a biased version of this correlation function, in order to describe different types of tracers. In this case, we can define the Gaussian correlation function as

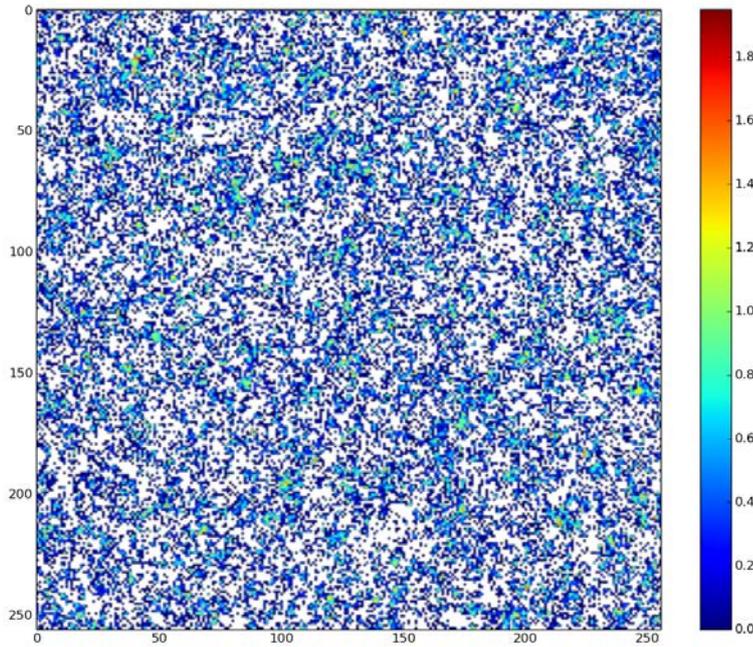
$$b^2 \xi^{(G)}(x) = \ln [1 + b^2 \xi^{(ph)}(x)]. \quad (4.72)$$

Given a fiducial cosmology, we obtain the  $z = 0$  matter power spectrum  $P_0(k)$  from the Boltzmann code `CAMB`<sup>5</sup> [56], and inverse-Fourier transform it to get the physical correlation function  $\xi^{(ph)}(x)$ . We then convert the physical (assumed lognormal) correlation function to the correlation function of the corresponding Gaussian field, and Fourier-transform that correlation function into a power spectrum for the Gaussian field. This is the power spectrum which is employed to generate the Gaussian random modes for the density contrast.

The next step is the generation of biased lognormal maps for each galaxy type. We

---

<sup>5</sup><http://CAMB.info>



**Figure 4.1:** A slice of a lognormal map for the denser tracer ( $\bar{n} = 10^{-2} h^3 \text{Mpc}^{-3}$ ). White pixels in this  $256 \times 256$  grid correspond to empty positions of the grid, and the color bar is in logarithmic scale. The biased lognormal maps have a second moment that corresponds to an input power spectrum, but with small deviations (see text). Since we will always compare the spectra we estimate from these maps with the slightly deviated version of the input, our analysis is not affected.

define the lognormal maps as

$$1 + \delta_{\mu}^{(L)}(\vec{x}) = \exp[b_{\mu} \delta^{(G)}(\vec{x}) - b_{\mu}^2 \sigma_G^2 / 2]. \quad (4.73)$$

One issue must be made clear at this point. According to (4.70), this whole procedure, when using equations (4.69) and (4.71), is guaranteed to give us approximately the same second moment of the underlying distribution of the input. In other words, for one unbiased tracer, the lognormal mocks are quite accurate.

When considering different tracers, however, we must use different values of  $b_{\mu}$  according to eq. (4.73). Notice, however that the Gaussian field that was generated as shown in eq. (4.72) with an already fixed bias. The moments of the lognormal correlation function will have higher order terms in the biased Gaussian correlation function:

$$\begin{aligned} \langle (\delta_{\mu}^{(L)})^2 \rangle &= e^{\xi_{\mu}^{(G)}} - 1 \approx b_{\mu}^2 \xi^{(G)} + \mathcal{O}\{b_{\mu}^n \xi^{(G)m}\} \\ \langle (\delta^{(L)})^2 \rangle &= e^{\xi^{(G)}} - 1 \approx b^2 \xi^{(G)} + \mathcal{O}\{b^n \xi^{(G)m}\}. \end{aligned}$$

As it can be seen, the higher order terms will not be linear relations between each correlation function. This prescription, then, is only entirely consistent for one tracer, with one bias. For a multi-tracer case, the result of is a slight distortion in the obtained spectra, for the moments of the biased lognormal field will be slightly different from the ones in the Gaussian field. In other words, the second moment in equation (4.70) will not follow exactly that relation when we consider  $\langle (\delta_\mu^{(L)})^2 \rangle$ .

This means that the actual power spectrum contained in our maps is not the exact input, but rather a slightly distorted one, as if it had a scale-dependent bias  $b = b(k)$ . We can estimate this spectrum directly from the Gaussian field, before introducing galaxies through a Poisson process, and simply take this to be the input power spectrum. This issue could have greater importance if we were comparing our results against the initial CAMB spectrum, but we are circumventing this problem. For all our applications, the main idea is that we are assuredly recovering the correct power spectrum contained in the map.

Going back to the map generation, we finally create the galaxy maps as independent Poisson realizations over the lognormal fields. Each tracer has its own spatial number density (or selection function)  $\bar{n}_\mu(\vec{x})$  and bias  $b_\mu$ , so that the maps for each tracer is given by integer numbers in each cell of volume  $dV$  through a Poisson sampling,  $N_\mu(\vec{x}) = \mathbb{P}\{\bar{n}_\mu(\vec{x})[1 + \delta_\mu^{(L)}(\vec{x})]dV\}$ , where  $\mathbb{P}\{\lambda\}$  is a Poisson distribution with mean  $\lambda$ . An example of one of the generated lognormal maps is shown in figure 4.1, which represents a dense tracer.

In the three cases explained above we considered cubic  $256^3$  grids with a fiducial cosmology characterized by a flat  $\Lambda$ CDM model with  $\Omega_b h^2 = 0.0226$ ,  $\Omega_{CDM} h^2 = 0.112$  and  $h = 0.72$ . Each cube has a physical (comoving) volume of  $1280 h^{-3}$  Mpc<sup>3</sup>. It is important to note that lognormal maps created this way do not show the usual effect of suppression in power at small scales when a smoothing algorithm is applied to convert from a continuous distribution to a discrete grid, such as Nearest Grid Point (NGP). In any case, the formalism is general enough to accommodate this necessity. Even though any discretization scheme could be used, the square grid is required in order to employ an implementation in terms of a Fast Fourier Transform (FFT), which is the only practical way to perform a Fourier analysis of large data sets.

With the galaxy maps  $n_\mu(\vec{x})$  as input, along with an initial guess for the biases  $b_\mu$ , we can start to deploy the machinery developed in the previous chapters. A previous step, in case we had not explicitly generated maps with constant, uniform number densities, would be to appropriately estimate  $\bar{n}_\mu(\vec{x})$ <sup>6</sup>. We start by building up random maps,  $n_\mu^r(\vec{x})$ , for each tracer as a Poisson process in each cell of the grid, with the same shape for the mean number density as the data, but with a larger number of objects,  $\bar{n}_\mu^r = \bar{n}_\mu/\alpha_\mu$ , with  $\alpha_\mu$  small constants. We then construct the density contrasts  $\delta_\mu(\vec{x}) = \delta_\mu^d(\vec{x}) - A_\mu \delta_\mu^r(\vec{x})$ .

---

<sup>6</sup>Estimating the selection function is a complicated task by itself, outside the scope of this work.

With an initial guess for the biases and for the amplitude of the power spectrum, we can construct  $\mathcal{P}_\mu$  and  $\mathcal{P} = \sum_\mu \mathcal{P}_\mu$ , plug them into the weights (4.49), and calculate the weighted density contrasts of eq. (4.66). We then perform an FFT over  $f(\vec{x})$  and  $f_\mu(\vec{x})$  to obtain the integrand of eq. (4.50). Taking proper care of the volume factors (in real and in Fourier space), this step should be analogous to the average over modes in equation (2.4.5) of FKP [36].

We then subtract the bias of the estimators using eq. (4.68). Since we are taking the effective biases  $B_\mu$  to be constants  $b_1$  and  $b_2$  (see table 4.1), we can rewrite (4.68) as

$$\delta Q_{\mu,i} = \frac{1 + \alpha}{2} \int_{V_i} \frac{d^3x d^3k}{(2\pi)^3} \frac{\bar{n}_\mu}{(1 + \mathcal{P})^2}, \quad (4.74)$$

which is what we actually use. Our choice for the matching between random and real catalogs is  $\alpha = 10^{-6}$ . The final spectra are then obtained with eq. (4.53).

Our estimates were evaluated in  $\sim 200$  evenly spaced bins with  $\Delta k = 0.005 h \text{ Mpc}^{-1}$ , from  $k = 0.005 h \text{ Mpc}^{-1}$  up to  $k \approx 1.0 h \text{ Mpc}^{-1}$ . For our analysis, however, we consider only the band estimates up to  $k = 0.2 h \text{ Mpc}^{-1}$ , slightly into the nonlinear regime but still below the Nyquist frequency, so that our results are not affected by discretization effects. We employed the same common simplification used in FKP and PVP: fixing  $P_m(k)$  as a constant for the weights of all bandpowers. In our case, we used  $P_m(k) = P_m = 10^4 h^{-3} \text{ Mpc}^3$ , and found no significant changes in our estimates for values different by a factor 2 or 1/2.

An instructive summary of the algorithm could be represented as follows:

1. Define input parameters:  $b_1, b_2, \bar{n}_1, \bar{n}_2$ , physical size of the grid, number of cells, number of realizations, input (theoretical)  $P_0(k)$ ;
2. Generate Gaussian field;
3. Generate two lognormal realizations, one for each tracer – see detailed description above;
4. Generate catalog by using the Poisson distribution for each lognormal realization;
5. Estimate the power spectrum for each tracer;
6. Store spectra for subsequent covariance analysis;
7. Repeat from step 2.

Notice that, in this scheme, it is crucial that both catalogs follow the same underlying Gaussian field. The galaxy catalog (resulting from the Poisson distribution) of, say, tracer 1 of a given iteration, cannot be used along with the catalog for the tracer 2 created in

another iteration. The pair of catalogs that emerged from a given Gaussian realization must necessarily be analyzed together, otherwise they will not follow the same underlying structure and the whole argument of the present work is flawed.

Step 5 is where the whole multi-tracer formalism comes into action. For all our tests, with the purpose of making a direct comparison between the different methods, we performed both the multi-tracer estimation and the usual FKP routine, exemplified in section 3.3.2. We find that the most time-consuming step is the average over shells in  $k$ -space, both for FKP and the multi-tracer. This is due to the fact that, in the present implementation, a vectorized operation could not be carried out with the  $\sim 200$  matrices of  $256^3$  cells. The alternative was to loop over all elements <sup>7</sup>, and the CPU time for 1000 realizations was of  $\sim 24$ h with the available hardware<sup>8</sup>.

## 4.4.2 Results

In this section, we present the results obtained with the estimators developed in the present work, as well as a direct comparison between their auto-covariances and covariance matrices. Our results are also reported in [52].

We first present, in figure 4.2, the estimated spectra for all the three cases. The error bars, as expected, are enlarged in the cases with low signal-to-noise. Notice that, as explained in section 4.3, even though tracer 2 in case B is much sparser than tracer 1, its noise does not propagate to tracer 1, whose error bars are the same size as they are in case A. Figure 4.3 is an even better indicative of that. The error bars represented in figure 4.2 are the theoretical ones, obtained using the fiducial power spectrum. These are, then, the error bars for the fiducial cosmology. As explained by the end of section 3.3.1, we could have also found the self-consistent error bars by replacing the fiducial  $P_0(k)$  by the estimated  $\hat{P}(k)$  for a single realization. We performed this test and found a slightly noisier pattern for the error bars (i.e. they did not decrease monotonically depicted in fig. 4.2), but no striking change in their magnitude.

The dashed line is the input spectrum  $P_0(k)$  from **CAMB**, but corrected for the deviations that arise due to the inconsistency of the lognormal “recipe” (see section 4.4.1). These spectra are, nevertheless, the actual second moment of the galaxy distribution contained in the lognormal maps. Since red circles and blue squares are the ensemble average of the estimated spectra over 1000 realizations, we clearly see that the multi-tracer estimators successfully recover the expected galaxy power spectrum. We checked that the deviations between the average band power estimate and the theoretical spectrum are, at most,

<sup>7</sup>Both estimators were fully implemented in **Python**. Even though direct iterations are notoriously slow in this language and we were dealing with excessively large matrices, there are smarter ways to perform this simple computation.

<sup>8</sup>A desktop computer with 8Gb RAM memory separated in 4 cores. The code was not paralelized to make use of the whole available computing power.

around  $\sim 2\%$ .

With 1000 realizations and storing each pair of spectra obtained using the FKP method and ours, we can perform a direct comparison between both. Furthermore, we can test the theoretical predictions for the covariance of the multi-tracer against the actual, empirically obtained covariance. We expect the theoretical error bars to be a good estimate of the empirical ones. Our prediction for the covariance and error bars, are, revising eq. (4.55), given by

$$Cov(P_{\mu,i}, P_{\nu,j}) = \delta_{ij} \left[ \frac{1}{P_{\mu,i} P_{\nu,j}} \int_{V_i} \frac{d^3x d^3k}{(2\pi)^3} \mathcal{F}_{\mu\nu} \right]^{-1}, \quad (4.75)$$

where  $\mathcal{F}_{\mu\nu}$  is given by eq. (4.44).

In figure 4.3 we present a comparison between the theoretical and empirical error bars of FKP, and the theoretical and empirical errors of the multi-tracer estimators. For FKP and the multi-tracer, the theoretical error bars were obtained, respectively, with eqs. (3.54) and (4.75). The empirical error bars are the standard deviation of each bin estimate, for each method separately.

We find that the theoretical expectations and the empirical variances generally agree very well. In some of them, the theoretical errors slightly underestimate the empirical values, which is consistent with the notion that the Fisher matrix is an underestimate of the true covariance, or, in other words, the ideal covariance. This is also found in the FKP implementation. In cases B and C, the multi-tracer method performs significantly better than FKP, both regarding the theoretical and empirical variances. We pay special attention to case B, in which the difference between FKP and the multi-tracer is of almost an order of magnitude around  $k = 0.1 h \text{ Mpc}^{-1}$ . We attribute the better performance of multi-tracer to the fact that it contains information regarding the cross-spectra, which is simply not taken into account by FKP, since it assumes single galaxy sets are sampling the underlying field.

In figure 4.4 we compare the theoretical and empirical variances for the cross-spectra of the two-tracers and for the ratio of the spectra,  $P_1/P_2$ . Since the FKP method cannot make a prediction in this case, we show only the multi-tracer results. To perform this test, we follow the technique exploited by [48], reproduced below.

We can diagonalize the multi-tracer Fisher matrix (4.44) to have a better insight on the cross-correlations. The set of variables which diagonalize the Fisher matrix are defined

as

$$\begin{aligned}\mathcal{Y}_1 &= \mathcal{S}_1 = \mathcal{P} \\ \mathcal{Y}_2 &= \frac{\mathcal{P}_1}{\mathcal{S}_1} \\ &\vdots \\ \mathcal{Y}_N &= \frac{\mathcal{P}_{N-1}}{\mathcal{S}_N},\end{aligned}$$

where  $N$  is the number of tracers and  $\mathcal{S}$  are *effective* powers given by

$$\mathcal{S}_a = \sum_{\alpha=a}^N \mathcal{P}_\alpha. \quad (4.76)$$

We can then regard  $\mathcal{Y}$  as *relative* effective powers. For 2 types of tracers, then, the variables which diagonalize the  $2 \times 2$  Fisher matrix are  $\mathcal{P} = \mathcal{P}_1 + \mathcal{P}_2$  and  $\mathcal{P}_1/\mathcal{P}_2$ . The corresponding Fisher matrix per unit of phase space for  $\log(\mathcal{P}_1/\mathcal{P}_2)$  is then

$$F_{\text{ratio}} = \frac{\mathcal{P}_1 \mathcal{P}_2}{4(1 + \mathcal{P}_1 + \mathcal{P}_2)}, \quad (4.77)$$

from which follows that the relative covariance of that ratio is, for a bin  $i$ ,

$$\left[ \int_{V_i} F_{\text{ratio}} d^3x d^3k / (2\pi)^3 \right]^{-1}.$$

Figure 4.4, then, evinces the power of the multi-tracer approach in measuring the quantity

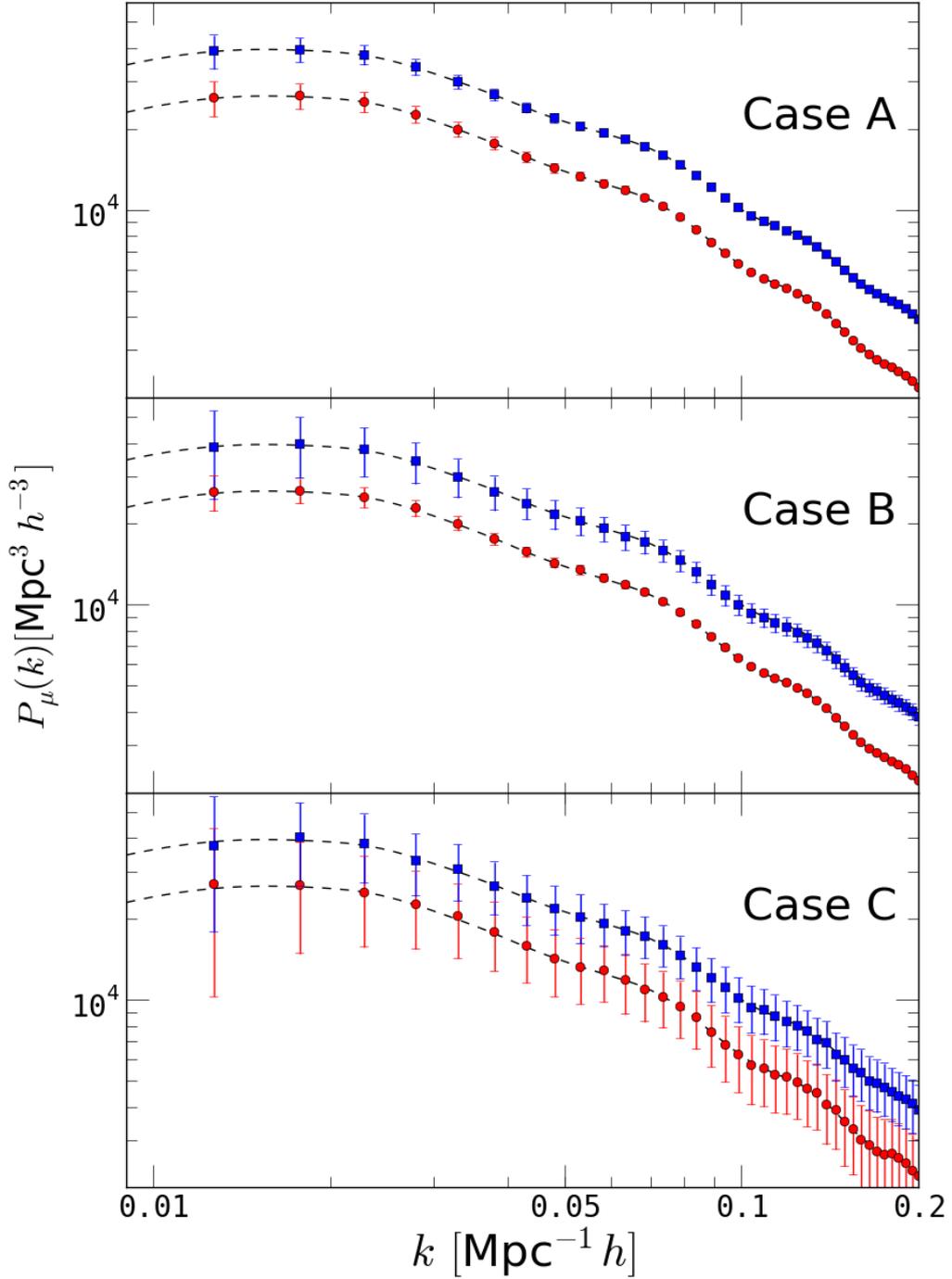
$$\frac{P_1}{P_2} = B_1^2(k, z, \mu_k) / B_2^2(k, z, \mu_k),$$

which can be used to place tighter constraints not only on the biases of each species but also on redshift-space distortions, non-gaussianities, etc. Notice that, for the case A, on the scales around  $k = 0.1 h \text{ Mpc}^{-1}$ , even though the uncertainties in each species are of order  $10^{-3}$  (fig. 4.3), the uncertainty in their ratio is close to  $5 \times 10^{-5}$ , more than an order of magnitude below. The uncertainty in the ratio is also much smaller than the one for the cross-spectra, in the same figure. Notice also that we are plotting the absolute value of the cross-covariance, since it runs negative for low signal-to-noise regimes (see section 4.3).

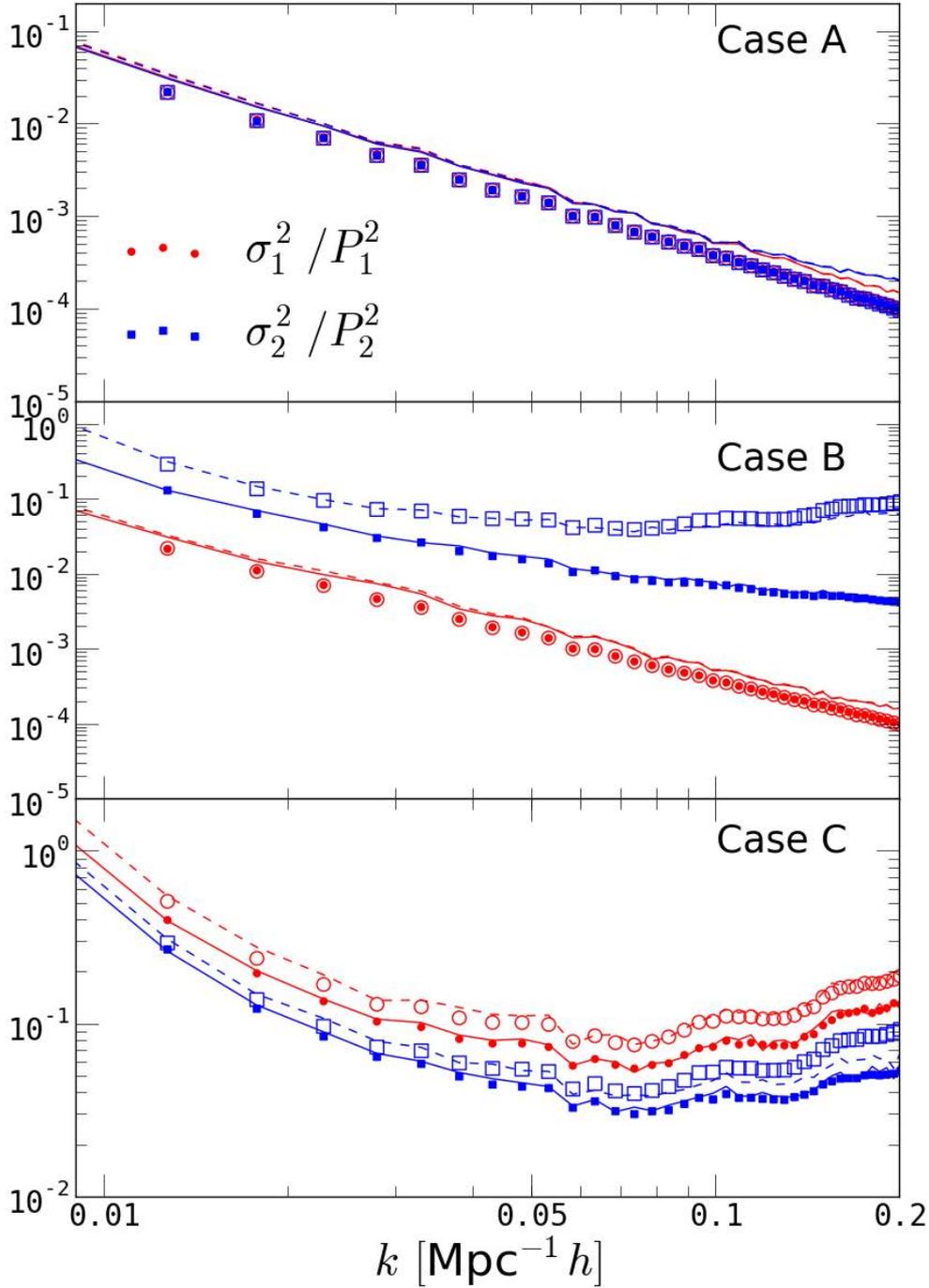
Figure 4.5 (upper panel) shows the full covariance matrix, empirically obtained from the lognormal realizations. We exploit the symmetry of the the matrix ( $C_{ij} = C_{ji}$ , as can be seen in eq. (3.8)) to plot both FKP and the multi-tracer together, with the diagonal subtracted. This figure represents the case B for tracer 2. The conclusion that can be

drawn is roughly the same we obtained from figure 4.3: FKP yields a larger covariance for the sparse tracer, while the multi-tracer estimators take into account the cross-correlations to suppress the covariance.

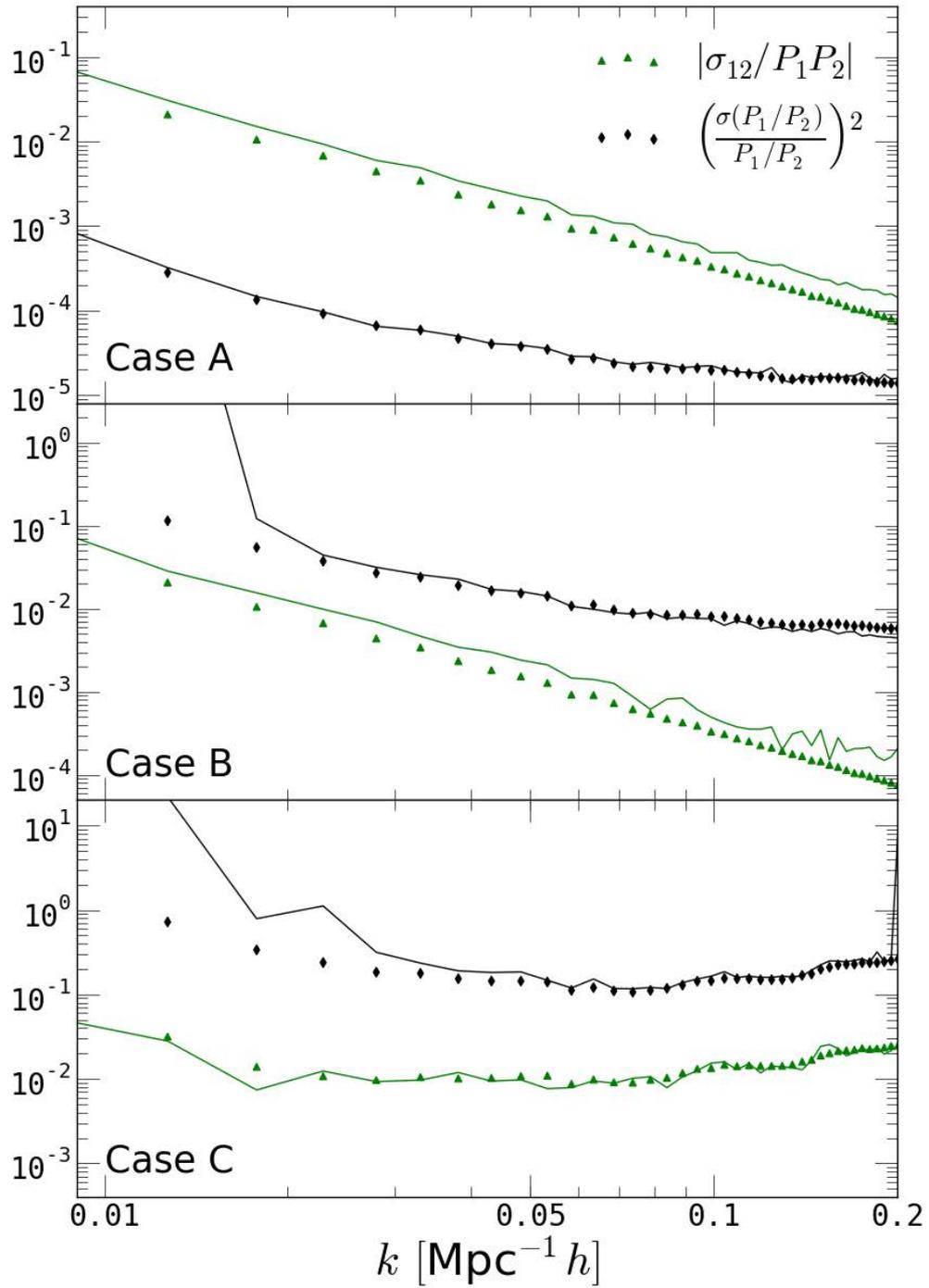
Figure 4.5 (lower panel) shows the correlation matrix for tracer 2 in case A. This matrix is defined in eq. (3.9) and basically subtracts the structure in the covariance matrix, showing more clearly the correlations between different modes. Values close to 0 mean that modes are perfectly uncorrelated, while 1.0 means complete correlation. By definition, the diagonal will show perfect correlation ( $Corr_{ii} = 1.0$ ) of a mode with itself. We see in this figure a similar behavior in the correlations of FKP and the multi-tracer among their modes. We performed the same test for the other tracers and cases and see a similar result. This means that, even though in general the FKP method yields larger uncertainties, its band estimates are roughly as uncorrelated as the multi-tracer method's.



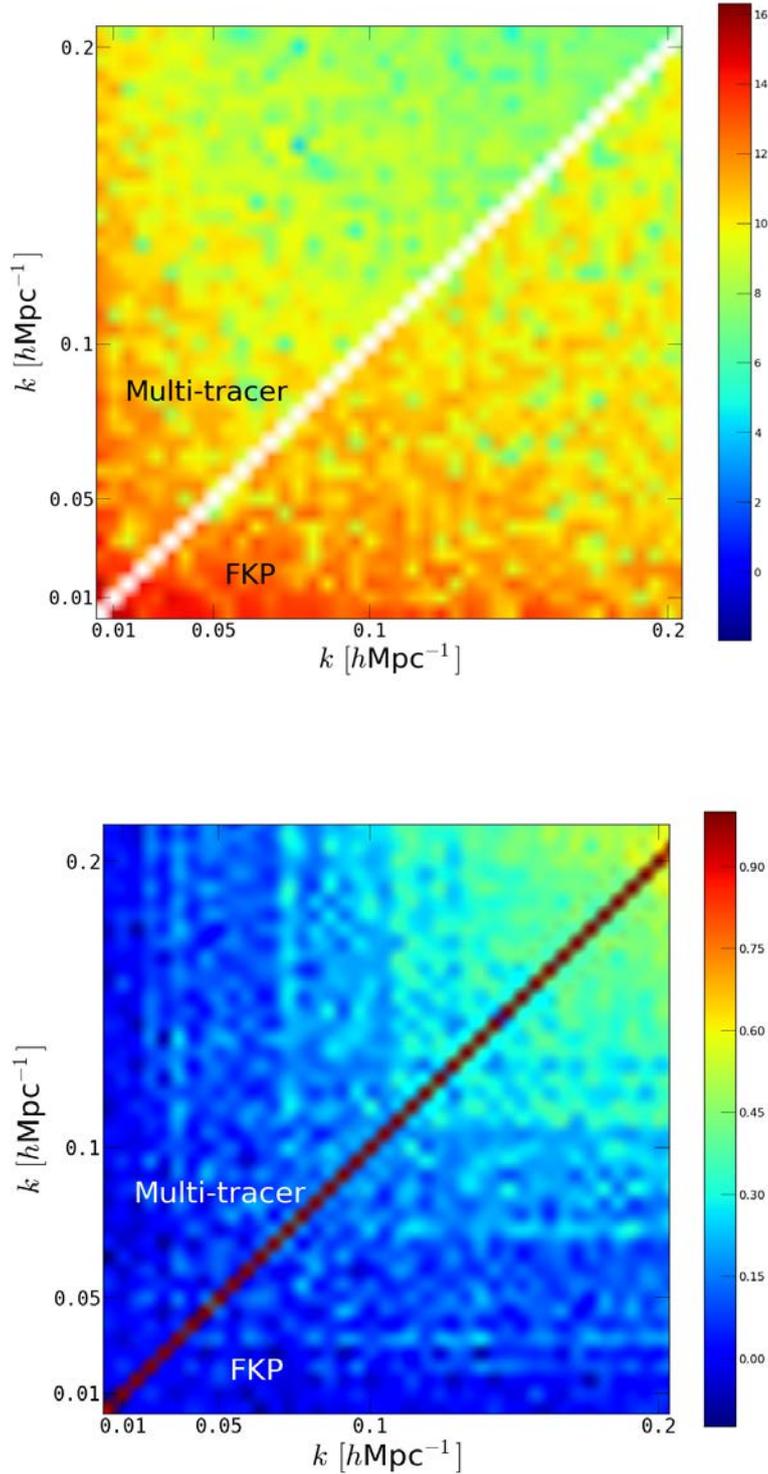
**Figure 4.2:** Galaxy power spectra estimated using the multi-tracer method. Dashed lines correspond to input power spectra from CAMB properly corrected for the deviation in the lognormal scheme. Red dots are the estimated spectra for the tracer with bias  $b_1 = 1.0$ , while blue squares are for the tracer with  $b_2 = 1.2$ . We are showing the average spectra of 1000 realizations for each case, which represents a different signal-to-noise situation (see table 4.1). The error bars ( $1\sigma$ ) are the theoretical ones, obtained by inverting the multi-tracer Fisher matrix (eq. (4.55)). This figure encapsulates 3 important results. Firstly, the multi-tracer method developed in this work manages to recover the power spectrum and it is robust when the signal is low. Second, the error bars for tracer 1 in cases A and B do not change significantly, which means the uncertainty in tracer 2 does not propagate into the other estimate. Finally, the error bars are enhanced when signal-to-noise is depleted, as expected.



**Figure 4.3:** Analysis of the relative variance  $[\sigma(k)/P(k)]^2$  for each tracer in each case. Dashed lines correspond to empirically obtained errors – the standard deviation of 1000 realizations. Solid symbols are theoretical predictions from the multi-tracer method, while empty symbols are theoretical predictions from FKP. Red curves and symbols are for tracer 1 ( $b_1 = 1.0$ ), and blue for tracer 2 ( $b_2 = 1.2$ ). All curves and symbols overlap in case A, so both estimators behave roughly the same way in this signal-to-noise regime (see table 4.1). The multi-tracer performs significantly better than FKP in all scales for cases B and C. The most interesting result is in case B, where the reduction in covariance is very clear for the sparser tracer2, when using the multi-tracer methods. We attribute this to the gain in information of using 2 tracers and taking into account their cross-correlations.



**Figure 4.4:** Theoretical and empirical covariances of the cross-spectra and ratio between spectra. Green triangles and curves are theoretical and empirical estimates, respectively, for the cross-spectra. Black diamonds and curves are, respectively, theoretical and empirical errors for the ratios between spectra. Case A shows clearly the gains of using the multi-tracer approach for constraining the ratio between effective biases (which may contain physical information such as redshift-space distortions and etc): the relative error in the ratio is much smaller than the relative error on the spectrum of each species separately (see fig. 4.3) and on the cross-correlation (in green). We see that this advantage only holds for the high signal-to-noise case.



**Figure 4.5: Upper panel:** Covariance matrix for tracer 2 in case B. The upper triangle is the result using the multi-tracer method, while the lower triangle results from FKP. The lower triangle shows higher values, indicating a larger covariance as estimated by FKP. The multi-tracer method performs significantly better. **Lower panel:** Correlation matrix for tracer 2, case A. In the definition of this matrix, the diagonal is unitary (see text and eq. (3.9)) and values close to 1 are correlated between each other. We see that both methods yield similar results.

# Chapter 5

## Conclusions

The current and forthcoming generation of astronomical instruments and redshift surveys will deliver overwhelming quantities of data. Control over systematics of various sorts is indispensable and have been occupying a central role in the research scenario. Nevertheless, minute understanding of the statistical properties of data is a powerful tool, and should not be disregarded. The goal of cosmology for the years to come is to test our current paradigm to exhaustion, and thoroughly look for flaws and imperfections in our theories. Ultimately, what we want is a consistent and comprehensive description of the universe. With accurate data and rigorous analysis, we are in the right direction to achieve that.

In the present work, we have explored the two sides of this endeavor. We began by developing the theoretical framework of the smooth, unperturbed universe, an ideal construct that serves amazingly well its objective of cementing the basic notions of spacetime and cosmological evolution. Assuming isotropy and the Copernican principle, both encapsulated in the FLRW metric, we studied the evolution of several components of the universe using the Einstein's Field Equations (section 2.1).

We then perturbed the smooth distributions of matter and energy to obtain a more realistic picture of the cosmos. With Newtonian fluid dynamics in an expanding background and the Boltzmann equations formalism, we were able to describe the evolution of radiation and dark matter in thermodynamical regimes which are slightly out of equilibrium. Taking into account the expanding causal horizon, we were able to track the evolution of small- and large-scale perturbations, and ultimately describe how they affected the gravitational potentials in the early universe. Along with a specific prescription for the initial conditions, we were able to make a theoretical prediction for the Fourier transform of the second moment of the matter distribution, which we called power spectrum. This quantity and its characteristic features were the starting point for developing a link between physical cosmology and observations (sec. 2.2).

Keeping the theoretical background in mind, we started reviewing the existing efforts

on how to properly – in fact, optimally – extract the power spectrum from actual observations of matter in the universe. We begin by describing the main sources of statistical (not systematic) noise that affect our analysis, to wit, the discreteness of the galaxy fields and the intrinsically poor statistics of finite volume samples (sec. 3.1).

Our desire is to exploit the rich available data to its limits, so we continue our work tackling the problem of optimally estimating the power spectrum from a given galaxy distribution. Following the existing literature, we studied well-known statistical tools based on the Cramér-Rao theorem and reproduced a general framework by which data sets can be optimally analyzed. During this study, we found that a major role is played by the Fisher matrix (sec. 3.2).

We then reviewed two popular and widely used sources in the literature that took a different but effective route in constructing optimal estimators. Feldman, Kaiser & Peacock (1994; FKP) designed a scheme in which they weighted the density contrast, built by subtracting from the actual observed galaxy distribution a synthetic one. They required that the relative error be minimum, and by doing so constrained the optimal weights (sec. 3.3). We subsequently tested this estimator in a variety of signal-to-noise regimes. Percival, Verde & Peacock (2004; PVP) later generalized this estimator by considering that different galaxies are biased in different ways, and obtained the optimal weights following the same idea of FKP (sec. 3.4).

Our next step was to refine the general formalism of optimal quadratic estimators. For that, we needed to better understand the derivation of the Fisher matrix for a single species of tracer. We then reviewed the generalization of this tool for an arbitrary number of tracers (secs. 4.1 and 4.2).

After gathering knowledge on the different existing methods, we created a set of optimal estimators that follow all the necessary conditions, including feasible implementation and reasonable computing time. We derived and thoroughly tested the expressions for the bias of the estimator, the subtleties of the random catalogs, the window functions and the theoretical auto- and cross-covariances. Our framework is also in tune with the recent advances of the large-scale structure research field, being a potential tool for developments in non-gaussianities and redshift distortions (sec. 4.3).

Finally, we extensively test our predictions in regimes with realistic physical motivations. We find remarkably good agreement between our predictions and the results of our tests. Not only our estimators robustly recover the power spectrum of matter the same way as other works in the literature, but also we find that they reduce the covariance and shrink error bars by significant factors (sec. 4.4). We publicly report our results in [52]. The specific contribution of the author of this dissertation was in implementing the estimators described throughout the work and performing the relevant tests. Many of the derivation steps were also discussed within the group as a whole.

Several future lines of work become clear at this point. Firstly, we would certainly want to apply our formalism to ever more realistic data sets. We could start by using N-body simulation mocks, which are the data sets that more closely resemble the physical properties of an actual observation. A distribution of this sort can easily be turned into a square grid with some smoothing scheme. To do that, we should be able to use grids preferably larger than  $256^3$  and study the distortions that appear in the power spectrum due to that convolution. The N-body mocks also allow for the possibility of changing the geometry of the survey into any arbitrary shape and employing an arbitrarily realistic selection function. In this case, we should be able to find efficient means to deconvolve the final power spectrum. We already have clear ideas on how to do that using Markov chain Monte Carlo methods [57]. These would be intermediate steps to apply our estimators to actual observational data. After careful analysis of these topics, we expect our pipeline to be used by the J-PAS collaboration community.

Secondly, we could introduce more theoretical subtleties to the estimators. One of them would be to account for the fact that, at least in photometric surveys, the best one can hope for is a probability distribution for the redshift of each object. Our weights should then reflect that uncertainty, while currently they assume well-defined positions for each object. We expect the largest scales to be only weakly affected by the photo- $z$  uncertainties, but the small scales would probably suffer a greater impact, since blurring the structure on these scales would generally cause a significant suppression of power. Another important research line would be to take a closer look to redshift distortions in our formalism. Even though our expressions are general enough to account for that, we have not performed tests in that direction. Perhaps the most direct question to be asked is how well we can constrain the redshift distortions parameters. To accomplish that, we would certainly have to introduce an expansion in Legendre polynomials of the power spectrum, and test it accordingly. This is also a reason why N-body simulations would be useful, because it is not clear how peculiar velocities could be introduced in the simple lognormal realizations.

Other alternatives exist and many questions remain open: how do the multi-tracer behave with reconstruction algorithms? What are the actual gains of the estimators when including refinements such as the 1-halo term? Can we figure out cross-applications between the multi-tracer and other probes such as, say, SNIa? What are the constraints on non-gaussianities from the multi-tracer estimators on real data sets?

Only one certainty exists: the territory is ripe for exploration.



# Appendices



# Appendix A

## Brief Thermal History

In the previous section, we have noted that the evolution of the energy density of a photon bath decreases with  $a(t)^{-4}$ . A heuristic argument to explain this fact is that not only photons will decrease in number density with  $a(t)^{-3}$ , following the volume expansion, but also their wavelengths will grow proportionally to the characteristic size of the box that contains them,  $a(t)$ , so their energy will fall with another factor of  $a(t)^{-1}$ .

The temperature of the photon bath, which is proportional to the energy of the photons, should then decrease as  $a(t)^{-1}$ . In other words, we can unambiguously define the evolution of the cosmic temperature as

$$T = \frac{T_0}{a(t)} = T_0(1+z). \quad (\text{A.1})$$

What equation (A.1) tells us is that the universe was much hotter in the past. It is natural to think that, when hot enough, this bath would hinder the formation of light elements or even protons and neutrons.

As the universe expands, however, the slowly cooling bath won't be energetic enough to ionize the existing atoms, allowing them to form. As long as the interactions between atoms or particles in the bath is efficient, the plasma will be in equilibrium and the production or annihilation of the constituents will continue. If, however, the rate of an interaction falls below the rate of expansion, the lack of physical contact will bring the production of elements to a halt, called *freeze-out*. Frozen abundances will be preserved and we can try to measure them today.

A brief thermal history along the lines of [58] can be summarized in the following timeline:

- Before  $10^{-14}$  s ( $T \gg 10$  TeV) : we may hope to probe these scales directly in future colliders, but maybe most of our knowledge on energies well above 10TeV, in the near future, will come from a possible B-mode polarization detection on the CMB.

If the result had held true, the BICEP2 announcement of the B-mode detection [59] would have yielded cosmological information on timescales before  $10^{-30}$  s. We also do not expect the usual particle physics and general relativity to be valid on scales before  $10^{-40}$  s or so;

- $\sim 10^{-5} - 10^{-14}$  s ( $T \approx 200$  MeV – 10 TeV): formation of the first baryons, and the standard model of particle physics is widely confirmed by experiments;
- $\sim 0.2$  s ( $T \approx 1 - 2$  MeV): the neutrino bath falls out of equilibrium and starts propagating without further scatterings. Also, the ratio of protons to neutrons freezes out too, determining the future element abundances that we can observe;
- $\sim 200$  s – 300 s ( $T \approx 0.05$  MeV): free protons and neutrons start forming helium and other light elements as nuclear reactions become efficient. This epoch is responsible for most of the predictions of Big Bang nucleosynthesis (BBN);
- $\sim 10^{12}$  s –  $10^{13}$  s ( $T < 1$  eV): nearly all free electrons and protons combine to form hydrogen. Curiously, this only happens when the temperature of the thermal bath is much below the binding energy of the H atom, 13.6 eV. This can be explained by the fact that a blackbody spectrum with temperature slightly below 13.6 eV has a long tail of high frequency photons, which can easily ionize newly formed hydrogen atoms. The universe abruptly becomes transparent to radiation, creating the last scattering surface that we observe in the CMB.

In general, BBN predictions have been very successful when confronted with experimental results. A more thorough description of the theory and observations of BBN can be found in [60].

# Appendix B

## The Multi-tracer Algorithm

We present here the actual implementation of the multi-tracer algorithm described in Section 4.4.1. The entire source code is written and explained below. The source code was written in Python 2.7.6<sup>1</sup>, and makes extensive use of the numerical package `numpy`<sup>2</sup>. This version is far from optimal in terms of speed, but nevertheless it obtains the correct results. We will omit here the implementation of the routines that generate the lognormal mocks used. In principle, it should suffice to use the code in this Appendix – written as a Python external *function* – in a loop that creates as many mocks as desired.

Even though the code transcribed below is in a specific programming language, it should be straightforward to be implemented in others. When using any specific feature of Python/numpy, we will make clear its meaning, so that translation to another language becomes easier. Comments are also widely used in the code, with the symbol `#`. Each block of code is in between red lines.

We start with the usual Python headers and definition of inputs of our external function:

---

```
#!/usr/bin/env python
# -*- coding: utf-8 -*-
import numpy as np
def multi(gridsize,nbar1,nbar2,nr1,nr2,ng1,ng2,b1,b2,phsize,bins):
```

---

The definition of each of the variables declared inside the `multi()` function is:

---

<sup>1</sup><https://www.python.org/download/releases/2.7.6/>

<sup>2</sup><http://www.numpy.org/>

- `gridsize`: The number of cells of the grid, assumed cubic (we used 256 in our applications);
- `nbar1,nbar2`:  $\bar{n}_1$  and  $\bar{n}_2$  for the tracers, this is supposed to be a cubic (`gridsize`<sup>3</sup>) matrix;
- `nr1,nr2`: the random catalogs for each tracer (a `gridsize`<sup>3</sup> matrix);
- `ng1,ng2`: the real catalog for each tracer (a `gridsize`<sup>3</sup> matrix);
- `b1,b2`: the initial guess for the bias of each tracer;
- `phsize`: the physical size of a side of the (cubic) grid;
- `bins`: the number of bins in which the power will be estimated.

Notice that, implemented this way, the function already assumes the random catalog was created and, therefore,  $\alpha$  is already set (see section 4.3). We did not write it as an input of this function, but the value used for this variable in what follows must be the same that was used to create the maps. In the next block, we define the overdensity fields  $\delta_\mu$  and the clustering strengths  $\mathcal{P}_\mu$  (called `Pc1` and `Pc2`) and the  $P_\mu$  (`Ps1` and `Ps2`).

---

```

nbar1=np.mean(nr1) #the number density of the random cat.
nbar2=np.mean(nr2)

P0= 10000.0*((gridsize/phsize)**3) #initial guess for
the underlying power spectrum

delta1= (ng1-alpha*nr1)/(alpha*nbar1) #overdensity for tracer 1
delta2= (ng2-alpha*nr2)/(alpha*nbar2) #overdensity for tracer 2

Pc1= nbar1*(b1**2)*P0 #P 'curved' for tracer 1
Pc2= nbar2*(b2**2)*P0
Ps1= (b1**2)*P0 #P 'straight' for tracer 1
Ps2= (b2**2)*P0

Pt= Pc1 + Pc2 #total P curved

```

---

Notice that here we make the explicit choice of a fiducial, constant power  $P_0$ , called  $P_0$ . The factor  $((\text{gridsize}/\text{phsize})**3)$  serves only to convert from “real” units to “grid” units. The returned quantities will be multiplied by the inverse factor to have proper physical units of volume. Notice also that the total clustering strength  $\mathcal{P} = \sum_{\beta} \mathcal{P}_{\beta}$  is called  $P_t$ .

Another important thing in the previous block is that `delta1`, `delta2`, `Pc1`, `Pc2` and `Pt` are all matrices. In the next block we define the overdensity fields  $f_{\mu}$  and  $f = \sum_{\mu} f_{\mu}$  using eqs. (4.48) and (4.49) explicitly, and take their  $n$ -dimensional Fourier transform using the `numpy` package called as `np.fft.fftn()`.

---

```
f1= (1.0 - Pc1/(1+Pt))*nbar1*b1*delta1 - (Pc1/(1+Pt))*nbar2*b2*delta2
f2= -(Pc2/(1+Pt))*nbar1*b1*delta1 + (1.0-Pc2/(1+Pt))*nbar2*b2*delta2
f= f1 + f2

f1k= np.fft.fftn(f1)
f2k= np.fft.fftn(f2)
fk= np.fft.fftn(f)
```

---

The FFT employed in the block above is very fast and takes approximately  $10^{-3}$ s for the  $256^3$  grids we’ve been using. For this step, however, the RAM required is roughly twice the size of the grid being transformed. In the next block we define  $\hat{Q}_{\mu}$  from eq. (4.50) and the inverse Fisher matrix from eqs. (4.55) and (4.56).

---

```
Q1= (1.0/(4*b1**2)) * 2* (f1k*np.conjugate(fk)).real
Q2= (1.0/(4*b2**2)) * 2* (f2k*np.conjugate(fk)).real

F1_11= Ps1*Ps1 * (4*(1+Pt)/Pt) * (1/Pc1 + (Pt-1)/(2*Pt))
F1_12= Ps1*Ps2 * (4*(1+Pt)/Pt) * ((Pt-1)/(2*Pt))
F1_21= Ps1*Ps2 * (4*(1+Pt)/Pt) * ((Pt-1)/(2*Pt))
F1_22= Ps2*Ps2 * (4*(1+Pt)/Pt) * (1/Pc2 + (Pt-1)/(2*Pt))

estP1= (F1_11*Q1 + F1_12*Q2)/(gridsize**3)
estP2= (F1_21*Q1 + F1_22*Q2)/(gridsize**3)
```

---

A few things should be noticed here. First, what we are calling `Q1` and `Q2` are only the integrands of eq. (4.50), without the  $(2\pi)^3$  term which is unnecessary in the FFT implementation of `numpy` and without subtracting shot noise, which we can do later. After defining the inverse Fisher matrix terms `F1...`, we multiply them by `Q1` and `Q2`, obtaining `estP1` and `estP2` according to equation (4.53), with a volume factor as normalization.

Notice that we skipped the average over shells in  $k$ -space to obtain the `estP`, and will carry out this calculation in a later block. This choice was made for convenience, to make the code a bit easier to debug. It is mathematically correct to commute the average over shells with the matrix multiplication only in the case we explored, where the grids for  $\bar{n}$  and  $F^{-1}$  are uniform and constant.

In the next block we define the bias to be subtracted from the estimated power, according to eqs. (4.59) and (4.68). We also called  $\delta P_\mu$  by `Pshot` just as a reminder that it should be the shot noise in the 1 tracer, FKP case with which we can compare.

---

```

deltaQ1= (1+alpha)*(0.5)*nbar1/((1+Pt)**2)
deltaQ2= (1+alpha)*(0.5)*nbar2/((1+Pt)**2)

Pshot1= (1.0/gridsize**3) * np.sum(F1_11*deltaQ1 + F1_12*deltaQ2)
Pshot2= (1.0/gridsize**3) * np.sum(F1_21*deltaQ1 + F1_22*deltaQ2)

```

---

We now define the frequencies sampled in the FFT and the edges of the bins of  $k$ . In the last line we define the  $\Delta k$ .

---

```

kfft=np.fft.fftfreq(gridsize) #generates the fft frequencies
kminfft=np.amin(np.abs(kfft)) #finds the lowest frequency (should be 0)
kmaxfft=np.amax(np.abs(kfft)) #finds the highest frequency (Nyquist)
kmax=np.sqrt(3)*kmaxfft #higher absolute value of k is given by this
kmin=kminfft #lowest possible absolute value of k
kNy=kmaxfft #Nyquist freq.
k_bins=np.linspace(kmin,kmax,bins+1) #edges for bins in k
delta_k=k_bins[4]-k_bins[3]

```

---

In the next block, we perform the average of shells in  $k$ -space, which is the bottleneck of the code because it uses some long loops along with the fast, vectorized operations. In principle, this whole block can be exchanged by something faster.

---

```

P1=np.zeros(bins) #initializing final spectrum
P2=np.zeros(bins)
counts1=np.zeros(bins) #initializing counts vector
counts2=np.zeros(bins)

kfft2=kfft**2
for ii in range(len(kfft)):
    kx2=kfft[ii]**2
    for jj in range(len(kfft)):
        ky2=kfft[jj]**2
        k_sum = np.sqrt(kx2 + ky2 + kfft2) #absolute value of k
        m = np.asarray(k_sum/delta_k-0.000001).astype(int) #gridsize/2
indices
        zcounter=0
        for ind in m: #iterating over the indices to attribute the power
to the correct bins

            P1[ind]=P1[ind]+estP1[ii,jj,zcounter]
            P2[ind]=P2[ind]+estP2[ii,jj,zcounter]
            counts1[ind]=counts1[ind]+1
            counts2[ind]=counts2[ind]+1

            zcounter=zcounter+1

P1=P1/counts1 - Pshot1 #mean power on each bin and shot noise correction
P2=P2/counts2 - Pshot2

```

---

In the last line of the block above, we divided each spectrum by the volume of the shell (the number of counts in each mode) and subtract the bias of the estimator. What the parts inside the loop are doing is finding the bin to which each mode belongs and add the corresponding power to that bin.

In the final block of the code, we re-define the bins such that they are in between each of the bin edges, and multiply by the appropriate factors to return  $k$ ,  $P1$  and  $P2$  in physical units ( $h^{-3} \text{ Mpc}^3$  and  $h \text{ Mpc}^{-1}$ ).

```
k=np.zeros(len(P1))
for i in range(len(k_bins)-1):
    k[i]=(k_bins[i]+k_bins[i+1])/2.0

    return k*(2*np.pi*gridsize/phsize),P1*((phsize/gridsize)**3),
P2*((phsize/gridsize)**3)
```

---

For the error bars, it suffices to take the average over  $k$ -shells of the matrices  $F1...$  define in the previous blocks. When doing this, one can change the initial guess  $P0$  for the theoretical spectrum in each bin of  $k$ , in order to get the error bars for the corresponding theory.

As a final remark, notice that the current implementation depends heavily on the uniformity of most of the grids. In fact, the only grids which are not constant in the studied case are those involving `delta1` and `delta2`. For a general application, this source code should be reviewed.

# Bibliography

- [1] The Dark Energy Survey Collaboration, *The Dark Energy Survey*, *ArXiv Astrophysics e-prints* (Oct., 2005) [[astro-ph/0510346](#)].
- [2] D. G. York and SDSS Collaboration, *The Sloan Digital Sky Survey: Technical Summary*, *The Astronomical Journal* **120** (Sept., 2000) 1579–1587, [[astro-ph/0006396](#)].
- [3] LSST Science Collaboration, P. A. Abell, J. Allison, S. F. Anderson, J. R. Andrew, J. R. P. Angel, L. Armus, D. Arnett, S. J. Asztalos, T. S. Axelrod, and et al., *LSST Science Book, Version 2.0*, *ArXiv e-prints* (Dec., 2009) [[arXiv:0912.0201](#)].
- [4] N. Benitez et al., *J-PAS: The Javalambre-Physics of the Accelerated Universe Astrophysical Survey*, *ArXiv e-prints* (Mar., 2014) [[arXiv:1403.5237](#)].
- [5] M. Takada, R. S. Ellis, M. Chiba, et al., *Extragalactic science, cosmology, and Galactic archaeology with the Subaru Prime Focus Spectrograph*, **66** (Feb., 2014) 1, [[arXiv:1206.0737](#)].
- [6] R. Laureijs, J. Amiaux, S. Arduini, J. . Auguères, J. Brinchmann, R. Cole, M. Cropper, C. Dabin, L. Duvet, A. Ealet, and et al., *Euclid Definition Study Report*, *ArXiv e-prints* (Oct., 2011) [[arXiv:1110.3193](#)].
- [7] R. J. Bouwens et al., *Discovery of  $z \sim 8$  Galaxies in the Hubble Ultra Deep Field from Ultra-Deep WFC3/IR Observations*, *The Astrophysical Journal Letters* **709** (Feb., 2010) L133–L137, [[arXiv:0909.1803](#)].
- [8] L. R. Abramo et al., *Measuring large-scale structure with quasars in narrow-band filter surveys*, *Monthly Notices of the Royal Astronomical Society* **423** (July, 2012) 3251–3267, [[arXiv:1108.2657](#)].
- [9] T. Budavri, *A unified framework for photometric redshifts*, *The Astrophysical Journal* **695** (2009), no. 1 747.
- [10] R. A. Sunyaev and I. B. Zeldovich, *Microwave background radiation as a probe of the contemporary structure and history of the universe*, **18** (1980) 537–560.

- [11] S. W. Allen, A. E. Evrard, and A. B. Mantz, *Cosmological Parameters from Observations of Galaxy Clusters*, **49** (Sept., 2011) 409–470, [[arXiv:1103.4829](#)].
- [12] R. A. Battye, I. W. A. Browne, C. Dickinson, G. Heron, B. Maffei, and A. Pourtsidou, *H I intensity mapping: a single dish approach*, *Monthly Notices of the Royal Astronomical Society* **434** (Sept., 2013) 1239–1256, [[arXiv:1209.0343](#)].
- [13] J.-P. Uzan, *The fundamental constants and their variation: observational and theoretical status*, *Rev. Mod. Phys.* **75** (Apr, 2003) 403–455.
- [14] **Supernova Search Team** Collaboration, A. G. Riess et al., *Observational evidence from supernovae for an accelerating universe and a cosmological constant*, *Astron.J.* **116** (1998) 1009–1038, [[astro-ph/9805201](#)].
- [15] **Supernova Cosmology Project** Collaboration, S. Perlmutter et al., *Measurements of Omega and Lambda from 42 high redshift supernovae*, *Astrophys.J.* **517** (1999) 565–586, [[astro-ph/9812133](#)].
- [16] G. Bertone, D. Hooper, and J. Silk, *Particle dark matter: evidence, candidates and constraints*, *Physics Reports* **405** (Jan., 2005) 279–390, [[hep-ph/0404175](#)].
- [17] K. G. Begeman, A. H. Broeils, and R. H. Sanders, *Extended rotation curves of spiral galaxies - Dark haloes and modified dynamics*, *Monthly Notices of the Royal Astronomical Society* **249** (Apr., 1991) 523–537.
- [18] M. Markevitch, *Chandra Observation of the Most Interesting Cluster in the Universe*, in *The X-ray Universe 2005* (A. Wilson, ed.), vol. 604 of *ESA Special Publication*, p. 723, Jan., 2006. [astro-ph/0511345](#).
- [19] S. Carroll, *Spacetime and Geometry: An Introduction to General Relativity*. Addison Wesley, 2004.
- [20] S. Dodelson, *Modern cosmology*. Academic Press, 1 ed., 2003.
- [21] S. M. Carroll, *The Cosmological constant*, *Living Rev.Rel.* **4** (2001) 1, [[astro-ph/0004075](#)].
- [22] D. W. Hogg, *Distance measures in cosmology*, *ArXiv Astrophysics e-prints* (May, 1999) [[astro-ph/9905116](#)].
- [23] D. Baumann, *TASI Lectures on Inflation*, *ArXiv e-prints* (July, 2009) [[arXiv:0907.5424](#)].
- [24] H. Kodama and M. Sasaki, *Cosmological Perturbation Theory*, *Progress of Theoretical Physics Supplement* **78** (1984) 1.

- [25] J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay, *The statistics of peaks of Gaussian random fields*, *The Astrophysical Journal* **304** (May, 1986) 15–61.
- [26] P. Peter and J. Uzan, *Primordial cosmology*. Oxford graduate texts. Oxford University Press, 2009.
- [27] L. Anderson et al., *The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Data Release 9 spectroscopic galaxy sample*, *Monthly Notices of the Royal Astronomical Society* **427** (Dec., 2012) 3435–3467, [arXiv:1203.6594].
- [28] Planck Collaboration, P. A. R. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, C. Baccigalupi, A. J. Banday, and et al., *Planck 2013 results. XV. CMB power spectra and likelihood*, *Astronomy and Astrophysics* **571** (Nov., 2014) A15, [arXiv:1303.5075].
- [29] D. J. Eisenstein, H.-J. Seo, E. Sirko, and D. N. Spergel, *Improving Cosmological Distance Measurements by Reconstruction of the Baryon Acoustic Peak*, *The Astrophysical Journal* **664** (Aug., 2007) 675–679, [astro-ph/0604362].
- [30] M. Tegmark, A. J. S. Hamilton, M. A. Strauss, M. S. Vogeley, and A. S. Szalay, *Measuring the Galaxy Power Spectrum with Future Redshift Surveys*, *The Astrophysical Journal* **499** (May, 1998) 555–576, [astro-ph/9708020].
- [31] M. Tegmark, *How to measure CMB power spectra without losing information*, *Physical Review D* **55** (May, 1997) 5895–5907, [astro-ph/9611174].
- [32] M. Tegmark, A. N. Taylor, and A. F. Heavens, *Karhunen-Loève Eigenvalue Problems in Cosmology: How Should We Tackle Large Data Sets?*, *The Astrophysical Journal* **480** (May, 1997) 22–35, [astro-ph/9603021].
- [33] M. S. Vogeley and A. S. Szalay, *Eigenmode Analysis of Galaxy Redshift Surveys. I. Theory and Methods*, *The Astrophysical Journal* **465** (July, 1996) 34, [astro-ph/9601185].
- [34] M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics*, vol. 2. 1969.
- [35] M. Tegmark, M. R. Blanton, M. A. Strauss, F. Hoyle, D. Schlegel, R. Scoccimarro, M. S. Vogeley, D. H. Weinberg, I. Zehavi, and SDSS Collaboration, *The Three-Dimensional Power Spectrum of Galaxies from the Sloan Digital Sky Survey*, *The Astrophysical Journal* **606** (May, 2004) 702–740, [astro-ph/0310725].

- [36] H. A. Feldman, N. Kaiser, and J. A. Peacock, *Power-spectrum analysis of three-dimensional redshift surveys*, *The Astrophysical Journal* **426** (May, 1994) 23–37, [astro-ph/9304022].
- [37] A. J. S. Hamilton, *Towards optimal measurement of power spectra - I. Minimum variance pair weighting and the Fisher matrix*, *Monthly Notices of the Royal Astronomical Society* **289** (Aug., 1997) 285–294, [astro-ph/9701008].
- [38] G. Efstathiou, N. Kaiser, W. Saunders, A. Lawrence, M. Rowan-Robinson, R. S. Ellis, and C. S. Frenk, *Largescale Clustering of IRAS Galaxies*, *Monthly Notices of the Royal Astronomical Society* **247** (Nov., 1990) 10P.
- [39] S. Cole, W. J. Percival, J. A. Peacock, P. Norberg, C. M. Baugh, C. S. Frenk, I. Baldry, J. Bland-Hawthorn, T. Bridges, R. Cannon, M. Colless, C. Collins, W. Couch, N. J. G. Cross, G. Dalton, V. R. Eke, R. De Propris, S. P. Driver, G. Efstathiou, R. S. Ellis, K. Glazebrook, C. Jackson, A. Jenkins, O. Lahav, I. Lewis, S. Lumsden, S. Maddox, D. Madgwick, B. A. Peterson, W. Sutherland, and K. Taylor, *The 2dF Galaxy Redshift Survey: power-spectrum analysis of the final data set and cosmological implications*, *Monthly Notices of the Royal Astronomical Society* **362** (Sept., 2005) 505–534, [astro-ph/0501174].
- [40] A. Dressler, *Galaxy morphology in rich clusters - Implications for the formation and evolution of galaxies*, *The Astrophysical Journal* **236** (Mar., 1980) 351–365.
- [41] A. Dressler, A. Oemler, Jr., W. J. Couch, I. Smail, R. S. Ellis, A. Barger, H. Butcher, B. M. Poggianti, and R. M. Sharples, *Evolution since  $z = 0.5$  of the Morphology-Density Relation for Clusters of Galaxies*, *The Astrophysical Journal* **490** (Dec., 1997) 577–591, [astro-ph/9707232].
- [42] G. P. Smith, T. Treu, R. S. Ellis, S. M. Moran, and A. Dressler, *Evolution since  $z = 1$  of the Morphology-Density Relation for Galaxies*, *The Astrophysical Journal* **620** (Feb., 2005) 78–87, [astro-ph/0403455].
- [43] L. Guzzo, M. A. Strauss, K. B. Fisher, R. Giovanelli, and M. P. Haynes, *Redshift-Space Distortions and the Real-Space Clustering of Different Galaxy Types*, *The Astrophysical Journal* **489** (Nov., 1997) 37–48, [astro-ph/9706150].
- [44] W. J. Percival, L. Verde, and J. A. Peacock, *Fourier analysis of luminosity-dependent galaxy clustering*, *Monthly Notices of the Royal Astronomical Society* **347** (Jan., 2004) 645–653, [astro-ph/0306511].
- [45] A. Dekel and O. Lahav, *Stochastic Nonlinear Galaxy Biasing*, *The Astrophysical Journal* **520** (July, 1999) 24–34, [astro-ph/9806193].

- [46] D. J. Eisenstein, I. Zehavi, D. W. Hogg, R. Scoccimarro, M. R. Blanton, R. C. Nichol, R. Scranton, H.-J. Seo, M. Tegmark, Z. Zheng, et al., *Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies*, *The Astrophysical Journal* **633** (Nov., 2005) 560–574, [astro-ph/0501171].
- [47] U. Seljak, *Extracting Primordial Non-Gaussianity without Cosmic Variance*, *Physical Review Letters* **102** (Jan., 2009) 021302, [arXiv:0807.1770].
- [48] L. R. Abramo and K. E. Leonard, *Why multitracer surveys beat cosmic variance*, *Monthly Notices of the Royal Astronomical Society* **432** (June, 2013) 318–326, [arXiv:1302.5444].
- [49] P. McDonald and U. Seljak, *How to evade the sample variance limit on measurements of redshift-space distortions*, *Journal of Cosmology and Astroparticle Physics* **10** (Oct., 2009) 7, [arXiv:0810.0323].
- [50] L. R. Abramo, *The full Fisher matrix for galaxy surveys*, *Monthly Notices of the Royal Astronomical Society* **420** (Mar., 2012) 2042–2057, [arXiv:1108.5449].
- [51] N. Hamaus, U. Seljak, and V. Desjacques, *Optimal constraints on local primordial non-Gaussianity from the two-point statistics of large-scale structure*, *Physical Review D* **84** (Oct., 2011) 083509, [arXiv:1104.2321].
- [52] L. R. Abramo, L. F. Secco, and A. Loureiro, *Fourier analysis of multi-tracer cosmological surveys*, *ArXiv e-prints* (May, 2015) [arXiv:1505.0410].
- [53] M. White, Y.-S. Song, and W. J. Percival, *Forecasting cosmological constraints from redshift surveys*, *Monthly Notices of the Royal Astronomical Society* **397** (Aug., 2009) 1348–1354, [arXiv:0810.1518].
- [54] J. A. Peacock and D. Nicholson, *The large-scale clustering of radio galaxies*, *Monthly Notices of the Royal Astronomical Society* **253** (Nov., 1991) 307–319.
- [55] P. Coles and B. Jones, *A lognormal model for the cosmological mass distribution*, *Monthly Notices of the Royal Astronomical Society* **248** (Jan., 1991) 1–13.
- [56] A. Lewis, A. Challinor, and A. Lasenby, *Efficient Computation of Cosmic Microwave Background Anisotropies in Closed Friedmann-Robertson-Walker Models*, *The Astrophysical Journal* **538** (Aug., 2000) 473–476, [astro-ph/9911177].
- [57] A. Loureiro, *Galaxy power spectrum analysis: A monte-carlo approach*, Master’s thesis, Universidade de São Paulo, 2015.

- [58] V. Mukhanov, *Physical Foundations of Cosmology*. Cambridge University Press, 2005.
- [59] **BICEP2** Collaboration, P. A. R. Ade et al., *Detection of B-Mode Polarization at Degree Angular Scales by BICEP2*, *Physical Review Letters* **112** (June, 2014) 241101, [[arXiv:1403.3985](#)].
- [60] K. A. Olive, G. Steigman, and T. P. Walker, *Primordial nucleosynthesis: theory and observations*, *Physics Reports* **333** (Aug., 2000) 389–407, [[astro-ph/9905320](#)].