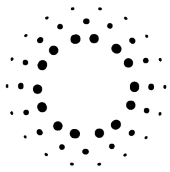

IFT - UNESP
INSTITUTO DE FÍSICA TEÓRICA

DISSERTATION IFT–D.005/2023

Cosmology with an uncertain galaxy-matter relation

Abdias Aires de Queiroz Neto

Advisor
Dr. Rogerio Rosenfeld

Co-Advisors
Dr. Nickolas Kokron
Dr. Felipe Andrade-Oliveira

September 2023

Queiroz Neto, Abdias Aires de
Q3c Cosmology with an uncertain galaxy-matter relation / Abdias Aires de Queiroz Neto. – São Paulo, 2024
67 f.: il. (algumas color.)

Dissertação (mestrado) – Universidade Estadual Paulista (Unesp), Instituto de Física Teórica (IFT), São Paulo
Orientador: Rogério Rosenfeld
Coorientador: Nickolas Kokron
Coorientador: Felipe Andrade-Oliveira

1. Cosmologia. 2. Matéria escura (Astronomia). 3. Estrutura em larga escala (Astronomia). I. Título

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de Física Teórica (IFT), São Paulo. Dados fornecidos pelo autor(a).

COSMOLOGY WITH AN UNCERTAIN

GALAXY-MATTER RELATION

Dissertação de Mestrado apresentada ao Instituto de Física Teórica do Câmpus de São Paulo, da Universidade Estadual Paulista "Júlio de Mesquita Filho", como parte dos requisitos para obtenção do título Mestre em Física, Especialidade Física Teórica.

Comissão Examinadora:

Prof. Dr. ROGÉRIO ROSENFELD (Orientador)
Instituto de Física Teórica/UNESP

Prof. Dr. MARCUS VINÍCIUS BORGES TEIXEIRA LIMA
IFUSP/Instituto de Física da USP

Prof. Dr. LUIS RAUL WEBER ABRAMO
IFUSP/Instituto de Física da USP

Conceito: Aprovado

São Paulo, 15 de setembro de 2023.

Para Juarez, Kalina e Mateus Aires

Agradecimentos

Sou grato a Rogerio Rosenfeld pela orientação calorosa e questionadora. Agradeço a Nickolas Kokron, Felipe Andrade-Oliveira e Vivian Miranda pelas muitas discussões e crítica atenciosa.

Dedico este trabalho a Juarez, Kalina e Mateus Aires, cujo apoio amoroso tornou estes passos em Física mais firmes e céleres.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Resumo

Para extrair informação da estrutura em grande escala do universo, um requisito importante é conectar o campo de galáxias observado à distribuição de matéria escura. O modelo mais conhecido dessa relação, *bias* linear de galáxias, perde acurácia em escalas $k \sim 0.1h\,\mathrm{Mpc}^{-1}$. A solução usual é o chamado "corte de escalas", que consiste em descartar dados da análise com o fim de limitar efeitos sistemáticos.

Testamos neste trabalho uma abordagem alternativa: mantemos modos k superiores a $k_{\mathrm{max}} \sim 0.1h\,\mathrm{Mpc}^{-1}$ na estimativa, mas com peso estatístico reduzido. Este último é avaliado conforme a credibilidade do modelo – o nível de incerteza teórica – no regime em questão. Isso pode ser visto como um corte gradual de escalas e pode produzir, como pretendemos mostrar, uma precisão maior sobre parâmetros cosmológicos para o mesmo nível de acurácia.

A fim de dar contexto às investigações acima, apresentamos também uma introdução à formação de estruturas em teoria de perturbações. Os resultados usuais são baseados na dinâmica de fluidos perfeitos. Descrevemos correções sobre essa teoria baseadas numa regularização de seu parâmetro de expansão, o que resulta em uma teoria de campos efetivos para a formação de estruturas. Considerações adicionais permitem relacionar a distribuição de matéria escura ao campo de galáxias observado, o que torna este último uma sonda de importância em cosmologia observacional.

Palavras Chaves: Cosmologia; *bias* de galáxias; matéria escura; formação de estruturas; estrutura de grande escala.

Áreas do conhecimento: Física; cosmologia.

Abstract

In order to extract information from the large-scale structure of the universe, a major requirement is to connect the observed galaxy field to the distribution of dark matter. The best-known model for that relationship, linear galaxy bias, loses accuracy on scales $k \sim 0.1h\,\mathrm{Mpc}^{-1}$. The usual approach to that problem is the so-called "scale cut", and it consists of discarding data from the analysis to prevent systematic effects.

We here test an alternative approach, keeping those k-modes higher than $k_{\mathrm{max}} \sim 0.1h\,\mathrm{Mpc}^{-1}$ in the estimation with reduced statistical weight. The latter is assessed from the model's credibility – the level of theoretical uncertainty – on the regime in question. This may be viewed as a gradual scale cut, and it can lead, as we aim to show, to higher precision on cosmological parameters for the same level of estimation accuracy.

To put the above inquiries in context, we also present an overview of structure growth in perturbation theory. The standard results are based on the perfect-fluid dynamics. We describe corrections to that theory based on a regularization of the expansion parameter, which leads to an effective field theory of structure growth. Further considerations permit the connection of the matter distribution to the observable galaxy field, making the latter a fundamental probe in observational cosmology.

Keywords: Cosmology; galaxy bias; dark matter; structure formation; large-scale structure.

Subject areas: Physics; cosmology.

Contents

Chapter 1

Introduction

Large-scale structure contains key information on the universe and its constituents. It can complement, and in some cases supersede, what can be gauged from the Cosmic Microwave Background – about the sum of neutrino masses or primordial non-Gaussianities, for example [6]. Present-day observational programs, such as the Legacy Survey of Space and Time (LSST) [16], are expected to probe that structure with unprecedented precision, relying chiefly on data for galaxy clustering, gravitational lensing, and their cross-correlations. The increase in experimental precision must be matched with an effort on the theoretical side. This raises the need, for example, to efficiently treat data on scales for which the available models are inaccurate, so that experimental information isn't greatly sub-utilized.

In this work, we consider theoretical uncertainty on galaxy clustering and galaxy-galaxy lensing, specially in the range of non-linear scales. A simple model permits the connection of those observables to the matter power spectrum, through the addition of one "linear-bias" parameter per redshift bin. Predictions of this model, however, begin to deviate significantly from the true spectra at scales near $k \sim 0.1h\,\mathrm{Mpc}$.

A current approach to the problem – used, for example, in the Dark Energy Survey [19] – is to discard from the analysis those higher k-modes for which the model is deemed unreliable. This method, the "scale cut", has the advantage of guaranteeing the model's high accuracy on the range of scales under analysis, which prevents systematic effects. On the other hand, discarding data reduces the estimation's precision.

We study an alternative approach to the scale cut. Modes for which a model is inaccurate can be kept in the analysis with reduced statistical weight, provided that weight matches the level of theoretical uncertainty in that regime. The uncertainty can be assessed in several ways – for example, from the error between the model's prediction and what is deemed to be the "true spectrum" in a typical scenario. This, in our work, will be assumed to be the nonlinear-bias spectrum at one-loop

order. If theoretical uncertainty is included in the analysis in a consistent way, as we aim to show, the maximum mode available can be typically increased from near $0.1h\,\mathrm{Mpc}^{-1}$ to near $0.3h\,\mathrm{Mpc}^{-1}$, without adding error to the results, and still under the linear-bias approximation. One can then use the same model to make analyses of higher precision.

Knowledge of the relation between galaxy and matter power spectra, together with a theory that connects the latter to cosmological models, allows those models to be constrained. The fluctuations in the matter density can be predicted from the fluid equations, which can be solved perturbatively in terms of a workable expansion parameter – the "large-scale" matter overdensity. The latter is obtained by removing from the matter field those modes for which perturbation theory is expected to fail. This leads to an effective-field (EFT) correction to the so-called *standard perturbation theory*. Further considerations permit the expression of the galaxy overdensity in terms of the matter overdensity with a minimal number of new parameters. This leads to a theory of *galaxy bias*.

To put the analyses above in context, we treat structure growth and galaxy bias in chapters 2 and 3. They are relevant to our purpose for two reasons: (1) our model for the galaxy power spectrum relies on linear bias, but is based on the matter power spectrum of EFT at one-loop order; (2) what we consider the "true" galaxy power spectrum for the assessment of theoretical uncertainty is given by the nonlinear bias expansion up to second order in the matter overdensity. Finally, in chapter 4, we make use of these concepts to analyze mock data of large-scale structure whilst taking account of theoretical uncertainty. We present our conclusions in chapter 5.

Chapter 2

Structure growth in standard perturbation theory

To put the main work in context, we here review the standard approach to structure growth through perturbation theory. The aim is to obtain the evolution of matter-density fluctuations, $\delta(\boldsymbol{k})$, across cosmic time. This will be based on the fluid equations expressed in the Newtonian regime, which is a valid approximation on scales smaller than the Hubble radius – that is, for $k \gg aH \sim 3 \cdot 10^{-4} h\,\mathrm{Mpc}^{-1}$.

The equation of motion of dark-matter particles in the gravitational field, together with Liouville's theorem on the conservation of phase-space density, leads to the "collisionless Boltzmann equation" governing how the distribution function evolves in phase space. From the latter one obtains the equations for the density and velocity fields – the continuity and Euler equations –, which, under the assumption of a "pressureless perfect fluid", form a closed system. That system can be solved order by order on the density field.

The above analysis proves inadequate for scales below $k \sim 0.1h\,\mathrm{Mpc}^{-1}$, on which the overdensity approaches unity and ceases to function as an expansion parameter. Solutions to that problem will be considered in chapter 3.

2.1 Isotropy and homogeneity

In this section, we recall a few of the standard results in background cosmology. The large-scale isotropy and homogeneity of the universe imply the Friedmann-Lemaître-Robertson-Walker (FLRW) metric:

$$\mathrm{d}s^2 = \mathrm{d}t^2 - a^2(t)\delta_{ij}^{(\mathrm{K})}\mathrm{d}x^i\mathrm{d}x^j \tag{2.1}$$

where the *scale factor*, $a(t)$, is an arbitrary function of time and $\delta_{ij}^{(\mathrm{K})}$ is the Kronecker symbol. Summation over repeated indices is adopted. We have assumed no spatial curvature in the metric, which is in agreement with observational evidence [2].

The above equation is written in terms of the "comoving" coordinates x^i – i.e., such that accompany the expansion of the universe. Physical distances are then defined as $\boldsymbol{r} = a\boldsymbol{x}$. It is also convenient to define the *conformal time* η through $a\mathrm{d}\eta = \mathrm{d}t$, so that the metric can be written as:

$$\mathrm{d}s^2 = a^2(\eta)\left(\mathrm{d}\eta - \delta^{(\mathrm{K})}_{ij}\mathrm{d}x^i\mathrm{d}x^j\right) \tag{2.2}$$

The most general stress-energy tensor compatible with spatial symmetry is that of a perfect fluid [8]. In the comoving frame, it can be written as:

$$T^\mu_\nu = \mathrm{diag}\,(\rho, -P, -P, -P) \tag{2.3}$$

and, in general:

$$T^\mu_\nu = (\rho + P)U^\mu U_\nu - P\delta^\mu_\nu \tag{2.4}$$

where P is pressure, ρ is the energy density, and $U^\mu \equiv \mathrm{d}x^\mu/\mathrm{d}s$.

The Einstein equations, given the metric and the stress-energy tensor defined above, yield the two Friedmann equations:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho \tag{2.5}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P) \tag{2.6}$$

where G is the gravitational constant. We denote $\dot{f} \equiv \mathrm{d}f/\mathrm{d}t$.

We likewise define the *Hubble parameter*, $H \equiv \dot{a}/a$, and its "conformal" counterpart, $\mathcal{H} \equiv a'/a$. Here, $(\bullet)'$ denotes a derivative with respect to conformal time: $f' \equiv \mathrm{d}f/\mathrm{d}\eta$. The Friedmann equations can then be put in the useful form:

$$\mathcal{H}^2 = \frac{8\pi G}{3}a^2\rho \tag{2.7}$$

$$\mathcal{H}' = -\frac{4\pi G}{3}a^2(\rho + 3P) \tag{2.8}$$

2.2 Departures from homogeneity

To study structure formation, one must consider departures from a homogeneous background. In perturbation theory, this is done by writing the metric as

$g_{\mu\nu} = \bar{g}_{\mu\nu} + \delta g_{\mu\nu}$, where $\bar{g}_{\mu\nu}$ is the metric of the background, (2.1). We will focus on the scalar perturbations within $\delta g_{\mu\nu}$. In the Newtonian gauge, the perturbed metric can be written as [8]:

$$\mathrm{d}s^2 = a^2(\eta)\left[(1+2\Psi)\mathrm{d}\eta^2 - (1-2\phi)\delta^{(\mathrm{K})}_{ij}\mathrm{d}x^i\mathrm{d}x^j\right] \tag{2.9}$$

where Ψ and ϕ are scalar perturbations and can be thought of as gravitational potentials.

The stress-energy tensor, in the perturbative regime, may be written as $T^\mu_\nu = \bar{T}^\mu_\nu + \delta T^\mu_\nu$, where $\bar{T}^\mu_\nu$ is its background component (2.3) and δT^μ_ν is given by:

$$\delta T^\mu_\nu = (\delta\rho + \delta P)\bar{U}^\mu\bar{U}_\nu + (\bar{\rho} + \bar{P})(\delta U^\mu \bar{U}_\nu + \bar{U}^\mu \delta U_\nu) - \delta P \delta^\mu_\nu - \Pi^\mu{}_\nu \tag{2.10}$$

where Π^μ_ν is the *anisotropic stress*. It is a good approximation to ignore it after matter-radiation equality, when neutrinos become unimportant [8]. Perturbations to the stress-energy tensor, therefore, resolve into density and pressure perturbations, $\delta\rho$ and δP.

Einstein's equations then connect $\delta g_{\mu\nu}$ to $\delta\rho$ and δP. The absence of anisotropic stress implies $\Psi = \phi$ in (2.9) [8]. As a standard result at linear order (see [8] for a derivation), we obtain the modified Poisson equation:

$$\nabla^2\phi - 3\frac{a'}{a}\left(\phi' + \phi\frac{a'}{a}\right) = 4\pi G a^2 \rho\delta \tag{2.11}$$

where we have defined the *density contrast*, or *overdensity*, as $\delta \equiv \delta\rho/\bar{\rho}$.

2.3 The Newtonian regime

We will now consider the dynamics of matter under suitable approximations. We seek, in particular, a solution for the matter overdensity $\delta(\boldsymbol{x}, \eta)$ and the matter velocity $\boldsymbol{v}(\boldsymbol{x}, \eta)$ beyond linear order.

At first order in perturbations, as we have seen, Einstein's equations yield (2.11), which describes metric perturbations in terms of the density contrast. This equation gives an accurate description for the field ϕ even when the overdensity δ, which acts as its source, becomes nonlinear. Indeed, a standard result of linear perturbation theory is that the gravitational potential remains constant (thus small) during matter domination, whereas density perturbations grow with the scale

factor. One thus finds that the potential ϕ remains small ($\sim 10^{-4}$) on all scales [17]. This will allow the linear equation (2.11) to be used in the theory to be developed below, whereas density and velocity perturbations will require a treatment at nonlinear order.

Equation (2.11) can be further simplified by noting that the scales under study are much smaller than the Hubble radius: $k \gg aH \sim 3 \cdot 10^{-4} h\,\mathrm{Mpc}^{-1}$. Thus the first term is dominant in (2.11), and the latter reduces to the Poisson equation of Newtonian gravity [17]:

$$\nabla^2 \phi = 4\pi G a^2 \bar{\rho} \delta \tag{2.12}$$

where, in matter domination, $\bar{\rho}$ and δ refer to the mean density and density contrast of matter, respectively. The equation above can also be written as:

$$\nabla^2 \phi = 4\pi G a^2 \bar{\rho}_t \Omega_m \delta \tag{2.13}$$

where $\bar{\rho}_t$ is the total mean density and $\Omega_m \equiv \bar{\rho}/\bar{\rho}_t$. Hence, by noting that the first Friedmann equation, (2.7), implies:

$$\bar{\rho}_t = \frac{3}{8\pi G} \frac{\mathcal{H}^2}{a^2} \tag{2.14}$$

the Poisson equation can be put in the convenient form:

$$\boxed{\nabla^2 \phi = \frac{3}{2} \Omega_m \delta \mathcal{H}^2} \tag{2.15}$$

Since all velocities we consider are non-relativistic, the geodesic equation of a matter particle reduces to the Newtonian law:

$$\begin{aligned} \ddot{\boldsymbol{r}} &= -\boldsymbol{\nabla}_r \phi \\ \boldsymbol{x}'' + \mathcal{H}\boldsymbol{x}' + \mathcal{H}' x &= -\boldsymbol{\nabla}\phi \end{aligned} \tag{2.16}$$

where, again, $\boldsymbol{r} \equiv a\boldsymbol{x}$. If we define the canonical momentum as:

$$\boldsymbol{p} = am\boldsymbol{x}' \tag{2.17}$$

the equation of motion of a matter particle can be written as:

$$\begin{aligned} \boldsymbol{p}' &= a' m \boldsymbol{x}' + am\boldsymbol{x}'' \\ &= -am\left(\frac{a'}{a}\boldsymbol{x}' + \boldsymbol{x}''\right) \end{aligned}$$

which, in view of (2.16), gives:

$$\boxed{\boldsymbol{p}' = -am\boldsymbol{\nabla}\phi} \tag{2.18}$$

2.4 Fluid equations

Equations (2.15) and (2.18) will be used in our derivation of the fluid equations. The fluid can be conveniently described in phase space by a distribution function $f(\eta, \boldsymbol{x}, \boldsymbol{p})$. For the case of collisionless dark matter, Liouville's theorem yields the collisionless Boltzmann equation, or Vlasov equation:

$$\frac{\mathrm{d}f}{\mathrm{d}\eta} = \frac{\partial f}{\partial \eta} + \frac{\boldsymbol{p}}{ma}\cdot\frac{\partial f}{\partial \boldsymbol{x}} - am\boldsymbol{\nabla}\phi\cdot\frac{\partial f}{\partial \boldsymbol{p}} = 0 \tag{2.19}$$

From the above, one can study the evolution of quantities such as density and bulk velocity at arbitrary order. These are obtained as moments of the distribution function. We thus define the density, velocity, and the stress tensor, respectively, as:

$$\rho(\boldsymbol{x}, \eta) = \frac{m}{a^3}\int \mathrm{d}^3p\, f(\boldsymbol{x}, \boldsymbol{p}, \eta) \tag{2.20}$$

$$v_i(\boldsymbol{x}, \eta) = \int \mathrm{d}^3p\, \frac{p_i}{am} f(\boldsymbol{x}, \boldsymbol{p}, \eta) \,/\, \int \mathrm{d}^3p\, f(\boldsymbol{x}, \boldsymbol{p}, m) \tag{2.21}$$

$$\sigma_{ij}(\boldsymbol{x}, \eta) = \int \mathrm{d}^3p\, \frac{p_i}{am}\frac{p_j}{am} f(\boldsymbol{x}, \boldsymbol{p}, \eta) \,/\, \int \mathrm{d}^3p\, f(\boldsymbol{x}, \boldsymbol{p}, m) - v_i(\boldsymbol{x})v_j(\boldsymbol{x}) \tag{2.22}$$

The zeroth moment of the Vlasov equation – that is, its integral over momenta – yields, when multiplied by m:

$$\begin{aligned} 0 &= m\frac{\partial}{\partial\eta}\int \mathrm{d}^3p\, f + m\int \mathrm{d}^3p\, \frac{p^i}{ma}\frac{\partial f}{\partial x^i} - am^2\int \mathrm{d}^3p\, \frac{\partial\phi}{\partial x^i}\frac{\partial f}{\partial p_i} \\ &= \frac{\partial}{\partial\eta}(a^3\rho) + a^3\rho v^i - am^2\int \mathrm{d}^3p\, \frac{\partial}{\partial p_i}\left(\frac{\partial\phi}{\partial x^i} f\right) + am^2\int \mathrm{d}^3p\, f\frac{\partial\phi}{\partial x^i \partial p_i} \end{aligned}$$

In the line above, the fourth term vanishes because the potential ϕ doesn't depend on the momentum. The third term can be expressed as an integral over a boundary, which also vanishes because $f(\boldsymbol{x}, \boldsymbol{p}, \eta)$ is zero at infinite p. Hence we

have:

$$\begin{aligned} 0 &= \frac{\partial}{\partial\eta}(a^3\rho) + m\frac{\partial}{\partial x^i}\left(\frac{a^3}{m}\rho v^i\right) \\ &= 3a^2a'\rho + a^3\rho' + a^3\frac{\partial}{\partial x^i}(\rho v^i) \\ &= 3\mathcal{H}\rho + \rho' + \frac{\partial}{\partial x^i}(\rho v^i) \end{aligned}$$

Writing $\rho = (1+\delta)\bar{\rho}$, we obtain $\rho' = \delta'\bar{\rho} + (1+\delta)\bar{\rho}'$, so that:

$$\begin{aligned} 0 &= 3\mathcal{H}(1+\delta)\bar{\rho} + \delta'\bar{\rho} + (1+\delta)\bar{\rho}' + \frac{\partial}{\partial x^i}[\bar{\rho}(1+\delta)v^i] \\ &= 3\mathcal{H}(1+\delta) + \delta' + (1+\delta)\frac{\bar{\rho}'}{\bar{\rho}} + \frac{\partial}{\partial x^i}[v^i(1+\delta)] \end{aligned}$$

Mass conservation implies $\bar{\rho} \propto a^{-3}$, so $\bar{\rho}'/\rho = -3\mathcal{H}$. Cancellation between the first and the third terms above yields, finally, the exact continuity equation:

$$\delta' + \frac{\partial}{\partial x^i}[v^i(1+\delta)] = 0$$

or:

$$\boxed{\delta' + \boldsymbol{\nabla}\cdot[(1+\delta)\boldsymbol{v}] = 0} \tag{2.23}$$

Furthermore, the first moment of (2.19) – i.e., its momentum-average – gives:

$$\begin{aligned} 0 &= \int \mathrm{d}^3p\, p^i\frac{\partial f}{\partial\eta} + \int \mathrm{d}^3p\, p^i\frac{p^j}{ma}\frac{\partial f}{\partial x^j} - am\int \mathrm{d}^3p\, p^i\frac{\partial\phi}{\partial x^j}\frac{\partial f}{\partial p_j} \\ &= \frac{\partial}{\partial\eta}\left(am\,v^i\int \mathrm{d}^3p\, f\right) + \frac{\partial}{\partial x^j}\left[a^4\rho(\sigma^{ij} + v^iv^j)\right] - am\int \mathrm{d}^3p\,\frac{\partial}{\partial p^j}\left(p^i\frac{\partial\phi}{\partial x^j}f\right) \\ &+ am\int \mathrm{d}^3p\, f\frac{\partial}{\partial p^j}\left(p^i\frac{\partial\phi}{\partial x^j}\right) \end{aligned}$$

where the third integral in the first line has been computed by parts. As before, the third term in the last line generates a vanishing boundary component. We are left with:

$$\begin{aligned} 0 &= \frac{\partial}{\partial\eta}(a^4\rho v^i) + \frac{\partial}{\partial x^j}\left[a^4\rho(\sigma^{ij} + v^iv^j)\right] + am\int \mathrm{d}^3p\, f\frac{\partial\phi}{\partial x_i} \\ &= \frac{\partial}{\partial\eta}(a^4\rho v^i) + \frac{\partial}{\partial x^j}\left[a^4\rho(\sigma^{ij} + v^iv^j)\right] + a^4\rho\frac{\partial\phi}{\partial x_i} \end{aligned} \tag{2.24}$$

The first term above can be written as:

$$\begin{aligned}\frac{\partial}{\partial\eta}(a^4\rho v^i) &= a^4(\rho v^i)' + 4(\rho v^i)a^3a' \\ &= a^4\bar{\rho}'(1+\delta)v^i + a^4\bar{\rho}\delta' v^i + a^4\bar{\rho}(1+\delta)v'^i + 4\bar{\rho}(1+\delta)v^i a^3 a' \\ &= a^4\bar{\rho}\left[\frac{\bar{\rho}'}{\bar{\rho}}(1+\delta)v^i + \delta' v^i + (1+\delta)v'^i + \frac{4a'}{a}(1+\delta)v^i\right]\end{aligned}$$

Once again, we use $\bar{\rho}'/\bar{\rho} = -3\mathcal{H}$, so that:

$$\frac{\partial}{\partial\eta}(a^4\rho v^i) = a^4\bar{\rho}\left[\mathcal{H}(1+\delta)v^i + \delta' v^i + (1+\delta)v'^i\right]$$

Replacing the above in (2.24), we obtain:

$$\begin{aligned}0 &= \mathcal{H}(1+\delta)v^i + \delta' v^i + (1+\delta)v'^i \\ &\quad + \frac{\partial}{\partial x^j}\left[(1+\delta)(\sigma^{ij} + v^i v^j)\right] + (1+\delta)\frac{\partial\phi}{\partial x_i}\end{aligned} \tag{2.25}$$

The continuity equation, (2.23), implies:

$$\delta' v^i = -v^i\frac{\partial}{\partial x^j}[(1+\delta)v^j] \tag{2.26}$$

Hence, (2.25) can be written as:

$$\begin{aligned}0 &= \mathcal{H}(1+\delta)v^i + (1+\delta)v^j\frac{\partial v^i}{\partial x^j} + (1+\delta)v'^i + \frac{\partial}{\partial x^j}\left[(1+\delta)\sigma^{ij}\right] \\ &\quad + (1+\delta)\frac{\partial\phi}{\partial x_i} \\ &= \mathcal{H}v^i + v^j\frac{\partial v^i}{\partial x^j} + v'^i + (1+\delta)^{-1}\frac{\partial}{\partial x^j}\left[(1+\delta)\sigma^{ij}\right] + \frac{\partial\phi}{\partial x_i} \\ &= \mathcal{H}v^i + v^j\frac{\partial v^i}{\partial x^j} + v'^i + [\bar{\rho}(1+\delta)]^{-1}\frac{\partial}{\partial x^j}\left[\bar{\rho}(1+\delta)\sigma^{ij}\right] + \frac{\partial\phi}{\partial x_i}\end{aligned} \tag{2.27}$$

We have thus found the exact Euler equation:

$$\boxed{v_i' + \mathcal{H}v_i + \boldsymbol{v}\cdot\boldsymbol{\nabla}v_i = -\nabla_i\phi - \frac{1}{\rho}\nabla^j(\rho\sigma_{ij})} \tag{2.28}$$

In standard perturbation theory (SPT), one makes the assumption of a "pressureless perfect fluid", which implies $\sigma_{ij} = 0$. This involves discounting the effect

of non-gravitational forces on baryonic matter[1]. As shown in hydrodynamic simulations [23], it begins to cause percent-level deviations on the matter power spectrum only at reasonably small scales, $k \sim 0.2h\,\mathrm{Mph}^{-1}$. (The case of an imperfect fluid will be considered in chapter 3.) With the assumption $\sigma_{ij} = 0$, equations (2.23) and (2.28), together with the Poisson equation (2.15), form a closed system.

Before proceeding to the full perturbative treatment, we will seek an understanding of the solutions' behavior in the linear regime. We thus decompose $\boldsymbol{v}$ into scalar and vector parts, $\boldsymbol{v} = \boldsymbol{v}_{\parallel} + \boldsymbol{v}_{\perp}$, with $\boldsymbol{\nabla} \times \boldsymbol{v}_{\parallel} = 0$ and $\boldsymbol{\nabla} \cdot \boldsymbol{v}_{\perp} = 0$, and define the field's divergence $\theta \equiv \boldsymbol{\nabla} \cdot \boldsymbol{v}$ and vorticity $\boldsymbol{w} \equiv \boldsymbol{\nabla} \times \boldsymbol{v}$.

It can be shown that the Euler equation (2.28), together with the assumption of a vanishing stress-tensor, implies that gravitational evolution creates no vorticity ($\boldsymbol{w} = 0$) provided none is initially present [5]. In the linear regime, the Euler and the continuity equations can be written as:

$$\delta' + \theta = 0 \tag{2.29}$$

$$\theta' + \mathcal{H}\theta = -\nabla^2\phi \tag{2.30}$$

Together with the Poisson equation, they finally imply:

$$\delta'' + \mathcal{H}\delta' - \frac{3}{2}\mathcal{H}^2\Omega_m(a)\delta = 0 \tag{2.31}$$

In the matter-dominated regime, this has the growing-mode solution:

$$\delta(\boldsymbol{k}, \eta) = D(\eta)\delta_0(\boldsymbol{k}) \tag{2.32}$$

with:

$$D(\eta) = D_0\mathcal{H}(\eta)\int_0^{a(\eta)} \frac{\mathrm{d}a'}{\mathcal{H}^3(a')} \tag{2.33}$$

In an Einstein-de Sitter universe, where $\Omega_m = 1$, one finds $D(\eta) = a(\eta)$ during matter domination.

As for the velocity field, the same pair, (2.29) and (2.30), implies:

$$\theta(\boldsymbol{k}, \eta) = -\mathcal{H}f\delta(\boldsymbol{k}, \eta) \tag{2.34}$$

[1]Gravitational interactions also cause the appearance of a non-vanishing σ_{ij} in the *coarse-grained* fluid equations, as will be seen in chapter 3.

where the linear-growth factor f is defined as:

$$f = \frac{\mathrm{d}\ln D}{\mathrm{d}\ln a} \tag{2.35}$$

2.5 Solutions in standard perturbation theory

We will now proceed to the full perturbative solution of the fluid equations (2.23) and (2.28). For convenience, one may isolate nonlinear terms on the right-hand side and express them in Cartesian-component form. The continuity equation can thus be rewritten as:

$$\delta' + \theta = -v^i \frac{\partial}{\partial x^i}\delta + \delta\theta \tag{2.36}$$

where we have used again $\theta \equiv \boldsymbol{\nabla}\cdot\boldsymbol{v}$. Assuming $\sigma_{ij} = 0$ and taking the divergence of the Euler equation, we obtain:

$$\theta' + \mathcal{H}\theta + \nabla^2\phi = -v^j \frac{\partial}{\partial x^j}\theta - (\partial_i v^j)(\partial_j v^i) \tag{2.37}$$

It is the pair (2.36) and (2.37) that we wish to solve for $\delta(\boldsymbol{x},\eta)$ and $\theta(\boldsymbol{x},\eta)$. When translated to Fourier space, the nonlinear products in the right-hand side are transformed to convolutions. We adopt the Fourier conventions:

$$\tilde{f}(\boldsymbol{k}) = \mathcal{F}\{f\}(\boldsymbol{k}) = \int \mathrm{d}x^3\, f(\boldsymbol{x}) e^{-i\boldsymbol{k}\cdot\boldsymbol{x}}, \quad f(\boldsymbol{x}) = \int \frac{\mathrm{d}k^3}{(2\pi)^3}\, \tilde{f}(\boldsymbol{k}) e^{i\boldsymbol{k}\cdot\boldsymbol{x}} \tag{2.38}$$

and the tilde in the notation of the Fourier-space function will be suppressed when no ambiguity is implied. So one has, for example:

$$\begin{aligned}
\mathcal{F}\left\{v^i\frac{\partial}{\partial x^i}\delta\right\}(\boldsymbol{k}) &= \int \mathrm{d}^3x\, e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} \int \frac{\mathrm{d}^3q}{(2\pi)^3}\, v^i(\boldsymbol{q}) e^{i\boldsymbol{q}\cdot\boldsymbol{x}} \frac{\partial}{\partial x^i} \int \frac{\mathrm{d}^3q'}{(2\pi)^3}\, \delta(\boldsymbol{q}') e^{i\boldsymbol{q}'\cdot\boldsymbol{x}} \\
&= \int \mathrm{d}^3x\, e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} \int \frac{\mathrm{d}^3q}{(2\pi)^3}\, v^i(\boldsymbol{q}) e^{i\boldsymbol{q}\cdot\boldsymbol{x}} \int \frac{\mathrm{d}^3q'}{(2\pi)^3}\, \delta(\boldsymbol{q}')(iq'_i)\, e^{i\boldsymbol{q}'\cdot\boldsymbol{x}} \\
&= \int \frac{\mathrm{d}^3q}{(2\pi)^3} \int \frac{\mathrm{d}^3q'}{(2\pi)^3} v^i(\boldsymbol{q})\delta(\boldsymbol{q}')(iq'_i) \int \mathrm{d}^3x\, e^{-i\boldsymbol{x}\cdot(\boldsymbol{k}-\boldsymbol{q}-\boldsymbol{q}')} \\
&= \int \frac{\mathrm{d}^3q}{(2\pi)^3} \int \frac{\mathrm{d}^3q'}{(2\pi)^3} \left[-i\frac{q^i}{q^2}\theta(\boldsymbol{q})\right] \delta(\boldsymbol{q}')(iq'_i) \times \\
&\quad \left[(2\pi)^3\delta^{(\mathrm{D})}(\boldsymbol{k}-\boldsymbol{q}-\boldsymbol{q}')\right] \\
&= \int \frac{\mathrm{d}^3q}{(2\pi)^3} \int \frac{\mathrm{d}^3q'}{(2\pi)^3} \left[(2\pi)^3\delta^{(\mathrm{D})}(\boldsymbol{k}-\boldsymbol{q}-\boldsymbol{q}')\right] \left(\frac{\boldsymbol{q}\cdot\boldsymbol{q}'}{q^2}\right) \theta(\boldsymbol{q})\delta(\boldsymbol{q}')
\end{aligned} \tag{2.39}$$

where we have used $v(\boldsymbol{k}) = -i(\boldsymbol{k}/k^2)\theta(\boldsymbol{k})$[2]. The equations (2.36) and (2.37) in Fourier space can thus be expressed as:

$$\begin{aligned}
\delta'(\boldsymbol{k}) + \theta(\boldsymbol{k}) &= -\int \frac{\mathrm{d}^3q}{(2\pi)^3} \int \frac{\mathrm{d}^3q'}{(2\pi)^3} \left[(2\pi)^3\delta^{(\mathrm{D})}(\boldsymbol{k}-\boldsymbol{q}-\boldsymbol{q}')\right] \\
&\quad \times \alpha(\boldsymbol{q},\boldsymbol{q}')\theta(\boldsymbol{q})\delta(\boldsymbol{q}') \\
\theta'(\boldsymbol{k}) + \mathcal{H}\theta(\boldsymbol{k}) + \frac{3}{2}\Omega_m(a)\mathcal{H}^2\delta(\boldsymbol{k}) &= -\int \frac{\mathrm{d}^3q}{(2\pi)^3} \int \frac{\mathrm{d}^3q'}{(2\pi)^3} \left[(2\pi)^3\delta^{(\mathrm{D})}(\boldsymbol{k}-\boldsymbol{q}-\boldsymbol{q}')\right] \\
&\quad \times \beta(\boldsymbol{q},\boldsymbol{q}')\theta(\boldsymbol{q})\delta(\boldsymbol{q}')
\end{aligned} \tag{2.40}$$

with kernels $\alpha(\boldsymbol{k}_1,\boldsymbol{k}_2)$ and $\beta(\boldsymbol{k}_1,\boldsymbol{k}_2)$ defined as:

$$\alpha(\boldsymbol{k}_1,\boldsymbol{k}_2) = \frac{\boldsymbol{k}_1\cdot(\boldsymbol{k}_1+\boldsymbol{k}_2)}{k_1^2} \tag{2.41}$$

$$\beta(\boldsymbol{k}_1,\boldsymbol{k}_2) = \frac{1}{2}(\boldsymbol{k}_1+\boldsymbol{k}_2)^2\frac{\boldsymbol{k}_1\cdot\boldsymbol{k}_2}{k_1^2k_2^2} = \frac{1}{2}\frac{\boldsymbol{k}_1\cdot\boldsymbol{k}_2}{k_1k_2}\left(\frac{k_2}{k_1}+\frac{k_1}{k_2}\right) + \frac{(\boldsymbol{k}_1\cdot\boldsymbol{k}_2)^2}{k_1^2k_2^2} \tag{2.42}$$

The kernel $\beta(\boldsymbol{k}_1,\boldsymbol{k}_2)$ has been symmetrized in its arguments.

It is now the pair (2.40) that we wish to solve for $\delta(\boldsymbol{k},\eta)$ and $\theta(\boldsymbol{k},\eta)$, thus providing the Fourier decomposition of the fields of interest. That system of equations can be solved iteratively by the procedure [17]:

[2]Note that, indeed, $\mathcal{F}\{\boldsymbol{\nabla}\cdot\boldsymbol{v}\} = i\boldsymbol{k}\cdot v(\boldsymbol{k}) = i\boldsymbol{k}\cdot[-i(\boldsymbol{k}/k^2)\theta(\boldsymbol{k})] = \theta(\boldsymbol{k})$.

$$
\begin{aligned}
(\delta^{(n)})'(\boldsymbol{k}) + (\theta^{(n)})(\boldsymbol{k}) &= -\int_{\boldsymbol{q},\boldsymbol{q}'} (2\pi)^3 \delta^{(\mathrm{D})}(\boldsymbol{k}-\boldsymbol{q}-\boldsymbol{q}') \times \\
&\qquad \alpha(\boldsymbol{q},\boldsymbol{q}')(\theta^{(n-1)})(\boldsymbol{q})(\delta^{(n-1)})(\boldsymbol{q}') \\
(\theta^{(n)})'(\boldsymbol{k}) + \mathcal{H}(\theta^{(n)})(\boldsymbol{k}) + \frac{3}{2}\Omega_m(a)\mathcal{H}^2(\delta^{(n)})(\boldsymbol{k}) &= -\int_{\boldsymbol{q},\boldsymbol{q}'} (2\pi)^3 \delta^{(\mathrm{D})}(\boldsymbol{k}-\boldsymbol{q}-\boldsymbol{q}') \times \\
&\qquad \beta(\boldsymbol{q},\boldsymbol{q}')(\theta^{(n-1)})(\boldsymbol{q})(\delta^{(n-1)})(\boldsymbol{q}')
\end{aligned}
\tag{2.43}
$$

That is, one invokes the solutions at nth order, $\delta^{(n)}$ and $\theta^{(n)}$, as source terms for the equations of order $n+1$. The complete solutions are then obtained as[3]:

$$
\begin{aligned}
\delta(\boldsymbol{k},\eta) &= \sum_{n=1}^{\infty} \delta^{(n)}(\boldsymbol{k},\eta) \\
\theta(\boldsymbol{k},\eta) &= \sum_{n=1}^{\infty} \theta^{(n)}(\boldsymbol{k},\eta)
\end{aligned}
\tag{2.44}
$$

The solutions at first order are obtained by setting the right-hand side of (2.43) to zero, thus yielding the fields already found under the linear approximation:

$$
\begin{aligned}
\delta^{(1)}(\boldsymbol{k},\eta) &= D(\eta)\delta_0(\boldsymbol{k}) \\
\theta^{(1)}(\boldsymbol{k},\eta) &= -\mathcal{H}fD(\eta)\delta_0(\boldsymbol{k})
\end{aligned}
\tag{2.45}
$$

By analogy with the functional forms above, we will suppose that the functions at every order can be separated into spatial and time-dependent parts, as $\delta^{(n)}(\boldsymbol{k},\eta) = \tilde{\delta}^{(n)}(\boldsymbol{k})D^n(\eta)$. This ansatz will prove exact only for an EdS model, for which $D(\eta) = a(\eta)$, but is still a good approximation in ΛCDM – the agreement of $\delta^{(2)}$ and $\theta^{(2)}$, for example, is at the sub-percent level [17]. Hence:

$$
\begin{aligned}
\delta(\boldsymbol{k},\eta) &= \sum_{n=1}^{\infty} D^n(\eta)\tilde{\delta}^{(n)}(\boldsymbol{k}) \\
\theta(\boldsymbol{k},\eta) &= -\mathcal{H}(\eta)\sum_{i=n}^{\infty} D^n(\eta)\tilde{\theta}^{(n)}(\boldsymbol{k})
\end{aligned}
\tag{2.46}
$$

By substitution of (2.46) into (2.43), one may construct the solutions at all orders recursively. In general, we have:

[3] This is known as the *Neumann series* method for the solution of integro-differential equations [4].

$$\tilde{\delta}^{(n)}(\boldsymbol{k}) = \prod_{m=1}^{n} \left\{ \int \frac{d^3 q_m}{(2\pi)^3} \delta^{(1)}(\boldsymbol{q}_m) \right\} F_n(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n)(2\pi)^3 \delta^{(D)}(\boldsymbol{k} - \boldsymbol{q}|_1^n) \tag{2.47}$$

$$\tilde{\theta}^{(n)}(\boldsymbol{k}) = \prod_{m=1}^{n} \left\{ \int \frac{d^3 q_m}{(2\pi)^3} \delta^{(1)}(\boldsymbol{q}_m) \right\} G_n(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n)(2\pi)^3 \delta^{(D)}(\boldsymbol{k} - \boldsymbol{q}|_1^n) \tag{2.48}$$

where $\boldsymbol{q}|_i^j \equiv \sum_{m=1}^{j} \boldsymbol{q}_l$. The convolution kernels are defined iteratively as:

$$\begin{aligned} F_n(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_m) = \sum_{m=1}^{n-1} \frac{G_m(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_m)}{(2n+3)(n-1)} \big[& (2n+1)\alpha(\boldsymbol{q}|_1^m, \boldsymbol{q}|_{m+1}^n) F_{n-m}(\boldsymbol{q}_{m+1}, \ldots, \boldsymbol{q}_n) \\ & + 2\beta(\boldsymbol{q}|_1^m, \boldsymbol{q}|_{m+1}^n) G_{n-m}(\boldsymbol{q}_{m+1}, \ldots, \boldsymbol{q}_n) \big] \end{aligned} \tag{2.49}$$

$$\begin{aligned} G_n(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_m) = \sum_{m=1}^{n-1} \frac{G_m(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_m)}{(2n+3)(n-1)} \big[& 3\alpha(\boldsymbol{q}|_1^m, \boldsymbol{q}|_{m+1}^n) F_{n-m}(\boldsymbol{q}_{m+1}, \ldots, \boldsymbol{q}_n) \\ & + 2n\beta(\boldsymbol{q}|_1^m, \boldsymbol{q}|_{m+1}^n) G_{n-m}(\boldsymbol{q}_{m+1}, \ldots, \boldsymbol{q}_n) \big] \end{aligned} \tag{2.50}$$

with $F_1 = G_1 \equiv 1$, so that the linear solutions (2.45) are recovered.

Having arrived at $\delta(\boldsymbol{k}, \eta)$, we may use it to calculate the power spectrum $P(k)$. This is defined through:

$$\langle \delta(\boldsymbol{k}, \eta) \delta(\boldsymbol{k}', \eta) \rangle = (2\pi)^3 \delta^{(\mathrm{D})}(\boldsymbol{k} + \boldsymbol{k}') P(k, \eta, \eta') \tag{2.51}$$

where $P(k, \eta, \eta')$ depends only on the modulus $k = |\boldsymbol{k}|$ by statistical isotropy. We thus compute the correlator [5]:

$$\begin{aligned} \langle \delta(\boldsymbol{k}, \eta) \delta(\boldsymbol{k}', \eta) \rangle &= \sum_{i,j} D^{i+j}(\eta) \langle \tilde{\delta}^{(i)}(\boldsymbol{k}) \tilde{\delta}^{(j)}(\boldsymbol{k}') \rangle \\ &\sim \sum_{i,j} D^{i+j}(\eta) \int F_i F_j \langle \tilde{\delta}^{(1)}(\boldsymbol{q}_1) \ldots \tilde{\delta}^{(1)}(\boldsymbol{q}_i) \tilde{\delta}^{(1)}(\boldsymbol{q}'_1) \ldots \tilde{\delta}^{(1)}(\boldsymbol{q}'_j) \rangle \end{aligned} \tag{2.52}$$

Since $\tilde{\delta}(\boldsymbol{k})$ is a Gaussian field, Wick's theorem implies that only products of an even number of terms will be non-zero [17]. At next-to-leading order, the correlator is then:

$$\langle\delta(\boldsymbol{k})\delta(\boldsymbol{k}')\rangle = \langle\delta^{(1)}(\boldsymbol{k})\delta^{(1)}(\boldsymbol{k}')\rangle + 2\langle\delta^{(1)}(\boldsymbol{k})\delta^{(3)}(\boldsymbol{k}')\rangle + \langle\delta^{(2)}(\boldsymbol{k})\delta^{(2)}(\boldsymbol{k}')\rangle \tag{2.53}$$

and the next-to-leading order, or *one-loop*, power spectrum is:

$$P_{1-\text{loop}} = P_{\text{lin}}(\boldsymbol{k}) + 2P_{13}(\boldsymbol{k}) + P_{22}(\boldsymbol{k}) \tag{2.54}$$

The components $P_{13}(k)$ and $P_{22}(k)$ are given explicitly in terms of the linear power spectrum and the convolution kernels as:

$$P_{13}(k) = 3P_{\text{lin}}(k)\int\frac{\mathrm{d}^3q}{(2\pi)^3}P_{\text{lin}}(q)F_3(\boldsymbol{k},\boldsymbol{q},-\boldsymbol{q}) \tag{2.55}$$

$$P_{22}(k) = 2\int\frac{\mathrm{d}^3q}{(2\pi)^3}P_{\text{lin}}(q)P_{\text{lin}}(|\boldsymbol{k}-\boldsymbol{q}|)|F_2(\boldsymbol{q},\boldsymbol{k},-\boldsymbol{q})|^2 \tag{2.56}$$

2.6 Problems with the standard approach

A comparison of the power spectrum obtained in standard perturbation theory with that found in simulations can be viewed in figure 2.1. (Analytical results have been computed with the Boltzmann code CLASS-PT [14].) It is clear that the standard approach fails at reasonably-large scales, near $k = 0.1h\,\text{Mpc}^{-1}$, where it was expected to work well. Moreover, inclusion of higher-loop corrections doesn't seem to make the result closer to the expected answer.

An explanation for this behavior can be found in the magnitude of $\delta(\boldsymbol{k})$, which must be small ($\delta \ll 1$) to function as an expansion parameter. Its typical size may be computed from its variance, and one finds that it crosses unity at $k_{NL} \sim 0.3h\,\text{Mpc}^{-1}$ [17]. Standard perturbation theory isn't expected to work near that point. Moreover, we took it as a starting assumption that the stress-tensor vanishes, $\sigma_{ij} = 0$, but this is known to be false on small scales. Indeed, this assumption had effect on the right-hand side of the Euler equation, (2.28), which in turn led to the solutions (2.40).

Chapter 3 presents how these problems can be avoided in an effective field theory of structure growth. The corrections to SPT are based on a modification of the expansion parameter and on the use of the Euler equation for the imperfect fluid.

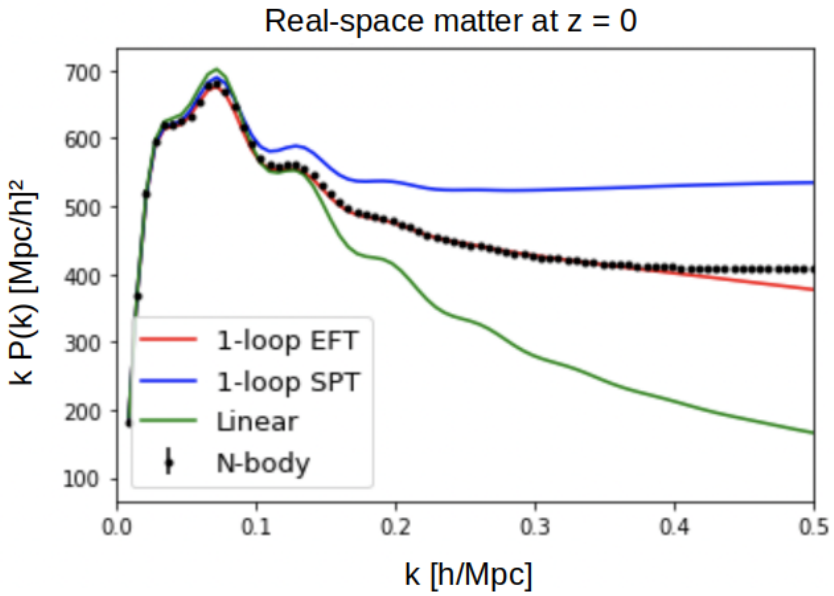

Figure 2.1: Matter power spectra obtained by various methods. The EFT spectrum is computed with the technique discussed in chapter 3. It shows good agreement with simulations down to scales $k \sim 0.3\,h$Mpc. The numerical results were obtained from the Quijote simulations [29] at redshift zero, whereas analytical spectra were computed with CLASS-PT [14]. Source: [24].

Chapter 3

An effective field theory of structure growth

We here consider a solution to the problem encountered at the end of the previous chapter. The matter overdensity can be made a suitable expansion parameter by a procedure that filters out small-scale fluctuations, the behavior of which one is not interested in. The "long-wavelength" overdensity, δ_l, is then guaranteed to be small on scales as small as $k \sim 0.3h\,\mathrm{Mpc}^{-1}$, which extends the validity of perturbation theory.

The aim of this chapter is to construct $\delta_l(\boldsymbol{k})$ and determine its evolution. The point of departure will be again the fluid equations in the Newtonian regime. One will find, however, that the regularization procedure introduces a viscosity term in the fluid equations – i.e., the matter fluid can no longer be considered ideal. Viscosity expresses the effect of small scales on larger-scale dynamics. Since it can't be predicted from the effective theory, it has to be fitted to the data.

We then establish the analogous theory in the Lagrangian frame – i.e., in terms of the initial coordinates of the fluid elements and their later displacement. This often allows an expansion in terms of the overdensity field in the initial positions, in which perturbations are smaller and more amenable to analytic treatment.

Lagrangian perturbation theory is well suited to the treatment of galaxy bias. We study the statistical relation between galaxy and matter overdensities by means of a simple model based on spherical collapse. This will allow the establishment of a "galaxy-bias expansion", i.e., the expression of the galaxy overdensity in terms of operators of the matter overdensity. We will then be prepared for the applications to be discussed in chapter 4.

3.1 Coarse-grained equations of motion

To establish a proper expansion parameter, we will smooth the field over a scale $1/\Lambda$, with $\Lambda \ll k_{NL}$. We expect the perturbative approach to work well in that regime. The long-wavelength component of a function $g(\boldsymbol{x})$ is defined through the convolution:

$$g_l(\boldsymbol{x}) \equiv [g]_\Lambda(\boldsymbol{x}) = \int \mathrm{d}^3\boldsymbol{x}'\, W_\Lambda(\boldsymbol{x}-\boldsymbol{x}')g(\boldsymbol{x}) \tag{3.1}$$

where the filtering function $W_\Lambda(\boldsymbol{x})$ is chosen as to be negligible for $x \gg 1/\Lambda$. The convolution above thus amounts to taking the average of $g(\boldsymbol{x})$ over a region of size $\sim 1/\Lambda$. A convenient choice for $W_\Lambda(\boldsymbol{x})$ is the Gaussian function:

$$W_\Lambda(\boldsymbol{x}) = \left(\frac{\Lambda}{\sqrt{2\pi}}\right)^3 \exp\left[-\frac{1}{2}\Lambda^2 x^2\right], \quad W_\Lambda(\boldsymbol{k}) = \exp\left[-\frac{1}{2}\frac{k^2}{\Lambda^2}\right] \tag{3.2}$$

As argued in chapter 2, conservation of density in phase space implies the Vlasov equation, (2.19), from which we derived the continuity and Euler equations:

$$\delta' + \boldsymbol{\nabla}\cdot[(1+\delta)\boldsymbol{v}] = 0 \tag{2.23}$$

$$v_i' + \mathcal{H}v_i + \boldsymbol{v}\cdot\boldsymbol{\nabla}v_i + \nabla_i\phi = 0 \tag{2.28}$$

We seek the analogous equations in terms of the long-wavelength variables δ_l and v_l. For that purpose, we define the long-wavelength density and momentum as:

$$\rho_l = \int \mathrm{d}^3 p f_l(\boldsymbol{x},\boldsymbol{p},\eta) \tag{3.3}$$

$$\boldsymbol{\pi}_l = \rho_l \boldsymbol{v}_l = \frac{m}{a^3}\int \mathrm{d}^3 p \frac{\boldsymbol{p}}{ma} f_l(\boldsymbol{x},\boldsymbol{p},\eta) \tag{3.4}$$

from which we define $\delta_l \equiv \rho_l/\bar{\rho} - 1$. One should note that the long-wavelength velocity is defined as $v_l \equiv \boldsymbol{\pi}_l/\rho_l$. That is, it is the particle's momentum, not its velocity, which is smoothed over.

By applying the filtering function to (2.23), one obtains the so-called "coarse-grained" continuity equation, which has the same form as its original counterpart:

$$\delta_l' + \partial_j\left[(1+\delta_l)v_{l,j}\right] = 0 \tag{3.5}$$

For the Euler equation, however, product terms cause the appearance of an additional component dependent on the small-scale variables v_s and ϕ_s, even in the equation for large-scale fields. The fundamental reason for this behavior is that filtered products receive corrections[1]:

$$[fg]_\Lambda = f_l g_l + [f_s g_s]_\Lambda + \frac{1}{\Lambda^2}\nabla f_l \nabla g_l + \dots \tag{3.6}$$

where the ellipsis denotes higher-derivative contributions [5]. By splitting the velocity v and the potential ϕ into their long- and short-wavelength parts, and considering the Taylor expansion of the former in the neighboring region of size $\sim 1/\Lambda$, Baumann et al [9] have been able to show that the coarse-grained Euler equation can be written as:

$$v'_{l,i} + \mathcal{H} v_{l,i} + \partial_i \phi_l + v_{l,j}\partial_j v_{l,i} = -\frac{1}{\rho_l}\partial_j [\tau_{ij}]_\Lambda \tag{3.7}$$

One notices the appearance of an effective stress-tensor $[\tau_{ij}]$ even if the original equations were supposed to represent an ideal fluid. Up to higher-derivative corrections, that tensor can be written as [9]:

$$\tau_{ij} = \rho\sigma_{ij} + \rho v_i^s v_j^s - \frac{\phi^s_{,k}\phi^s_{,k}\delta^{(K)}_{ij} - 2\phi^s_{,i}\phi^s_{,j}}{8\pi G} \tag{3.8}$$

where $\phi^s_{,i} \equiv \partial\phi^s/\partial x^i$. Equations (3.5) and (3.7) describe large-scale matter as an imperfect fluid, its "viscosity" arising from small-scale dynamics.

The tensor in (3.8) still depends on short-wavelength behavior, which we can't expect to treat under perturbation theory. We circumvent that problem by, first, replacing τ_{ij} by its expectation value. We allow $\langle\tau_{ij}\rangle$ at a point $\boldsymbol{x}$ to depend on the value of long-wavelength fields $\delta_l(\boldsymbol{x})$ and $v_l(\boldsymbol{x})$, and expand it in a Taylor series written (schematically) as [11]:

$$\langle [\tau^{ij}]_\Lambda \rangle_{\delta_l} = \langle [\tau^{ij}]_\Lambda \rangle_0 + \left.\frac{\partial \langle [\tau^{ij}]_\Lambda \rangle_{\delta_l}}{\partial \delta_l}\right|_0 \delta_l + \dots \tag{3.9}$$

The advantage of such a procedure is that the complicated dependence of τ^{ij} upon small-scale dynamics is hereby captured by parameters of the sort $\partial\langle[\tau^{ij}]_\Lambda\rangle_{\delta_l}/\partial\delta_l|_0$, which we can measure from observations or simulated data. Moreover, a Taylor expansion on δ_l and v_l is allowed because those variables are

[1] The product appearing in the continuity equation, (2.23), $v\delta$, doesn't generate such corrections because the long-wavelength velocity is defined through $v_l \equiv (\rho v)_l/\rho_l$. The coarse-graining of the term $v\delta$ thus gives exactly $v_l\delta_l$.

guaranteed to be small on large scales ($\Lambda \ll k_{\rm NL}$).

At linear order, one can write the effective tensor as [11]:

$$\begin{aligned}[\tau_{ij}]_\Lambda = p\delta^{(\rm K)}_{ij} + \bar\rho\tilde c^2_{\rm s}\delta^{(\rm K)}_{ij}\delta_l - \bar\rho\frac{\tilde c^2_{\rm v,b}}{\mathcal H}\delta^{(\rm K)}_{ij}\partial_m v_{l,m} \\ - \frac{3}{4}\bar\rho\frac{\tilde c^2_{\rm v,s}}{\mathcal H}\left[\partial_i v_{l,j} + \partial_j v_{l,i} - \frac{2}{3}\delta^{(\rm K)}_{ij}\partial_m v_{l,m}\right] + \Delta\tau_{ij}\end{aligned} \tag{3.10}$$

The remainder $\Delta\tau_{ij}$, the "stochastic term", is that part of τ^{ij} which deviates from its expectation value and is uncorrelated with long modes. It can be shown to be much smaller than the other terms in (3.10) and thus safely neglected [11]. In the above, p is the background pressure, $\tilde c_{\rm s}$ is the speed of sound[2], and $\tilde c_{v,b}$ and $\tilde c_{v,s}$ are "bulk" and "shear" viscosity parameters, named in analogy with the corresponding terms in the Navier-Stokes equation for imperfect fluids.

An equation for the divergence velocity field can be obtained by taking the divergence of (3.7). Its source term, which appears only in the effective theory, will be proportional to $\partial^i\partial^j[\tau_{ij}]_\Lambda \equiv \tau_\theta$. This is given by [5]:

$$\begin{aligned}\tau_\theta &= \bar\rho\left[\tilde c^2_{\rm s}\partial^2\delta_l - \frac{\tilde c^2_{\rm v,b}}{\mathcal H}\partial^2\theta_l - \frac{3}{4}\frac{\tilde c^2_{\rm v,s}}{\mathcal H}\partial^2\theta_l\right] + \Delta J \\ &= \bar\rho\left[\tilde c^2_{\rm s}\partial^2\delta_l - \frac{\tilde c^2_{\rm v}}{\mathcal H}\partial^2\theta_l\right] + \Delta J\end{aligned} \tag{3.11}$$

where $\Delta J \equiv \partial^i\partial^j\Delta\tau_{ij}$ and $\tilde c^2_{\rm v} \equiv \tilde c^2_{\rm v,b} + (3/4)c^2_{\rm v,s}$.

It remains to solve the equations (3.5) and (3.7) in the presence of the source term (3.11). The small size of variables δ_l, v_l, and ϕ_l allows one to safely apply the recursive approach used in the previous chapter to obtain (2.43). At third order, one has:

$$\delta(\boldsymbol{k},\eta) = \delta^{(1)}(\boldsymbol{k},\eta) + \delta^{(2)}(\boldsymbol{k},\eta) + \delta^{(3)}(\boldsymbol{k},\eta) - k^2c_{\rm s}^2(\eta)\delta^{(1)}(\boldsymbol{k},\eta) + \delta_J(\boldsymbol{k},\eta) \tag{3.12}$$

where, as anticipated above, the stochastic counterterm, δ_J, can be shown to be much smaller than the other corrections and thus safely ignored. The new parameter $c^2_{\rm s}$ has the effect of both $\tilde c^2_{\rm s}$ and $\tilde c^2_{\rm v}$, which appear in (3.11) multiplying divergence terms [5].

[2]The term $\bar\rho\tilde c^2_{\rm s}\delta_l$, which, in (3.10), is added to the diagonal elements of the stress tensor, can be regarded as an effective pressure perturbation. In the absence of viscosity ($\tilde c^2_{\rm v,b} = \tilde c^2_{\rm v,s} = 0$), one has $\tilde c^2_{\rm s} = \delta p/(\bar\rho\delta)$, which makes $\tilde c^2_{\rm s}$, indeed, a speed of sound.

From the density field one can finally obtain the power spectrum. At next-to-leading order ("one loop"), this is:

$$P(k) = P_{\rm lin}(k) + P_{22,\Lambda}(k) + 2P_{13,\Lambda}(k) - 2c^2_{{\rm s},\Lambda}k^2 P_{\rm lin}(k) + P_{JJ,\Lambda}(k) \tag{3.13}$$

The equation above looks dependent on the nonphysical cutoff Λ used to normalize the theory, but that dependence is only apparent. Indeed, the quantities $c^2_{{\rm s},\Lambda}$ and $P_{JJ,\Lambda}$ can be shown to have the right functional relation with Λ to absorb the dependence of $P_{22,\Lambda}$ and $P_{13,\Lambda}$ [11]. One can then formally take the cutoff to infinity, so that:

$$P(k) = P_{\rm lin}(k) + P_{22}(k) + 2P_{13}(k) - 2c^2_{{\rm s},\infty}k^2 P_{\rm lin}(k) + P_{JJ,\infty}(k) \tag{3.14}$$

where P_{13} and P_{22} are computed through the ordinary integrals found in chapter 2. The quantities $c^2_{{\rm s},\infty}$ and $P_{JJ,\infty}(k)$ capture the effect of small-scale physics on the large-scale dynamics. As mentioned above, $P_{JJ,\infty}(k)$ has a comparatively small size – smaller than two-loop corrections [5] – and can usually be ignored. Since they cannot be predicted within the EFT, these parameters must be derived from data – for example, by fitting $P(k)$ above to a nonlinear spectrum.

3.2 Lagrangian perturbation theory

The theory in this chapter and in chapter 2 has been developed in Eulerian coordinates. We now seek its Lagrangian counterpart – that is, in terms of the initial coordinates of the fluid's particles and their later displacement. This will allow expansions to be made in terms of the initial density field, which has smaller magnitude, thus leading to more accurate predictions.

The final position of a fluid element at time η is written in terms of the initial position, $\boldsymbol{q}$, and a displacement vector $\boldsymbol{\Psi}$ as:

$$\boldsymbol{x}(\boldsymbol{q},\eta) = \boldsymbol{q} + \boldsymbol{\Psi}(\boldsymbol{q},\eta) \tag{3.15}$$

The equation above defines a mapping from $\boldsymbol{q}$ to $\boldsymbol{x}$. Conservation of mass is expressed in the relation:

$$\begin{aligned}\rho(\boldsymbol{q})\mathrm{d}^3q &= \rho(\boldsymbol{x})\mathrm{d}^3x \\ &= \rho(\boldsymbol{x})\mathcal{J}\mathrm{d}^3q\end{aligned} \tag{3.16}$$

where volume distortion is described by the Jacobian:

$$\mathcal{J} = \left|\frac{d^3x}{d^3q}\right| = \det\left[\delta_{ij}^{(\mathrm{K})} + \frac{\partial\Psi_i}{\partial q_j}\right] = \det\left[\delta_{ij}^{(\mathrm{K})} + \Psi_{i,j}\right] \tag{3.17}$$

where $\Psi_{i,j} \equiv \partial\Psi_i/\partial x^j$. Equation (3.16) can be written as:

$$\begin{aligned}\rho(\boldsymbol{q})\mathrm{d}^3q &= [1+\delta(\boldsymbol{x})]\rho(\boldsymbol{q})\mathcal{J}\mathrm{d}^3q \\ \mathrm{d}^3q &= [1+\delta(\boldsymbol{x})]\mathcal{J}\mathrm{d}^3q\end{aligned} \tag{3.18}$$

That is, one is assuming no inhomogeneity is present in the initial field. It follows that $\delta(\boldsymbol{x}) = 1/\mathcal{J} - 1$, so $\delta^{(1)}(\boldsymbol{x}) = -\mathcal{J}^{(1)}$. Taking the divergence of the particle equation of motion and the Poisson equation, we obtain:

$$\mathcal{J}\boldsymbol{\nabla}_x \cdot \left[\frac{d^2\boldsymbol{x}}{d\eta^2} + \mathcal{H}\frac{d\boldsymbol{x}}{d\eta}\right] = \frac{3}{2}\Omega_m(a)\mathcal{H}^2(\mathcal{J}-1) \tag{3.19}$$

The equation above is not yet expressed exclusively in terms of Lagrangian variables, $\boldsymbol{q}$ and $\boldsymbol{\Psi}(\boldsymbol{q},\eta)$, for the gradient is still taken with respect to $\boldsymbol{x}$ (that is, $\nabla^i \equiv \partial/\partial x^i$). The gradients in Eulerian and Lagrangian variables are related by:

$$\frac{\partial}{\partial x^i} = (\delta_{ij}^{(\mathrm{K})} + \Psi_{i,j})^{-1}\frac{\partial}{\partial q^j} \tag{3.20}$$

where the derivative in $\Psi_{i,j}$ is taken with respect to q^j.

Equations (3.19) and (3.20) can finally be solved for the displacement field $\boldsymbol{\Psi}(\boldsymbol{q},\eta)$. As in the Eulerian case, we will adopt the ansatz [10]:

$$\boldsymbol{\Psi} = \boldsymbol{\Psi}^{(1)} + \boldsymbol{\Psi}^{(2)} + \boldsymbol{\Psi}^{(3)} + \ldots \tag{3.21}$$

which induces, order by order, an expansion of the Jacobian:

$$\mathcal{J} = 1 + \mathcal{J}^{(1)} + \mathcal{J}^{(2)} + \mathcal{J}^{(3)} + \ldots \tag{3.22}$$

One notices that the matrix in (3.17), whose determinant is the Jacobian, can be regarded as a perturbation of the identity matrix. To expand that determinant in terms of the displacement field, we will use the following result [10]:

$$\det(I + \epsilon A) = 1 + \epsilon\mathcal{K}(A) + \epsilon^2\mathcal{L}(A) + \epsilon^3\mathcal{M}(A) + \mathcal{O}(\epsilon^4) \tag{3.23}$$

with:

$$\begin{aligned} \mathcal{K} &= \mathrm{Tr}\, A \\ \mathcal{L} &= \frac{\mathrm{Tr}^2 A - \mathrm{Tr}(A^2)}{2} \\ \mathcal{M} &= \det A \end{aligned} \tag{3.24}$$

In the case at hand, where $A_{ij} \equiv \Psi_{i,j}$, we have:

$$\begin{aligned} \mathcal{K} &= \sum_i \Psi_{i,i} \\ \mathcal{L} &= \frac{1}{2}\sum_{i\neq j}(\Psi_{i,i}\Psi_{j,j} - \Psi_{i,j}\Psi_{j,i}) \\ \mathcal{M} &= \det[\Psi_{i,j}] \end{aligned} \tag{3.25}$$

Each term above can itself be expanded in orders of $\boldsymbol{\Psi}$, so that a final, perturbative expansion of the Jacobian can be written as:

$$\begin{aligned} \mathcal{J} &= \mathcal{J}^{(1)} + \mathcal{J}^{(2)} + \mathcal{J}^{(3)} + \dots \\ &= \mathcal{K}^{(1)} + (\mathcal{K}^{(2)} + \mathcal{L}^{(2)}) + (\mathcal{K}^{(3)} + \mathcal{L}^{(3)} + \mathcal{M}^{(3)}) + \dots \end{aligned} \tag{3.26}$$

where:

$$\mathcal{J}^{(1)} = \mathcal{L}^{(1)} = \sum_i \Psi^{(1)}_{i,i} \tag{3.27}$$

$$\mathcal{J}^{(2)} = \mathcal{L}^{(2)} + \mathcal{K}^{(2)} = \sum_i \Psi^{(2)}_{i,i} + \frac{1}{2}\sum_{i\neq j}\left\{\Psi^{(1)}_{i,i}\Psi^{(1)}_{j,j} - \Psi^{(1)}_{i,j}\Psi^{(1)}_{j,i}\right\} \tag{3.28}$$

$$\mathcal{J}^{(3)} = \mathcal{L}^{(3)} + \mathcal{K}^{(3)} + \mathcal{M}^{(3)} = \sum_i \Psi^{(3)}_{i,i} + \sum_{i\neq j}\left\{\Psi^{(2)}_{i,i}\Psi^{(1)}_{j,j} - \Psi^{(2)}_{i,j}\Psi^{(1)}_{j,i}\right\} + \det \Psi^{(1)}_{i,j} \tag{3.29}$$

3.3 Solution to the Lagrangian equations of motion

We will now proceed to the solution of the equations of motion (3.19) and (3.20) using the perturbative expansion given in (3.26). More specifically, we seek only

a solution for the divergence of $\boldsymbol{\Psi}$, since, by assumption, the final field has no vorticity (section 2.3). At first order, the pair (3.19) and (3.20) gives:

$$\left[\frac{d^2}{d\eta^2} + \mathcal{H}\frac{d}{d\eta} - \frac{3}{2}\Omega_m(a)\mathcal{H}^2\right]\Psi^{(1)}_{i,i} = 0 \tag{3.30}$$

which has the same form as the linear equation obtained in the Eulerian frame (chapter 2). The first-order solution is then:

$$\Psi^{(1)}(\boldsymbol{k},\eta) = i\frac{\boldsymbol{k}}{k^2}\delta^{(1)}(\boldsymbol{k})D(\eta) \tag{3.31}$$

In the EdS model ($\Omega_m = 1$) one has $\Psi^{(n)} \propto a^n$, so that:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}a}\Psi^{(n)} &= \frac{n\Psi^{(n)}}{a}\\ \frac{\mathrm{d}^2}{\mathrm{d}a^2}\Psi^{(n)} &= \frac{n(n-1)\Psi^{(n)}}{a^2}\end{aligned} \tag{3.32}$$

Equation (3.19) for $\Psi^{(n)}$ thus reads:

$$(\delta^{(\mathrm{K})}_{i,j} + \Psi_{i,j})^{-1}(n^2 + \frac{n}{2})\mathcal{H}^2\Psi^{(n)}_{i,j} = \frac{3}{2}\mathcal{H}^2(\mathcal{J} - 1) \tag{3.33}$$

To write the equation above strictly at nth order, one needs an expansion for the operator $(\delta^{(\mathrm{K})}_{i,j} + \Psi_{i,j})^{-1}$. Up to second order it can be written as [5]:

$$(\delta^{(\mathrm{K})}_{i,j} + \Psi_{i,j})^{-1} = \delta^{(\mathrm{K})}_{ij} - \Psi_{i,j} + \Psi_{i,l}\Psi_{l,j} \tag{3.34}$$

Equation (3.33) now gives:

$$\begin{aligned}&\mathcal{J}(\delta_{ij} + \Psi_{i,j})^{-1}\left[\frac{3}{2}\Psi^{(1)}_{i,j} + 5\Psi^{(2)}_{i,j}\right] = \frac{3}{2}(\mathcal{J} - 1)\\ \implies &\frac{3}{2}\mathcal{J}^{(1)}\Psi^{(1)}_{i,i} - \frac{3}{2}\Psi^{(1)}_{i,j}\Psi^{(1)}_{i,j} + 5\Psi^{(2)}_{i,i} = \frac{3}{2}\mathcal{J}^{(2)}\end{aligned} \tag{3.35}$$

$$\frac{3}{2}\mathcal{J}^{(1)}\Psi^{(1)}_{i,i} - \frac{3}{2}\Psi^{(1)}_{i,j}\Psi^{(1)}_{i,j} + 5\Psi^{(2)}_{i,i} = \frac{3}{2}\mathcal{J}^{(2)} \tag{3.36}$$

Using the expressions for $\mathcal{J}^{(1)}$ and $\mathcal{J}^{(2)}$, we finally obtain:

$$\frac{3}{2}\left[\Psi^{(1)}_{i,i}\Psi^{(1)}_{j,j} - \Psi^{(1)}_{i,j}\Psi^{(1)}_{i,j}\right] + 5\Psi^{(2)}_{i,i} = \frac{3}{2}\Psi^{(2)}_{i,i} + \frac{3}{4}\left[\Psi^{(1)}_{i,i}\Psi^{(1)}_{j,j} - \Psi^{(1)}_{i,j}\Psi^{(1)}_{i,j}\right] \tag{3.37}$$

which yields an expression for the second-order divergence:

$$\mathcal{K}^{(2)} = \Psi^{(2)}_{i,i} = -\frac{3}{14}\sum_{i,j}\left[\Psi^{(1)}_{i,i}\Psi^{(1)}_{j,j} - \Psi^{(1)}_{i,j}\Psi^{(1)}_{i,j}\right] = -\frac{3}{7}\mathcal{L}^{(2)} \tag{3.38}$$

Equations (3.31) and (3.38), together with the assumption of no vorticity, yield a solution for $\boldsymbol{\Psi}(\boldsymbol{q},\eta)$ at second order. One can proceed to an arbitrary nth order by recursion. In general, the solution can be written as:

$$\boldsymbol{\Psi}^{(n)}(\boldsymbol{k}) = \frac{i}{n!}\prod_{i=1}^{n}\left\{\int\frac{\mathrm{d}^3q_i}{(2\pi)^3}\delta^{(1)}(\boldsymbol{q}_i)\right\}\boldsymbol{L}_n(\boldsymbol{q}_1,\ldots,\boldsymbol{q}_n)(2\pi)^3\delta^{(\mathrm{D})}(\boldsymbol{k}-\boldsymbol{q}_1-\cdots-\boldsymbol{q}_n) \tag{3.39}$$

The first-order kernel, as we found above, is $L_1(\boldsymbol{k}) = \boldsymbol{k}/k^2$. The second- and third-order kernels are:

$$\boldsymbol{L}_2 = \frac{3}{7}\frac{\boldsymbol{k}}{k^2}\left(1-\frac{(\boldsymbol{q}_1\cdot\boldsymbol{q}_2)^2}{q_1^2q_2^2}\right) \tag{3.40}$$

$$\begin{aligned}\boldsymbol{L}_3 = {}& \frac{5}{7}\frac{\boldsymbol{k}}{k^2}\left[1-\left(\frac{\boldsymbol{q}_1\cdot\boldsymbol{q}_2}{q_1q_2}\right)^2\right]\left[1-\left(\frac{(\boldsymbol{q}_1+\boldsymbol{q}_2)\cdot\boldsymbol{q}_3}{|\boldsymbol{q}_1+\boldsymbol{q}_2|q_3}\right)^2\right]\\ & -\frac{1}{3}\frac{\boldsymbol{k}}{k^2}\left[1-3\left(\frac{\boldsymbol{q}_1\cdot\boldsymbol{q}_2}{q_1q_2}\right)^2+2\frac{\boldsymbol{q}_1\cdot\boldsymbol{q}_2\,\boldsymbol{q}_2\cdot\boldsymbol{q}_3\,\boldsymbol{q}_3\cdot\boldsymbol{q}_1}{q_1^2q_2^2q_3^2}\right]\end{aligned} \tag{3.41}$$

Having found $\boldsymbol{\Psi}$, we can now obtain the density field through:

$$\begin{aligned}\delta(\boldsymbol{k}) &= \int \mathrm{d}^3x\exp[-i\boldsymbol{k}\cdot\boldsymbol{x}]\delta(\boldsymbol{x})\\ &= \int \mathrm{d}^3x\exp[-i\boldsymbol{k}\cdot\boldsymbol{x}][1+\delta(\boldsymbol{x})] - \int \mathrm{d}^3q\exp[-i\boldsymbol{k}\cdot\boldsymbol{q}]\\ &= \int \mathrm{d}^3q\exp[-i\boldsymbol{k}\cdot(\boldsymbol{q}+\boldsymbol{\Psi})] - \int \mathrm{d}^3q\exp[-i\boldsymbol{k}\cdot\boldsymbol{q}]\\ &= \int \mathrm{d}^3q\exp[-i\boldsymbol{k}\cdot\boldsymbol{q}](\exp[-i\boldsymbol{k}\cdot\boldsymbol{\Psi}(\boldsymbol{q})]-1)\end{aligned} \tag{3.42}$$

where we have used the continuity relation $[1+\delta(\boldsymbol{x})]\,\mathrm{d}^3x = \mathrm{d}^3q$.

To obtain the power spectrum, we compute the correlator:

$$\begin{aligned}\langle\delta(\boldsymbol{k})\delta(\boldsymbol{k}')\rangle &= \left\langle \int \mathrm{d}^3q\, e^{-i\boldsymbol{k}\cdot\boldsymbol{q}}(e^{-i\boldsymbol{k}\cdot\boldsymbol{\Psi}(\boldsymbol{q})}-1)\int \mathrm{d}^3q'\, e^{-i\boldsymbol{k}'\cdot\boldsymbol{q}'}(e^{-i\boldsymbol{k}'\cdot\boldsymbol{\Psi}(\boldsymbol{q}')}-1)\right\rangle \\ &= \left\langle \int \mathrm{d}^3q\, e^{-i\boldsymbol{k}\cdot\boldsymbol{q}}(e^{-i\boldsymbol{k}\cdot\boldsymbol{\Psi}(\boldsymbol{q})}-1)\int \mathrm{d}^3r\, e^{-i\boldsymbol{k}'\cdot(\boldsymbol{q}-\boldsymbol{r})}(e^{-i\boldsymbol{k}'\cdot\boldsymbol{\Psi}(\boldsymbol{q}-\boldsymbol{r})}-1)\right\rangle \\ &= \int \mathrm{d}^3q\, e^{-i(\boldsymbol{k}+\boldsymbol{k}')\cdot\boldsymbol{q}}\int \mathrm{d}^3r\, e^{i\boldsymbol{k}'\cdot\boldsymbol{r}}\left\langle(e^{-i\boldsymbol{k}\cdot\boldsymbol{\Psi}(\boldsymbol{q})}-1)(e^{-i\boldsymbol{k}'\cdot\boldsymbol{\Psi}(\boldsymbol{q}-\boldsymbol{r})}-1)\right\rangle\end{aligned} \tag{3.43}$$

where we have defined the relative variable $\boldsymbol{r} \equiv \boldsymbol{q} - \boldsymbol{q}'$. Cosmic homogeneity implies that the expression between angle brackets cannot depend on $\boldsymbol{q}$. It is a function only of $\boldsymbol{r}$. The first integral can thus be evaluated separately, yielding a Dirac delta function. We have:

$$\begin{aligned}\langle\delta(\boldsymbol{k})\delta(\boldsymbol{k}')\rangle &= (2\pi)^3\delta_{\mathrm{D}}(\boldsymbol{k}+\boldsymbol{k}')\int \mathrm{d}^3r\, e^{-i\boldsymbol{k}\cdot\boldsymbol{r}}\times \\ &\qquad \left\langle e^{-i\boldsymbol{k}[\boldsymbol{\Psi}(\boldsymbol{q})-\boldsymbol{\Psi}(\boldsymbol{q}')]} - e^{-i\boldsymbol{k}\boldsymbol{\Psi}(\boldsymbol{q})} - e^{i\boldsymbol{k}\boldsymbol{\Psi}(\boldsymbol{q}-\boldsymbol{r})} + 1\right\rangle \\ &= (2\pi)^3\delta_{\mathrm{D}}(\boldsymbol{k}+\boldsymbol{k}')\int \mathrm{d}^3r\, e^{-i\boldsymbol{k}\cdot\boldsymbol{r}}\langle e^{-i\boldsymbol{k}\cdot\Delta\boldsymbol{\Psi}}-1\rangle\end{aligned} \tag{3.44}$$

where $\Delta\Psi(\boldsymbol{r}) \equiv \boldsymbol{\Psi}(\boldsymbol{q}) - \boldsymbol{\Psi}(\boldsymbol{q}-\boldsymbol{r})$. It follows that the power spectrum is given, to arbitrary order, by:

$$P(k) = \int \mathrm{d}^3r \exp[-i\boldsymbol{k}\cdot\boldsymbol{r}]\langle\exp[-i\boldsymbol{k}\cdot\Delta\boldsymbol{\Psi}]-1\rangle \tag{3.45}$$

If we assume the displacement field, $\Delta\boldsymbol{\Psi}$, to be Gaussian, the cumulant-expansion theorem can be used to bring the expectation value into the exponential: $\langle\exp[i\boldsymbol{k}\cdot\Delta\boldsymbol{\Psi}]\rangle = \exp[-1/2\langle(\boldsymbol{k}\cdot\Delta\boldsymbol{\Psi})^2\rangle]$. At first order, for ΛCDM, one gets the so-called Zel'dovich solution [24]:

$$P^{\mathrm{Zel}}(k) = \int \mathrm{d}q^3 e^{i\boldsymbol{k}\cdot\boldsymbol{q}}\left[\exp\left(D^2(\eta)k_ik_j\int_p \frac{p^ip^j}{p^4}P_{\mathrm{L}}(\boldsymbol{p})(e^{i\boldsymbol{q}\cdot\boldsymbol{p}}-1)\right)-1\right] \tag{3.46}$$

where $D(\eta)$ is the same as in the Eulerian case (section 2.3).

3.4 Galaxy bias

Galaxies form in dark matter halos, which are bound structures of dark matter. One can get an analytical understanding of the relationship between the galaxy and dark matter density distributions – the theory of "galaxy bias" – by studying

the statistical properties of dark matter halos. If the halo is atypically massive, one can reasonably suppose it to have evolved independently of its neighborhood, i.e., as an isolated system [17]. Hence the simplest model of halo formation – the "spherical-collapse model" – considers the collapse of an isolated, spherical region.

3.4.1 The spherical-collapse model

We consider a region of mass M and density slightly higher than the mean density in the universe. We can compute its comoving radius by considering the radius of a homogeneous region of same mass on the present day. Its density is:

$$\rho_m(t_0) = \Omega_m \rho_{\rm cr} = \frac{M}{4\pi R^3/3} \tag{3.47}$$

which gives the "Lagrangian radius":

$$R_L(M) = 1.40 h^{-1}{\rm Mpc} \left(\frac{M}{10^{12} h^{-1} M_\odot} \right)^{1/3} \tag{3.48}$$

That is, $R_L(M)$ is the comoving radius of a sphere of mass M in an unperturbed universe, with density differing only slightly from the mean matter density.

We now consider the dynamics of such a region, treating it as an isolated system. Its metric is the same as in FLRW cosmology with a changed matter density, $\tilde{\rho}_m$. The Friedmann equation thus reads:

$$\frac{\ddot{\tilde{a}}}{\tilde{a}} = -\frac{4\pi G}{3} \tilde{\rho}_m \tag{3.49}$$

The physical radius (not the comoving one) grows proportionally to the scale factor. This yields the equation:

$$\begin{aligned} \frac{\ddot{r}}{r} &= -\frac{4\pi G}{3} \frac{M}{4\pi r^3(t)/3} \\ \implies \ddot{r}(t) &= -\frac{GM}{r^2(t)} \end{aligned} \tag{3.50}$$

We wish to solve the above equation under the following initial conditions:

$$r(t_{\rm ini}) = a(t_{\rm ini}) R_L, \quad \dot{r}(t_{\rm ini}) = \dot{a}_{\rm ini} R_L \tag{3.51}$$

That is, at the initial time, $t_{\rm ini}$, the sphere is supposed to have had the comoving radius, R_L, and to have expanded together with the Hubble flow.

The solution is expressed parametrically as:

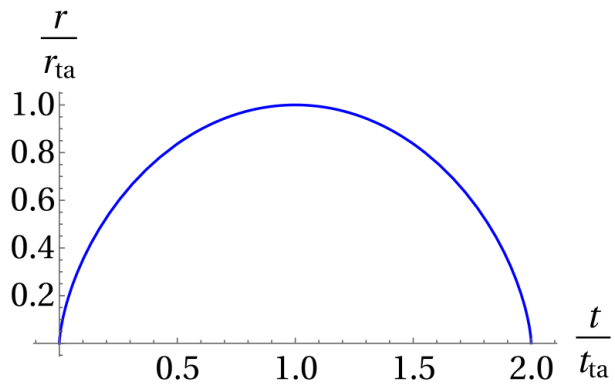

Figure 3.1: Solution of equation (3.50) describing the collapse of a sphere. It initially expands according to the Hubble flow. At the instant of "turn around", $t = t_{\rm ta}$, expansion halts and contraction begins. Source: [17].

$$\begin{aligned} r &= \frac{r_{\rm ta}}{2}(1-\cos\theta) \\ t &= \frac{t_{\rm ta}}{\pi}(\theta-\sin\theta) \end{aligned} \tag{3.52}$$

It is shown in figure 3.1. Collapse happens at $t = 2t_{\rm ta}$, where $t_{\rm ta}$ is the instant of "turn around": the sphere stops expanding and begins to contract. The parameters $t_{\rm ta}$ and $r_{\rm ta}$ depend on the initial size and overdensity of the region [17]. The chief result is: if the region had an initial overdensity, $\delta(t_{\rm ini})$, such that its linearly-evolved counterpart at time t exceeds the threshold

$$\delta^{(1)}_{R_L}(t) > \delta_c = \frac{3}{5}\left(\frac{3\pi}{2}\right)^{2/3} \approx 1.686 \tag{3.53}$$

then it has collapsed by time t [17]. By "linearly evolved", we mean an evolution according to the growth factor of linear theory, $D(t)$, presented in (2.33).

We now proceed to study the statistics of collapsed regions according to the above model. What fraction of matter has collapsed into halos of mass M by the present day? That is the fraction which, in the unperturbed universe, exhibited a density satisfying condition (3.53). In other words, it is the fraction of present-day linear density field such that $\delta(\boldsymbol{r}) \geq \delta_c$. The rms amplitude of fluctuations of size R_L in that field is:

$$\sigma^2(M) = \int \frac{{\rm d}^3 q}{(2\pi)^3} P(q) W_{\rm TH}(kR) \tag{3.54}$$

where $W_{TH}(kR_L)$ is a top-hat window of size R_L, $P(q)$ is the linear power spectrum

today, and M is the mass corresponding to $R_L(M)$.

Since the overdensity is a Gaussian field, the probability density of a given value δ is [17, 5]:

$$\mathbb{P}(\delta|M) = \frac{1}{\sqrt{2\pi\sigma^2(M)}} \exp\left[-\frac{1}{2}\frac{\delta^2}{\sigma^2(M)}\right] \tag{3.55}$$

The fraction of the density field to have collapsed is hence:

$$\mathbb{P}(>\delta|M) = \int_{\delta_c}^{\infty} \mathrm{d}\delta\, \mathbb{P}(\delta|M) = \int_{\nu}^{\infty} \mathrm{d}x \exp\left[-\frac{x^2}{2}\right] \tag{3.56}$$

with the *peak height* ν defined as $\nu \equiv \delta_c/\sigma$.

From this one can obtain the *halo mass function*, i.e, the population density of halos of a certain mass. By definition, it involves a derivative of (3.56) with respect to M. It is given, more precisely, by the mean number density of structures of mass M (that is, $\bar{\rho}/M$) multiplied by the fraction that has collapsed into halos of mass M [17]:

$$n(M) = -\sqrt{\frac{2}{\pi}}\frac{\bar{\rho}}{M^2}\nu \exp\left[-\frac{\nu^2}{2}\right]\frac{\mathrm{d}\ln\sigma}{\mathrm{d}\ln M} \tag{3.57}$$

3.4.2 Halo formation in the presence of a long-wavelength matter field

We now seek a relationship between the halo and matter overdensities. We expect such a relationship to exist, since the presence of a perturbation of the large-scale matter field at some point, $\delta_l(\boldsymbol{x})$, moves the overdensity closer or farther from the threshold δ_c.

One should recall that $n(M)$ in (3.57) is also a function of ν, and therefore of δ_c through the definition $\nu \equiv \delta_c/\sigma(M)$. In the presence of a long-wavelength perturbation of the matter field, the collapse threshold becomes effectively $\tilde{\delta}_c = \delta_c - \delta_l$. The effective peak height is then:

$$\tilde{\nu} = \frac{\delta_c - \delta_l}{\sigma} \tag{3.58}$$

We now seek $n(\tilde{\nu})$. Since δ_l is assumed to be small, one can expand $n(\tilde{\nu})$ around $\tilde{\nu}|_{\delta_l=0} = \nu$. Hence:

$$n(\tilde{\nu}) = \bar{n}(\nu) + \frac{\partial n(\tilde{\nu})}{\partial \delta_l}\delta_l + \frac{1}{2}\frac{\partial^2 n(\tilde{\nu})}{\partial \delta_l^2}\delta_l^2 + \dots \tag{3.59}$$

from which we obtain the *galaxy overdensity*:

$$\delta_g(\boldsymbol{q}) = \frac{n}{\bar{n}} - 1 = b_1^{(\mathrm{L})}\delta_l(\boldsymbol{q}) + \frac{b_2^{(\mathrm{L})}}{2}\delta_l^2(\boldsymbol{q}) + \frac{b_3^{(\mathrm{L})}}{3!}\delta_l^3(\boldsymbol{q}) + \dots \tag{3.60}$$

We have assumed that the halo and galaxy overdensities are the same – which, as shown in hydrodynamic simulations, is not far from the truth [30, 7]. The Lagrangian *bias parameters* are defined as:

$$b_i^{(\mathrm{L})} = \frac{1}{\bar{n}}\frac{\partial^i n}{\partial \delta_l^i} \tag{3.61}$$

With the mass function (3.57), one obtains:

$$\begin{aligned} b_1^{(\mathrm{L})} &= \frac{\nu^2 - 1}{\delta_c} \\ b_2^{(\mathrm{L})} &= \frac{\nu^2(\nu^2 - 3)}{\delta_c^2} \end{aligned} \tag{3.62}$$

3.4.3 Bias in Eulerian space

The treatment above was made in Lagrangian space – that is, in terms of the initial matter field (in the almost-unperturbed universe) rescaled according to linear theory. To arrive at the bias expansion in Eulerian space, one considers the continuity relation:

$$[1 + \delta(\boldsymbol{x})]\mathrm{d}^3x = \mathrm{d}^3q \tag{3.63}$$

which establishes mass conservation.

Assuming that the halos are comoving with matter and their number is conserved, we can write an analogous relation between the Lagrangian and Eulerian halo overdensities [5]:

$$\left[1 + \delta_g^{(\mathrm{E})}(\boldsymbol{x})\right]\mathrm{d}^3x = \left[1 + \delta_g^{(\mathrm{L})}(\boldsymbol{q})\right]\mathrm{d}^3q \tag{3.64}$$

Using (3.63), one can then relate Eulerian halo and matter overdensities:

$$\delta_g^{(\mathrm{E})}(\boldsymbol{x}) = \delta_g^{(\mathrm{L})}(\boldsymbol{q}) + \delta(\boldsymbol{x})\delta_g^{(\mathrm{L})}(\boldsymbol{q}) + \delta(\boldsymbol{x}) \tag{3.65}$$

At second order, this yields [5]:

$$\delta_g^{(E,2)}(\boldsymbol{x}) = \left(b_1^{(\mathrm{L})} + 1\right)\delta^{(2)}(\boldsymbol{x}) + \frac{1}{2}\left(b_2^{(\mathrm{L})} + \frac{8}{21}b_1^{(\mathrm{L})}\right)\left[\delta^{(1)}(\boldsymbol{x})\right]^2 - \frac{2}{7}b_1^{(\mathrm{L})}s^2(\boldsymbol{x}) \quad (3.66)$$

from which we can define Eulerian bias parameters, $b_i^{(\mathrm{E})}$, such that:

$$\delta_g^{(E,2)}(\boldsymbol{x}) = b_1^{(\mathrm{E})}\delta^{(2)}(\boldsymbol{x}) + \frac{b_2^{(\mathrm{E})}}{2}\left[\delta^{(1)}(\boldsymbol{x})\right]^2 + b_{s^2}^{(\mathrm{E})}s^2(\boldsymbol{x}) \quad (3.67)$$

In the above, $s^2 \equiv s_{ij}s^{ij}$, and s_{ij} is the tidal field:

$$s_{ij}(\boldsymbol{x}) = \left[\frac{\partial_i\partial_j}{\partial^2} - \frac{1}{3}\delta_{ij}^{\mathrm{K}}\right]\delta(\boldsymbol{x}) \quad (3.68)$$

The galaxy overdensity can also receive contributions uncorrelated with the matter field. At second order, these are described by a "stochastic" component, $\epsilon(\boldsymbol{x}, \eta)$, added to the bias expansion (3.67) [24]. If the galaxy distribution is Poissonian, it gives a contribution to the power spectrum – "shot noise" – described by:

$$\langle\epsilon(\boldsymbol{x}, \eta)\epsilon(\boldsymbol{x}', \eta)\rangle = P_\epsilon(\eta)\delta^{(\mathrm{D})}(\boldsymbol{x} - \boldsymbol{x}') \quad (3.69)$$

where $P_\epsilon(\eta) = 1/\bar{n}(\eta)$, and $\bar{n}(\eta)$ is the mean galaxy density.

Equation (3.67), with an additional stochastic component, had been the aim of this chapter. Substitution of $\delta^{(1)}$, $\delta^{(2)}$, and s^2 obtained in effective-field theory (see (3.12)) will yield an accurate prescription for the galaxy overdensity on scales as small as $k \sim 0.3h\,\mathrm{Mpc}^{-1}$. One must recall, however, that a number of free parameters have been introduced along the way, and that these must be fit to the data.

In the following, the theory reviewed in this chapter and in chapter 2 will be put to use. A simple model of galaxy bias – with one parameter, b_1 –, together with the EFT's correction to the matter overdensity at second order, will be used to constrain cosmological models in analyses of large-scale structure.

The accuracy of the linear-bias model is limited to scales larger than $k \sim 0.1h\,\mathrm{Mpc}^{-1}$. Data on smaller scales can still be included in the analysis, however, if one takes proper account of theoretical uncertainty in that regime. This will be based on assessing the linear model's ability to predict the galaxy power spectrum

for k-modes higher than $k \sim 0.1h\,\mathrm{Mpc}^{-1}$, and then including the quantified "theoretical uncertainty" in analyses of those scales.

Chapter 4

Theoretical uncertainty on galaxy bias

The linear-bias model for the galaxy power spectrum is adopted in present-day observational programs, such as the Dark Energy Survey [19]. It is known to lose accuracy, however, on scales below $k \sim 0.1h\,\mathrm{Mpc}^{-1}$. A standard approach to that problem, the "scale cut", consists in discarding data on k-modes above a certain k_{max}. In this chapter, we test the inclusion of those modes in the analysis with reduced statistical weight, which is assessed with respect the model's credibility – i.e., the theoretical uncertainty – in that regime. As we aim to show, this can typically raise k_{max} from near $0.1h\,\mathrm{Mpc}^{-1}$ to near 0.2-0.3 Mpc^{-1}, with beneficial effects on precision.

In our likelihood analyses, the theoretical prediction for the matter power spectra will be given by EFT at one-loop order and in the Lagrangian frame, as obtained in chapter 3. From this, the galaxy power spectra will be computed under the assumption of linear galaxy bias – that is, with only the b_1 parameter in the bias expansion. The mock data vectors used for validation, on the other hand, will be generated from the EFT power spectra of one loop and the galaxy-bias expansion at second order.

4.1 Power spectra of galaxy clustering and galaxy-galaxy lensing

We will assume the galaxy-bias expansion at second order in the matter overdensity. In the Lagrangian description, it can be written as:

$$\delta_g^{(\mathrm{L})}(\boldsymbol{q}) = b_1^{(\mathrm{L})}\delta^{(\mathrm{L})}(\boldsymbol{q}) + \frac{b_2^2}{2}(\delta^{(\mathrm{L})})^2(\boldsymbol{q}) + b_{s^2}(s^{(\mathrm{L})})^2(\boldsymbol{q}) \tag{4.1}$$

where $\boldsymbol{q}$ is the initial position of the fluid particle. As seen in chapter 3, $(s^{(\mathrm{L})})^2 \equiv s_{ij}^{(\mathrm{L})}(s^{(\mathrm{L})})^{ij}$, and $s_{ij}^{(\mathrm{L})}$ is the tidal field:

$$s_{ij}^{(\mathrm{L})}(\boldsymbol{q}) = \left[\frac{\partial_i \partial_j}{\partial^2} - \frac{1}{3}\delta_{ij}^{\mathrm{K}}\right] \delta^{(\mathrm{L})}(\boldsymbol{q}) \tag{4.2}$$

To obtain the late-time overdensity, $\delta_g(\boldsymbol{x}, \tau)$, one makes use of the continuity relation:

$$\left[1 + \delta_g^{(\mathrm{L})}(\boldsymbol{q})\right] \mathrm{d}^3 q = \left[1 + \delta_g(\boldsymbol{x}, \tau)\right] \mathrm{d}^3 x \tag{4.3}$$

where $\boldsymbol{x}$ is the particle's position at time τ. The initial and final coordinates are related by the displacement vector $\boldsymbol{\Psi}(\boldsymbol{q}, \tau)$, so that $\boldsymbol{x}(\tau) = \boldsymbol{q} + \boldsymbol{\Psi}(\boldsymbol{q}, \tau)$. It will also be convenient to define:

$$F(\boldsymbol{q}) \equiv 1 + \delta_g^{(\mathrm{L})}(\boldsymbol{q}) \tag{4.4}$$

Multiplying (4.3) by $e^{-\boldsymbol{k}\cdot\boldsymbol{x}}$ and integrating it, we obtain the Fourier transform of $[1 + \delta_g(\boldsymbol{x}, \tau)]$ as:

$$(2\pi)^3 \delta^{(\mathrm{D})}(\boldsymbol{k}) + \delta_g(\boldsymbol{k}, \tau) = \int \mathrm{d}q^3 \, e^{-i\boldsymbol{k}\cdot[\boldsymbol{q}+\boldsymbol{\Psi}(\boldsymbol{q},\tau)]} F(\boldsymbol{q}) \tag{4.5}$$

so that:

$$\begin{aligned} \delta_g(\boldsymbol{k}, \tau) &= \int \mathrm{d}q^3 \, e^{-i\boldsymbol{k}\cdot[\boldsymbol{q}+\boldsymbol{\Psi}(\boldsymbol{q},\tau)]} F(\boldsymbol{q}) - (2\pi)^3 \delta^{(\mathrm{D})}(\boldsymbol{k}) \\ &= \int \mathrm{d}q^3 \, e^{-i\boldsymbol{k}\cdot\boldsymbol{q}} \left[F(\boldsymbol{q}) e^{-i\boldsymbol{k}\cdot\boldsymbol{\Psi}(\boldsymbol{q},\tau)} - 1\right] \end{aligned} \tag{4.6}$$

which expresses the late-time, galaxy overdensity $\delta_g(\boldsymbol{k}, \tau)$ in terms of $F(\boldsymbol{q})$, and hence of the bias expansion in the Lagrangian basis, (4.1).

The power spectrum of galaxy clustering, $P_{gg}(\boldsymbol{k})$, is defined through:

$$P_{gg}(\boldsymbol{k})(2\pi)^3 \delta_D(\boldsymbol{k} + \boldsymbol{k}') = \langle \delta_g(\boldsymbol{k}) \delta_g(\boldsymbol{k}') \rangle \tag{4.7}$$

One can then use (4.6) and (4.4), together with the Lagrangian bias expansion (4.1), to write the galaxy power spectrum in terms of the Lagrangian bias parameters:

$$P_{gg}(\boldsymbol{k}) = \sum_i \sum_j b_{\mathcal{O}_i} b_{\mathcal{O}_j} P_{\mathcal{O}_i \mathcal{O}_j}(\boldsymbol{k}) \tag{4.8}$$

with $\mathcal{O}_i \in \{1, \delta, \delta^2, s^2, \nabla^2\delta\}$ and suitable definitions of $P_{\mathcal{O}_i\mathcal{O}_j}(\boldsymbol{k})$ – e.g., given in [12]. By adopting (4.1), we are neglecting cubic, Lagrangian bias parameters entering the one-loop $P(k)$ through the term $P_{13}(k)$. They have been shown to be negligible in simulations [1]. The stochastic contribution to the galaxy field will be assumed

to be described by shot-noise and accounted for at the level of the covariance matrix (see (4.13) below).

The galaxy-matter cross-spectrum, on the other hand, is written as:

$$P_{gm}(\boldsymbol{k}) = \sum_i b_{\mathcal{O}_i} P_{1\mathcal{O}_i}(\boldsymbol{k}) \tag{4.9}$$

We are interested, in fact, in the angular power spectrum of galaxy clustering and galaxy-galaxy lensing. These can be obtained from $P_{gg}(\boldsymbol{k})$ and $P_{gm}(\boldsymbol{k})$, in the Limber approximation [21], through:

$$C_{gg}^{ij} = \int_0^\infty d\chi \frac{\phi^i(\chi)\phi^j(\chi)}{\chi^2} P_{gg}\left(\frac{\ell}{\chi}, z(\chi)\right) \tag{4.10}$$

$$C_{g\kappa}^{ij} = \int_0^\infty d\chi \frac{\phi^i(\chi)g^j(\chi)}{\chi^2} P_{gm}\left(\frac{\ell}{\chi}, z(\chi)\right) \tag{4.11}$$

where $\phi^i(\chi)$ and $g^i(\chi)$ at bin i are the selection functions for clustering and lensing, respectively, and χ is the comoving radial distance. Specifically, $\phi^i(\chi)$ is the radial density of galaxies in the lens sample, $\phi^i(\chi) = dN^i/d\chi$, and $g^i(\chi)$ is the lens efficiency of a flat cosmology, given by:

$$g^i(\chi) \equiv \frac{3\Omega_m H_0^2}{2c^2 a(\chi)} \int_0^\infty dz\, n^i(z) \frac{(\chi'(z) - \chi)\chi}{\chi'(z)} \Theta(\chi'(z) - \chi) \tag{4.12}$$

where $n^i(z)$ is the radial density of galaxies in the source sample, Ω_m is the matter-density parameter, c is the speed of light, $a(\chi)$ is the expansion scale factor as a function of χ, H_0 is the Hubble parameter on present day, and $\Theta(\chi'(z) - \chi)$ is the Heaviside step function [20].

We will consider only the Gaussian components of the overdensity fields. The covariance matrix is then [18]:

$$\begin{aligned}\mathrm{Cov}\left[C_{AB}^{ij}(\ell_1), C_{CD}^{kl}(\ell_2)\right] = \frac{4\pi\delta_{\ell_1\ell_2}}{\Omega_s(2\ell_1+1)\Delta\ell_1} \times \\ \Big[(C_{AC}^{ik}(\ell_1) + \delta_{ik}\delta_{AC}N_A^i)(C_{BD}^{jl}(\ell_2) + \delta_{jl}\delta_{BD}N_B^j) + \\ (C_{AD}^{il}(\ell_1) + \delta_{il}\delta_{AD}N_A^i)(C_{BC}^{jk}(\ell_2) + \delta_{jk}\delta_{BC}N_B^j)\Big]\end{aligned} \tag{4.13}$$

Here, uppercase letters A to D represent the probes (either g, galaxy count, or κ,

lensing) and indices i to k identify the redshift bin. The parameter Ω_s denotes the survey area, which we will take as the full sky ($\Omega_s = 1$) for simplicity. $\Delta\ell$ is the length of ℓ binning. The expression above implies, for example, that there is nonzero covariance between spectra projected in different redshift bins, as well as between different spectra (C_{gg} and $C_{g\kappa}$) defined in the same redshift bin.

The parameters N_A^i are probe-specific noise terms, of which we will be concerned with shot noise and shape noise. The former will be given by $N_g^i = 1/\bar{n}^i_{\text{lens}}$, where $\bar{n}^i_{\text{lens}}$ is the mean galaxy density in the ith lens bin [20]. Shape noise is $N_\kappa^i = 2\sigma_\epsilon^2/\bar{n}^i_{\text{source}}$, where $\bar{n}^i_{\text{source}}$ is the mean galaxy density in the ith source bin and σ_ϵ is the rms shear arising from the intrinsic ellipticity of galaxies and measurement noise [18].

As to ℓ-binning and redshift binning, we will adopt the year-10 conventions of LSST-DESC [16]. These prescribe 20 ℓ-bins logarithmically spaced between 20 and 15000. The tomographic analysis of galaxy clustering is segmented in 10 redshift bins on the interval $0.2 \leq z \leq 1.2$, spaced by 0.1 in photo-z. Of these, we selected two for use in the estimations[1]: $\langle z \rangle = 0.26$ and $\langle z \rangle = 0.84$ (figure 4.1). The analysis of weak lensing is performed with 5 bins in that same interval, the photo-z limits being chosen so that each bin represents the same galaxy density. We chose bins $\langle z \rangle = 0.34$ and $\langle z \rangle = 0.86$ (figure 4.2).

The lens sample satisfies the distribution $dn/dz \propto z^2 \exp[-(z/z_0)^\alpha]$, with $z_0 = 0.28$ and $\alpha = 0.90$ and an overall normalization of 48 arcmin^{-2}. The photo-z scatter is $\sigma_z = 0.03(1+z)$. The source sample satisfies that same parametric distribution with $z_0 = 0.11$ and $\alpha = 0.68$, and an overall normalization of 27 arcmin^{-2}. The photo-z scatter is $\sigma_z = 0.05(1+z)$ in this case.

4.2 Marginalization over theoretical uncertainty

We will perform likelihood analyses of angular spectra under the assumption of some nonzero theoretical error – namely, the systematic error arising from the use of an incomplete bias model. The exact extent of that error can't be known with certainty, since nonlinear-bias parameters assume uncertain values. Therefore, the theoretical error $\boldsymbol{e}$ will be treated as a random variable with an yet-unspecified probability distribution $P(\boldsymbol{e})$.

[1] An analysis of two bins is simple enough to highlight the effects we aim to study. It also takes account of details not present in manipulations of one single bin, as cross-correlations between spectra at different redshifts. For future work, we recommend these tests to be extended to more realistic conditions, such as 10 and 5 redshift bins of lens and source samples, respectively.

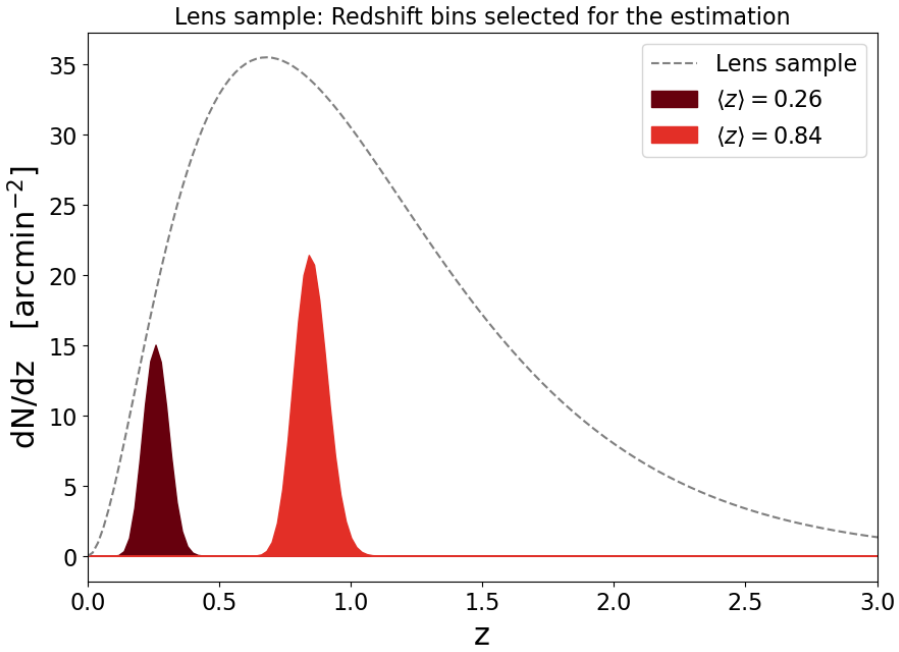

Figure 4.1: Redshift bins of galaxy clustering selected for analysis. The lens sample (dashed curve) satisfies the distribution $dn/dz \propto z^2 \exp[-(z/z_0)^\alpha]$, with $z_0 = 0.28$ and $\alpha = 0.90$ and an overall normalization of 48 arcmin^{-2}.

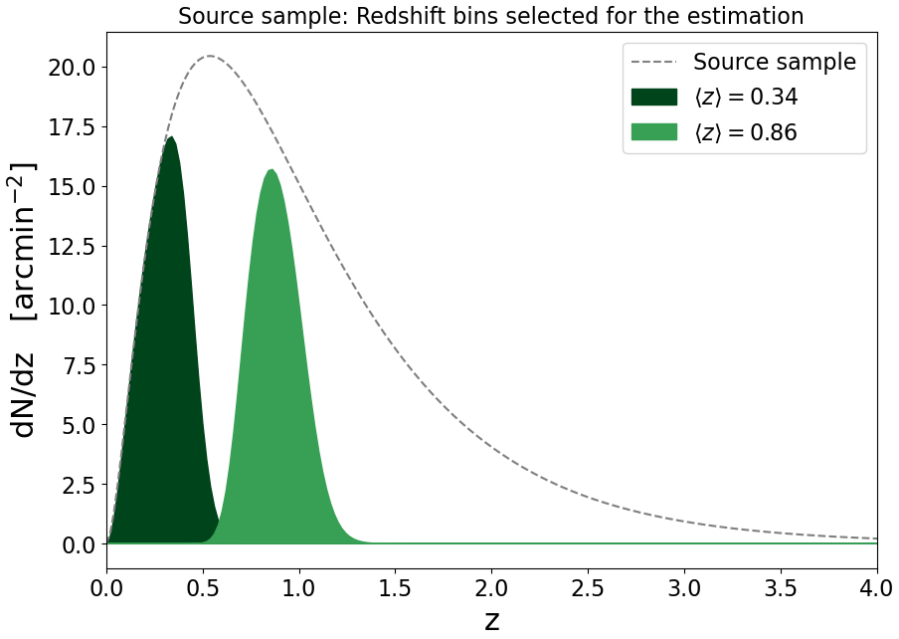

Figure 4.2: Redshift bins of galaxy lensing selected for analysis. The source sample (dashed curve) satisfies $dn/dz \propto z^2 \exp[-(z/z_0)^\alpha]$ with $z_0 = 0.11$ and $\alpha = 0.68$ and an overall normalization of 27 arcmin^{-2}.

In a given inference problem, let $\boldsymbol{d}$ be a set of m data points – for example, the angular spectrum of galaxy clustering, C_ℓ^{gg}, measured at an m number ℓ-values. Let $\boldsymbol{t}(\theta)$ be the corresponding theoretical prediction, which depends on a certain set of parameters θ that one wishes to infer from the data. We are then interested in the probability of θ given $\boldsymbol{d}$, $P(\theta|\boldsymbol{d})$. Using Bayes's prescription, this can be related to the probability (or "likelihood") of $\boldsymbol{d}$ given θ through:

$$P(\theta|\boldsymbol{d}) = \frac{\mathcal{L}(\boldsymbol{d}|\theta)\, P_{\rm prior}(\theta)}{P(\boldsymbol{d})} \tag{4.14}$$

The denominator in the above expression is independent of θ, so it can be computed from the normalization condition for the probability. After a prior on θ has been chosen, therefore, the probability $P(\theta|\boldsymbol{d})$ is fully determined by the likelihood $\mathcal{L}(\boldsymbol{d}|\theta)$.

Let $\boldsymbol{r}$ be the prediction of a complete, unknown model describing the phenomenon in question. For example, $\boldsymbol{r}$ would be the actual angular spectrum for galaxy clustering at certain redshift bin, whereas $\boldsymbol{d}$ would be the spectrum measured from the observed galaxy sample. Suppose the difference $\boldsymbol{d} - \boldsymbol{r}$ satisfies a Gaussian distribution, so that:

$$P(\boldsymbol{d}|\boldsymbol{r}) = \frac{1}{\sqrt{(2\pi)^m |\mathcal{C}_d|}} \exp\left[-\frac{1}{2}(\boldsymbol{d}-\boldsymbol{r})\mathcal{C}_d^{-1}(\boldsymbol{d}-\boldsymbol{r})^{\rm T})\right] \tag{4.15}$$

where $\mathcal{C}_d$ is the covariance matrix describing the probability distribution of the error between the data and the real quantity – in our example, the data and the real spectrum.

Suppose, now, there is a non-zero *theoretical* error between the prediction of a full model, $\boldsymbol{r}$, which is unknown, and the prediction of the incomplete model actually used in the analysis, $\boldsymbol{t}$. The theoretical error is then $\boldsymbol{e} = \boldsymbol{r} - \boldsymbol{t}$, and will be taken to be a random variable. Then the probability of obtaining the data given the theory *and* the error can be finally written as:

$$\begin{aligned} P(\boldsymbol{d}\,|\,\boldsymbol{t},\boldsymbol{e}) &\equiv P(\boldsymbol{d}\,|\,\boldsymbol{r}=\boldsymbol{t}+\boldsymbol{e}) \\ &= \frac{1}{\sqrt{(2\pi)^m |\mathcal{C}_d|}} \exp\left[-\frac{1}{2}(\boldsymbol{d}-\boldsymbol{t}-\boldsymbol{e})\mathcal{C}_d^{-1}(\boldsymbol{d}-\boldsymbol{t}-\boldsymbol{e})^{\rm T})\right] \end{aligned} \tag{4.16}$$

If one is only interested on the likelihood of the data given the prediction of an incomplete theory, $\boldsymbol{t}(\theta)$, the probability above should be be marginalized over the

error:

$$\begin{aligned}\mathcal{L}(\boldsymbol{d}|\boldsymbol{t}) &= \int d\boldsymbol{e}\, P(\boldsymbol{d}\,|\,\boldsymbol{t},\boldsymbol{e})P(\boldsymbol{e}) \\ &= \frac{1}{\sqrt{(2\pi)^m|\mathcal{C}_d|}} \int d\boldsymbol{e}\; P(\boldsymbol{e}) \exp\left[-\frac{1}{2}(\boldsymbol{d}-\boldsymbol{t}-\boldsymbol{e})\mathcal{C}_d^{-1}(\boldsymbol{d}-\boldsymbol{t}-\boldsymbol{e})^{\mathrm{T}})\right]\end{aligned} \tag{4.17}$$

Suppose, now, the theoretical error satisfies a Gaussian distribution with covariance $\mathcal{C}_e$ and mean $\bar{e}$. Then:

$$\begin{aligned}\mathcal{L}(\boldsymbol{d}|\boldsymbol{t}) = \frac{1}{(2\pi)^m\sqrt{|\mathcal{C}_d|\,|\mathcal{C}_e|}} \int d\boldsymbol{e}\, \exp\Bigg\{ -\frac{1}{2}\Big[(\boldsymbol{d}-\boldsymbol{t}-\boldsymbol{e})\mathcal{C}_d^{-1}(\boldsymbol{d}-\boldsymbol{t}-\boldsymbol{e})^{\mathrm{T}} \\ + (\boldsymbol{e}-\bar{\boldsymbol{e}})\mathcal{C}_e^{-1}(\boldsymbol{e}-\bar{\boldsymbol{e}})^{\mathrm{T}}\Big]\Bigg\}\end{aligned} \tag{4.18}$$

For simplicity of notation, let $\boldsymbol{f} \equiv \boldsymbol{d}-\boldsymbol{t}$. Then the term between square brackets can be written as:

$$\begin{aligned}(\boldsymbol{f}-\boldsymbol{e})\,\mathcal{C}_d^{-1}\,(\boldsymbol{f}-\boldsymbol{e})^{\mathrm{T}} + (\boldsymbol{e}-\bar{\boldsymbol{e}})\,\mathcal{C}_e^{-1}\,(\boldsymbol{e}-\bar{\boldsymbol{e}})^{\mathrm{T}} = \boldsymbol{e}\,(\mathcal{C}_e^{-1}+\mathcal{C}_d^{-1})\,\boldsymbol{e}^{\mathrm{T}} - 2(\boldsymbol{e}\,\mathcal{C}_e^{-1}\,\bar{\boldsymbol{e}}^{\mathrm{T}} \\ + \boldsymbol{e}\,\mathcal{C}_d^{-1}\,\boldsymbol{f}^{\mathrm{T}}) + \boldsymbol{f}\,\mathcal{C}_d^{-1}\,\boldsymbol{f}^{\mathrm{T}} + \bar{\boldsymbol{e}}\,\mathcal{C}_e^{-1}\,\bar{\boldsymbol{e}}^{\mathrm{T}}\end{aligned} \tag{4.19}$$

where we have used $\mathcal{C}_d^{\mathrm{T}} = \mathcal{C}_d$ and $\mathcal{C}_e^{\mathrm{T}} = \mathcal{C}_e$. The expression above can be cast into a simpler form by a procedure analogous to the completing of squares. It must be identified with:

$$(\boldsymbol{u}-\boldsymbol{a})\,M^{-1}\,(\boldsymbol{u}-\boldsymbol{a})^{\mathrm{T}} - b = \boldsymbol{u}\,M^{-1}\,\boldsymbol{u}^{\mathrm{T}} - 2\boldsymbol{u}\,M^{-1}\,\boldsymbol{a}^{\mathrm{T}} + \boldsymbol{a}\,M^{-1}\,\boldsymbol{a}^{\mathrm{T}} - b \tag{4.20}$$

with suitable variables $\boldsymbol{u}$, $\boldsymbol{a}$, b, and M and assuming $M^{\mathrm{T}} = M$. Identification of terms between (4.19) and (4.20) yields the relations:

$$\begin{aligned}M^{-1} &= \mathcal{C}_d^{-1} + \mathcal{C}_e^{-1} \\ \boldsymbol{u} &= \boldsymbol{e} \\ \boldsymbol{a} &= (\boldsymbol{f}\mathcal{C}_d^{-1} + \boldsymbol{e}\mathcal{C}_e^{-1})(\mathcal{C}_d^{-1}+\mathcal{C}_e^{-1})^{-1} \\ \boldsymbol{b} &= -(\boldsymbol{f}-\bar{\boldsymbol{e}})\,(\mathcal{C}_d+\mathcal{C}_e)^{-1}\,(\boldsymbol{f}-\bar{\boldsymbol{e}})^{\mathrm{T}}\end{aligned} \tag{4.21}$$

Then (4.18) can be written as:

$$\begin{aligned}\mathcal{L}(\boldsymbol{d}|\boldsymbol{t}) &= \frac{1}{(2\pi)^m\sqrt{|\mathcal{C}_d|\,|\mathcal{C}_e|}}\int d\boldsymbol{u}\,\exp\left\{-\frac{1}{2}\left[(\boldsymbol{u}-\boldsymbol{a})\,M^{-1}\,(\boldsymbol{u}-\boldsymbol{a})^{\mathrm{T}}-b\right]\right\}\\ &= \frac{1}{(2\pi)^m\sqrt{|\mathcal{C}_d|\,|\mathcal{C}_e|}}\exp\left(\frac{1}{2}b\right)\int d\boldsymbol{u}\,\exp\left\{-\frac{1}{2}\left[(\boldsymbol{u}-\boldsymbol{a})\,M^{-1}\,(\boldsymbol{u}-\boldsymbol{a})^{\mathrm{T}}\right]\right\}\\ &= \frac{1}{(2\pi)^m\sqrt{|\mathcal{C}_d|\,|\mathcal{C}_e|}}\exp\left(\frac{1}{2}b\right)\left(\sqrt{(2\pi)^m|M|}\right)\\ &= \frac{1}{\sqrt{(2\pi)^m|\mathcal{C}_T|}}\exp\left\{-\frac{1}{2}\left[(\boldsymbol{f}-\bar{\boldsymbol{e}})\,\mathcal{C}_T^{-1}\,(\boldsymbol{f}-\bar{\boldsymbol{e}})^{\mathrm{T}}\right]\right\}\end{aligned}$$

with $\mathcal{C}_T = \mathcal{C}_d + \mathcal{C}_e$. Hence [6]:

$$\boxed{\mathcal{L}(\boldsymbol{d}|\boldsymbol{t}) = \frac{1}{\sqrt{(2\pi)^m|\mathcal{C}_T|}}\exp\left\{-\frac{1}{2}\left[(\boldsymbol{d}-\boldsymbol{t}-\bar{\boldsymbol{e}})\,\mathcal{C}_T^{-1}\,(\boldsymbol{d}-\boldsymbol{t}-\bar{\boldsymbol{e}})^{\mathrm{T}}\right]\right\}} \tag{4.22}$$

After the assumption of a Gaussian theoretical error has been made, only $\mathcal{C}_e$ and $\bar{\boldsymbol{e}}$ remain to be specified for the likelihood in (4.22) to be characterized completely. Here some liberty of choice is inevitable, and akin to the freedom commonly found in the specification of parameter priors. We thus proceed in this way: we stipulate two reasonable choices, from a theoretical point of view, for $\mathcal{C}_e$ and $\bar{\boldsymbol{e}}$, and then put them to the test by analysing mock data with the likelihood (4.22). We note this procedure has the same logical basis as the ones commonly adopted in the choice of k_{max} for a scale cut [19].

Following [6], we will begin by writing the covariance matrix $\mathcal{C}_e$ as:

$$\mathcal{C}_e^{\ell\ell'} = E_\ell E_{\ell'}\rho_{\ell\ell'} \tag{4.23}$$

with $\rho_{\ell\ell} \equiv 1$ for all ℓ. The function E_ℓ, named the *envelope*, defines the diagonal elements of the covariance matrix and thus the variance of the theoretical error. It can be taken as a measure of the overall size of theoretical uncertainty at a given mode ℓ. The coefficients $\rho_{\ell\ell'}$, on the other hand, specify the correlation between the theoretical error at different modes. An obvious restriction on $\rho_{\ell\ell'}$ is that $\mathcal{C}_e$ must be positive-definite and symmetric, $\mathcal{C}_e^{\ell\ell'} = \mathcal{C}_e^{\ell'\ell}$.

The uncertainty on galaxy bias can be assessed through simulations. This is analogous to the procedure adopted by [23] to assess theoretical uncertainty on baryonic-feedback models, in this case based on a set of 13 hydrodynamic

simulations assumed to be representative of the phenomena of interest. Likewise, [15], for example, used LasDamas N-body simulations [22] to quantify the uncertainty on the one-loop perturbative model of dark matter and galaxies. In this work, we will use results of [7] on the relations between parameters b_2 and b_{s^2} and b_1. These were obtained by fitting the nonlinear bias model to a set of IllustrisTNG simulations. The best-fit values of b_2 and b_{s^2} are given in terms of b_1 as:

$$\begin{aligned} b_2 &= 0.24\, b_1^3 + 1.12\, b_1^2 - 0.44\, b_1 - 0.34 \\ b_{s^2} &= -0.28\, b_1 + 0.09 \end{aligned} \tag{4.24}$$

with standard deviations of

$$\sigma_2 = 0.44, \quad \sigma_{s^2} = 0.22 \tag{4.25}$$

independently of b_1. Since the authors of [7] quoted these relations in the Eulerian frame, we used the appropriate transformations at second order (for example, [1]) to express them in the Lagrangian frame, which is adopted in this work. These are:

$$\begin{aligned} b_1^{(\mathrm{E})} &= b_1^{(\mathrm{L})} + 1 \\ b_2^{(\mathrm{E})} &= \frac{4}{21} b_1^{(\mathrm{L})} + \frac{1}{2} b_2^{(\mathrm{L})} \\ b_{s^2}^{(\mathrm{E})} &= -\frac{2}{7} b_1^{(\mathrm{L})} \end{aligned} \tag{4.26}$$

which can be inverted to yield $b_i^{(\mathrm{L})}$ in terms of $b_j^{(\mathrm{E})}$.

As to the relation between the best-fit $b_{\nabla^2\delta}$ and b_1, which is not provided by [7], we use a result of [30] based on BACCO N-body simulations [3] and a subhalo-abundance matching technique for the creation of galaxy catalogs. The relation thus found is:

$$b_{\nabla^2\delta} = 0.2298\, b_1^3 - 2.096\, b_1^2 + 0.7816\, b_1 - 0.1545 \tag{4.27}$$

already in terms of Lagrangian parameters.

The authors of [30] report 100% of their $b_{\nabla^2\delta}$ samples to be contained in the interval $[b_{\nabla^2\delta}^{\mathrm{best-fit}} - 5, b_{\nabla^2\delta}^{\mathrm{best-fit}} + 8]$. To interpret that range in terms of a Gaussian distribution, we assume it equal to two times the 3σ range, or the 99.78% confidence interval, which implies:

$$\sigma_{\nabla^2\delta} = 2.17 \tag{4.28}$$

We will show this modelling choice to be sufficient for the recovery of the fiducial cosmology in an analysis with theoretical error.

The definitions of envelope and correlation coefficients adopted in this work will be described in section 4.3. We emphasize that the stipulations on the form of $P(\boldsymbol{e})$ made here will be put to the test in a standard way – by applying them to mock data. On a logical basis, they can be considered analogous to the various hypotheses entering the choice of the $k_{\rm max}$ of a scale cut, which are afterwards put to the test in the same way.

4.3 Models of theoretical-error covariance

In this section we discuss in more detail the models of theoretical-error likelihood adopted in this work. As already noted in section 4.2, the probability of theoretical error will be completely specified by three elements: (i) the *envelope*, E_ℓ, which controls the diagonal elements of the covariance matrix; (ii) the correlation coefficients, $\rho_{\ell\ell'}$, which define its non-diagonal elements; and (iii) the mean error, $\bar{\boldsymbol{e}}$, which enters the modified likelihood (4.22).

4.3.1 Correlation coefficients

We begin with the specification of $\rho_{\ell\ell'}$, which will have the same form irrespective of the choices of envelope and mean error. Since these coefficients determine the correlation between theoretical errors at different ℓ-modes – i.e., a correlation length –, they will be determined by the typical scale of variations of the spectra in question. This will be the scale of baryonic acoustic oscillations (BAOs) corresponding to each redshift bin; $\Delta\ell_{\rm BAO} \approx \chi(z_c)\Delta k_{\rm BAO}$, with $\Delta k_{\rm BAO} \approx 0.05h\,{\rm Mpc}^{-1}$ [6] and z_c, the central redshift of the given bin. This gives $\Delta\ell_{\rm BAO} \approx 50$ at $z_c \approx 0.5$. Furthermore, we define $\rho_{\ell\ell'}$ in terms of a well-behaved function that guarantees $\mathcal{C}_e$ to be positive-definite:

$$\rho_{\ell\ell'} = \exp\left[-\frac{(\ell-\ell')^2}{2\Delta\ell^2_{\rm BAO}}\right] \tag{4.29}$$

We also assume no correlation between theoretical errors in different redshift bins. This can be thought as a conservative choice, since adding such correlation

would increase the final precision on the estimated parameters. We will show it to be sufficient, however, for a significant gain in precision with respect to the scale cut.

4.3.2 Envelope

As to the envelope in (4.23), we will propose two definitions. A first choice, named Envelope 1, will be the standard deviation of the error $\boldsymbol{e} = \boldsymbol{d} - \boldsymbol{t}$ when (i) $\boldsymbol{d}$ is calculated with nonlinear-bias parameters satisfying the distributions (4.24) and (4.27), given a choice of b_1, and (ii) $\boldsymbol{t}$ is the corresponding best-fit model under linear bias, i.e., with only one b_1 parameter per redshift bin. Our routine for the computation of the best-fit model will be similar to that adopted in [15]. Specifically:

1. The linear-bias vector, $\boldsymbol{t}$, will be fitted to the data vector, $\boldsymbol{d}$, using all ell-bins from 20[2] to a certain variable $\ell_{\rm max}$. We will note the χ^2 statistic of the fit as a function of that maximum value, $\chi^2(\ell_{\rm max})$. Typically, there is a "breaking-point" $\ell^{\rm bp}$ after which χ^2 rises abruptly, indicating that the linear-bias model begins to fail. We will choose the fit of $\boldsymbol{t}$ made with an $\ell_{\rm max}$ smaller than $\ell^{\rm bp}$, that is, $\ell^{\rm fid}_{\rm max} = \ell^{\rm bp}_{\rm max} - \Delta\ell$, for some $\Delta\ell$. Since the specific value of the latter is not important – provided it is larger than zero, so the theoretical prediction is made on a valid range of scales – we accord with [15] by setting $\Delta\ell = 0.02/\chi(z_p)$, where $\chi(z_p)$ is the radial distance corresponding to the central value of a redshift bin.

2. We will vary the data vector by varying the nonlinear-bias parameters. These will be sampled according to the Gaussian distributions measured in simulations and specified by (4.24), (4.27), (4.25), and (4.28). For each such vector[3], we find the best-fitting linear-bias model in ℓ-bins between 20 and $\ell^{\rm bp}_{\rm max}$. We then compute the error $\boldsymbol{e} = \boldsymbol{d} - \boldsymbol{t}$, whose standard deviation is, by definition, Envelope 1. For this choice of envelope, moreover, we define $\bar{\boldsymbol{e}}$ as the average $\boldsymbol{e}$ in the distribution so generated. The range of relative errors on $C_\ell^{(gg)}$ and $C_\ell^{(g\kappa)}$ is shown in figures 4.3 and 4.4.

A second choice of envelope will be based on a looser stipulation. The error $\boldsymbol{e} = \boldsymbol{d} - \boldsymbol{t}$ will be computed in what is deemed to be a typical case of nonlinear

[2]The minimum ℓ is set at 20 to accord with the LSST-DESC conventions [16].

[3]We found the envelope to become stable – i.e., change little with each re-evaluation – when that computation is made with 150 samples.

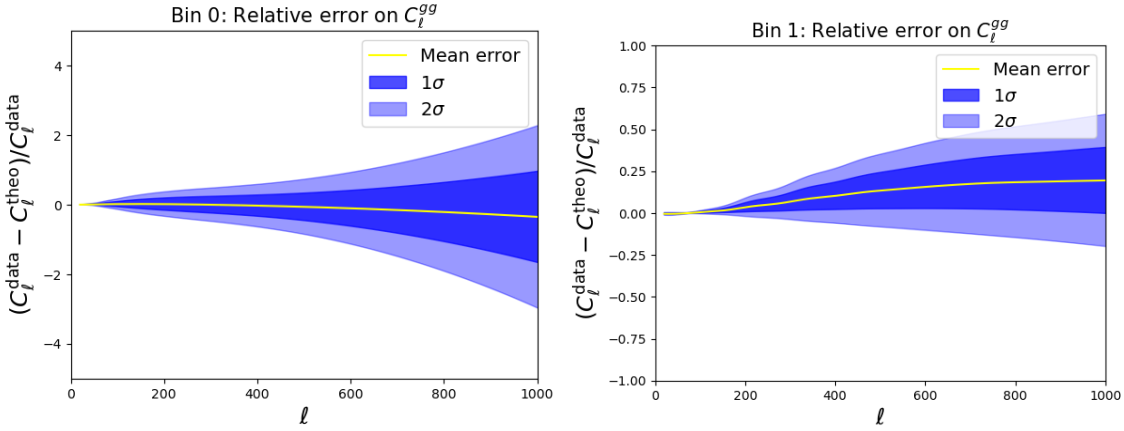

Figure 4.3: Distribution of relative errors on $C_\ell^{(gg)}$.

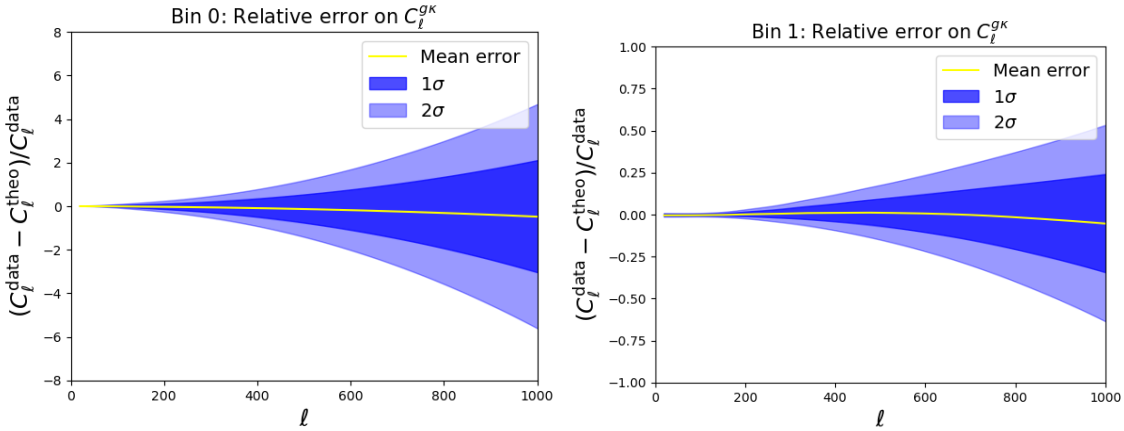

Figure 4.4: Distribution of relative errors on $C_\ell^{(g\kappa)}$.

Name	Envelope	Mean error
Envelope 1	Standard deviation of $\boldsymbol{e}(\theta) = \boldsymbol{d}(\theta) - \hat{\boldsymbol{t}}$	Average of $\boldsymbol{e}(\theta) = \boldsymbol{d}(\theta) - \hat{\boldsymbol{t}}$
Envelope 2	"Typical error": $\boldsymbol{e}(\bar{\theta}) = \boldsymbol{d}(\bar{\theta}) - \hat{\boldsymbol{t}}$	Same as the envelope

Table 4.1: Definition of envelopes and mean errors. Envelope 1 is the standard deviation of the error when parameters $\theta = \{b_2, b_{s^2}, b_{\nabla^2}\}$ satisfy the Gaussian distributions specified by (4.24), (4.25), (4.27), and (4.28). Envelope 2 is the error when these parameters assume the central values $\bar{\theta} = \{\bar{b}_2, \bar{b}_{s^2}, \bar{b}_{\nabla^2}\}$ given by (4.24) and (4.27). $\hat{\boldsymbol{t}}$ denotes the best-fit theoretical vector.

bias – namely, when parameters b_2, b_{s^2}, and $b_{\nabla^2\delta}$ satisfy the central values (4.24 and 4.27) found in simulations. We call that error Envelope 2, taken as the typical size of the uncertainty on nonlinear bias. We also consider it equal to the mean error $\bar{\boldsymbol{e}}$. Although seemingly less well-justified than the former envelope, we will show this proposal yields desirable results when applied to mock data.

A summary of the two definitions of envelope adopted in this work, with the accompanying definitions of the mean error, is given on table 4.1. They are depicted in figures 4.6 and 4.7.

The inclusion of theoretical error in the covariance matrix effectively increases the size of error bars, allowing the estimation to be performed on higher ℓ-modes. This is illustrated in figure 4.5, which shows χ-by-eye plots of estimations with Envelopes 1 and 2, respectively.

The augmented covariance matrix, $\mathcal{C}_T = \mathcal{C}_d + \mathcal{C}_e$, corresponding to the fiducial cosmology and the second choice of envelope, is shown on the right-hand side of figure 4.8, for two redshift bins. One sees the usual covariance matrix $\mathcal{C}_d$ on the left panel of the same figure.

4.4 Likelihood analysis

In this section, we discuss the methodology to be used in the analyses below. In particular, we seek to establish a common frame for comparison of results obtained with a scale cut and with theoretical error. As mock data vectors, we will use angular spectra of galaxy clustering and galaxy-galaxy lensing at two redshift bins, according to LSST's prescriptions summarized in section 4.1. These spectra will be generated in accordance with the nonlinear-bias model at one-loop order.

To perform comparisons in a range of realistic conditions of bias, we will use

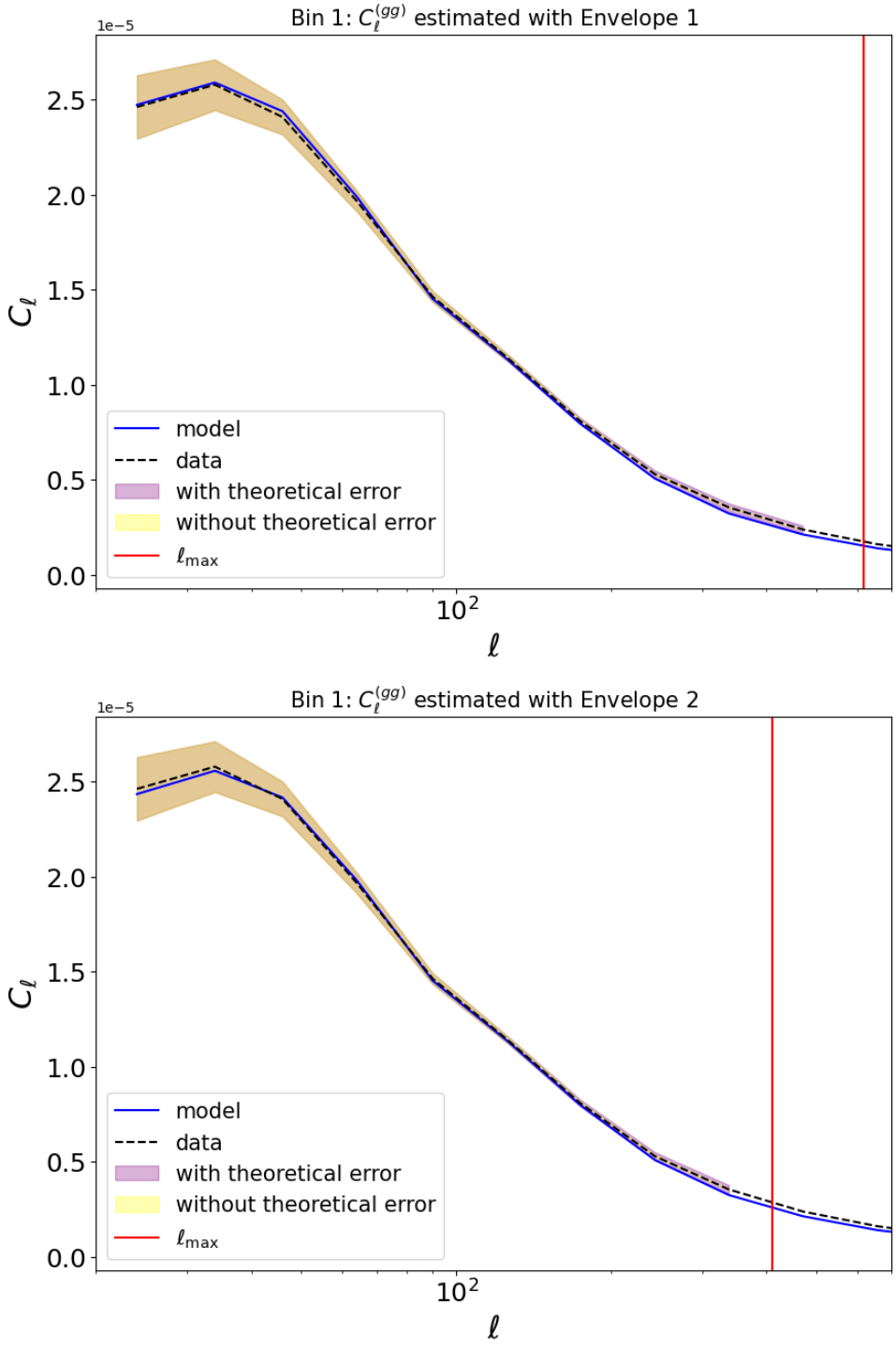

Figure 4.5: Upper and lower panels show plots of estimations with Envelopes 1 and 2, respectively. The inclusion of theoretical error increases the size of the error bars (purple shade), allowing larger deviations between data and model vectors to be accommodated. This is specially valid on small scales, where such deviations become significant. In contrast, the error bars without theoretical error (yellow shade) virtually fade for $\ell \gtrsim 200$.

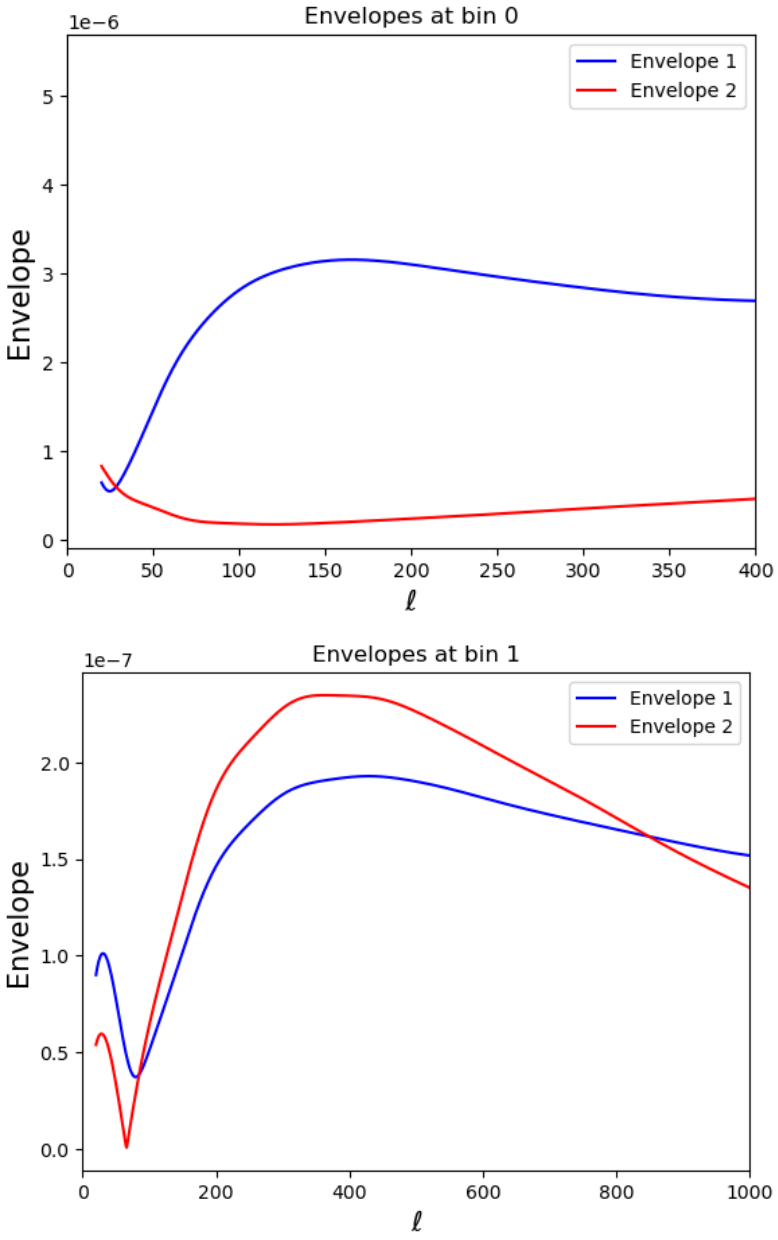

Figure 4.6: Envelopes of galaxy clustering at bins 0 and 1. Definitions are given on table 4.1.

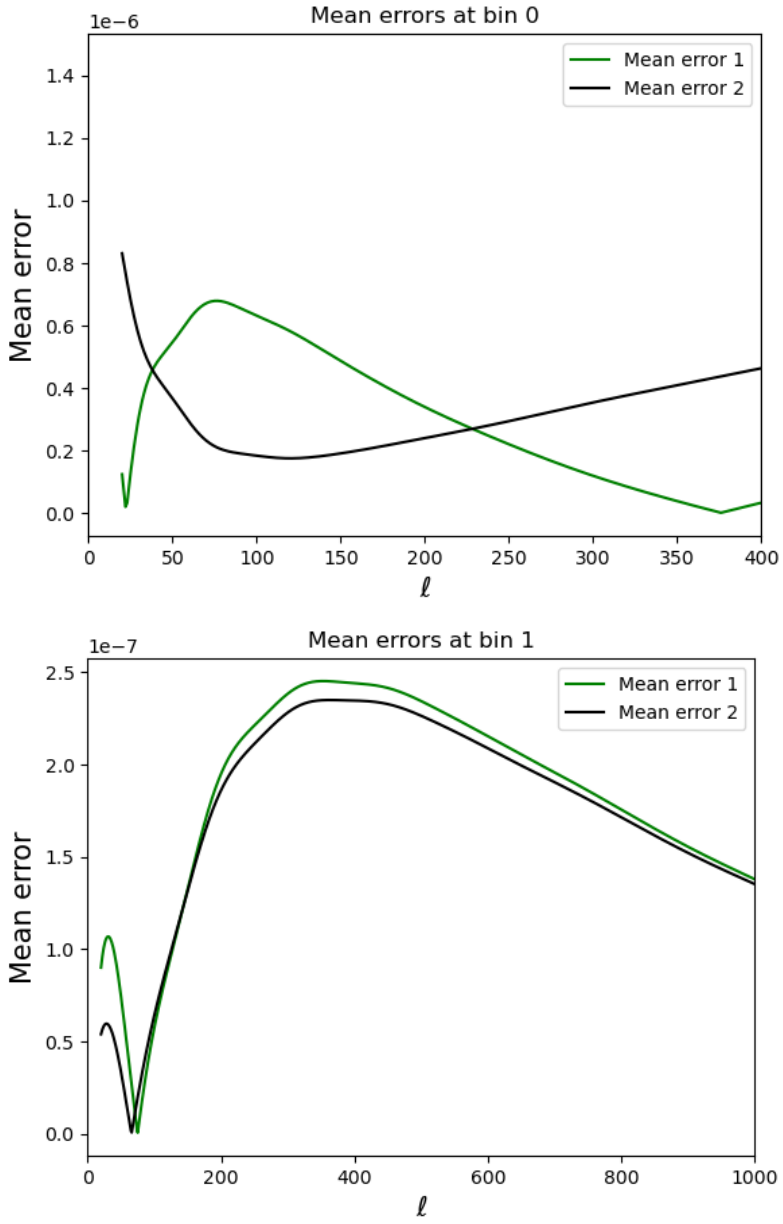

Figure 4.7: Mean errors at bins 0 and 1, in absolute values. Definitions are given on table 4.1.

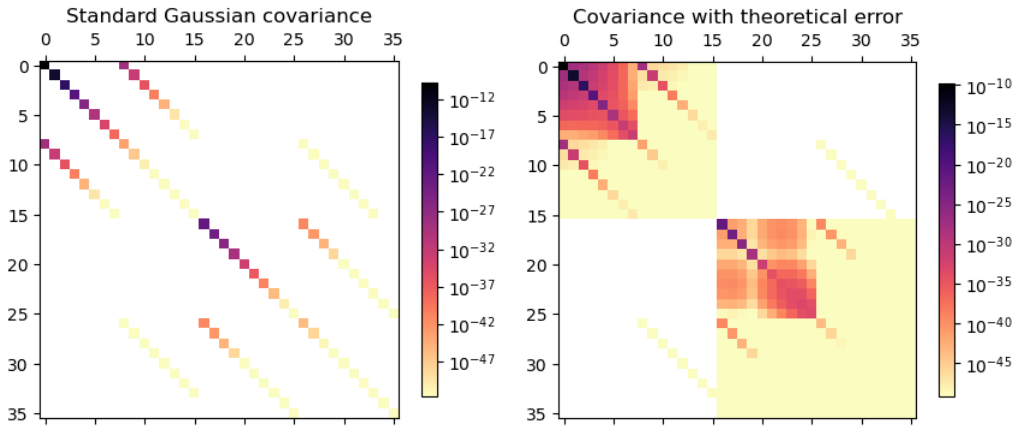

Figure 4.8: Left panel shows the usual cosmic covariance matrix for two redshift bins. Right panel shows the augmented matrix $\mathcal{C}_T = \mathcal{C}_d + \mathcal{C}_e$, with non-diagonal terms representing the correlation between theoretical errors on ℓ-modes of the same bin. We adopt the ℓ-binning prescribed by LSST-DESC and calculate $\ell^i_{\rm max}$ from $k_{\rm max} = 0.3h\,{\rm Mpc}^{-1}$.

three sets of data vectors differing in the strength of nonlinear bias. This is to avoid that we misrepresent the effectiveness of one method because of an exceptionally favorable (or unfavorable) data vector – e.g., a scale cut is expected to perform better in cases of weak nonlinear bias. That this care must be taken is exemplified by the results of [23], where the inclusion of theoretical error led to significant gains in situations of moderate-to-strong baryonic feedback, but was ineffectual (although harmless) when applied to models where these effects were particularly weak – e.g., the MassiveBlack-II model.

For the main data vector used for testing, we choose parameter values similar to those used to stress-test scale cuts in the Dark Energy Survey [19]. Our data vectors are defined as:

1. Typical bias: The parameter b_1 has values 0.7 and 1.0 in bins 0 and 1, respectively, which are slightly larger than usual for such redshifts ($z_p \sim 0.3$ and $z_p \sim 0.8$). These values (rather, their counterparts in the Eulerian frame) are used by DES to validate scale cuts [19]. The parameter b_2 used in DES, and also in this work, is interpolated from the $b_2 - b_1$ relation measured from catalogs of redMagic-like galaxies [25]. b_{s^2} is computed from b_1 through the co-evolution relation (in the Eulerian frame) $b_{s^2} = -4(b_1 - 1)/7$. The parameter b_{∇^2}, which is absent in the DES analysis, is computed from b_1 from the best-fit value (4.27).

Data vector	$\langle z \rangle = 0.24$				$\langle z \rangle = 0.86$			
	b_1	b_2	b_{s^2}	b_{∇^2}	b_1	b_2	b_{s^2}	b_{∇^2}
Typical bias	0.70	0.190	-0.200	-0.556	1.00	0.620	-0.280	-1.239
Low bias	0.70	-0.118	-0.354	-2.075	1.00	0.312	-0.434	-2.758
High bias	0.70	0.498	-0.046	0.963	1.00	0.928	-0.126	0.280

Table 4.2: Bias parameters defining the three pairs of data vectors used in the analyses below. "Typical bias" has the same set of b_1, b_2, and b_{s^2} values as the vector used for validation in DES Year 3 [19], when translated to the Lagrangian frame. The value of b_{∇^2} is computed through (4.27), given b_1. The parameters entering "low bias" and "high bias" are obtained from the first set by applying deviations of $-0.7\sigma_i$ and $+0.7\sigma_i$, respectively, where σ_i is provided in (4.25) and (4.28).

2. High bias: All nonlinear-bias parameters (i.e., b_2, b_{s^2}, and b_{∇^2}) are 0.7-σ deviated upwardly, according to the standard deviations (4.25) and (4.28). We have checked that this results in $C_\ell^{(gg)}$ and $C_\ell^{(g\kappa)}$ nearly one standard deviation stronger than the central spectra – that is, those computed with the central parameters (4.24) and (4.27). This fulfills our requisite of strongly-biased, but still realistic data vectors.

3. Low bias: All nonlinear-bias parameters are 0.7-σ deviated downwardly.

The bias parameters defining each data vector are presented on table 4.2. We emphasize that the "high bias" and "low bias" terminology refers to the signed value of the bias parameters, not to their absolute values. For example, for bin 1, the modulus $|b_{s^2}|$ is larger in the "low bias" than in the "high bias" case (table 4.2). In this sense, both the "low bias" and "high bias" vectors can be said to represent cases of strong nonlinear bias.

The fiducial cosmology is specified on table 4.3. We will take Ω_c and σ_8 as the only free cosmological parameters, whereas the linear bias b_1 at all redshift bins will play the role of nuisance parameters. We assume uninformative priors on both cosmological and nuisance parameters. The evolution of $b_1(z)$ within a redshift bin will be neglected, an approximation that [19] has found satisfactory.

Implicit in the considerations below is that theoretical errors on galaxy bias are fairly independent of the fiducial cosmology: the envelope is adjusted with mock data set at the chosen cosmology to be later applied to data describing the actual universe. One commonly assumes the same independence when computing the $k_{\max}$ of a scale cut [26].

Parameter	Fiducial value
Ω_c	0.266
Ω_b	0.05
Ω_k	0
σ_8	0.831
h	0.67
n_s	0.9645

Table 4.3: Fiducial cosmology. Only Ω_c and σ_8 (besides b_1) are free parameters in the estimations below, and uninformative priors are assumed.

The likelihood used in the analyses with the scale cut is Gaussian and specified by the covariance matrix $\mathcal{C}_d$ given in (4.13). The modified expression used in analyses with theoretical error is that given by (4.22), where $\mathcal{C}_e$ and $\bar{\boldsymbol{e}}$ depend on the model of theoretical error adopted. In both cases, one needs to choose a set of $\ell_{\rm max}^{(i)}$ for the analyses. We reduce the range of possibilities by fixing $\ell_{\rm max} = k_{\rm max}\chi(z_p)$, so that all $\ell_{\rm max}^{(i)}$ are defined from a single choice of $k_{\rm max}$.

The quality of results of one method depends on more than one factor: e.g., estimation bias, overall uncertainty, or quality of the fit. To facilitate comparisons, we will establish a common frame: the values of $k_{\rm max}$ will be adjusted so that both methods can recover the fiducial cosmology (Ω_c and σ_8) with a tension near 0.3σ in the marginalized, two-dimensional posterior. Fixing the desired tension allows a simple criterion for the evaluation of relative effectiveness – namely, if two estimation methods yield the same tension, the one which produces tighter constraints will be deemed more effective. Because of the large granularity of the ℓ-binning of LSST – where only 20 ℓ-bins are used –, it won't always be possible, however, to select values of $k_{\rm max}$ yielding a tension of exactly 0.30σ with both methods. In that case, we will chose a level close to 0.30σ where that equality can be approximately attained.

One may wonder why we still need to specify a $k_{\rm max}$ (or, equivalently, a set of $\ell_{\rm max}^{(i)}$ values) in the estimation with theoretical error. Ideally, theoretical uncertainty, as expressed in the envelope, would rise steadily with the growth of $\ell_{\rm max}$ and become very great on scales at which the linear-bias model is known to fail, somewhere around $k = 0.2h\,\mathrm{Mpc}^{-1}$. Such an envelope would lead to a smooth, automatic suppression of modes where the model is unreliable, and one wouldn't have the trouble of specifying $k_{\rm max}$. Although such a behavior is not hard to conceptualize, we haven't observed it in neither of the envelopes discussed in section 4.3. Specifically, they don't increase quickly enough with ℓ, so that even ℓ

modes corresponding to k values beyond $0.3h\text{Mpc}^{-1}$ have significant weight in the statistical analysis and still contribute to estimation error. Hence, they have to be removed by the *ad hoc* procedure of specifying a k_{max}.

To see more clearly that large k-modes still contribute to the analysis with theoretical uncertainty, one may check the behavior of the figure of merit (FoM) of the posterior with respect to k_{max}. That statistic is defined as:

$$\text{FoM} \equiv \frac{1}{\sqrt{\det[S_{\alpha\beta}/\theta_{\text{fid},\alpha}\theta_{\text{fid,fi}}]}} \tag{4.30}$$

where $S_{\alpha\beta}$ is the covariance matrix of the posterior, marginalized over the nuisance parameters $b_1^{(i)}$, and $\theta_{\text{fid},\alpha}$ is the fiducial value of parameter θ_α, either Ω_c or σ_8. The FoM is a measure of the estimation's precision. We define, in addition, the *figure of bias* as:

$$\text{FoB} \equiv \left[\sum_{\alpha,\beta}(\bar{\theta}_\alpha - \theta_{\text{fid},\alpha})S^{-1}_{\alpha\beta}(\bar{\theta}_\beta - \theta_{\text{fid},\beta})\right]^{1/2} \tag{4.31}$$

which gives a measure of estimation error with respect to the overall uncertainty. Here, $\bar{\theta}_\alpha$ is the mean value of θ_α.

If large k-modes didn't contribute to the estimation, one would expect the functions $\text{FoM}(k_{\text{max}})$ and $\text{FoB}(k_{\text{max}})$ to saturate for sufficiently-high k_{max}. This would mean that the statistical weight of large k-modes was automatically suppressed by the envelope, so that an increase in k_{max} would have no effects on either estimation bias or the overall uncertainty. We haven't observed this behavior with neither of the envelopes defined in section 4.3, however. Figure 4.9 presents the graphs of $\text{FoM}(k_{\text{max}})$ and $\text{FoB}(k_{\text{max}})$ in an analysis with Envelope 1, for illustration. That they show no sign of saturation demonstrates that the envelope can't be relied upon to automatically suppress modes of small scales, so a k_{max} must be chosen in the analysis with theoretical error as well. One can imagine the envelope growing effectively to infinity at that point.

The necessity of fixing a k_{max} eliminates one of the expected advantages of the theoretical-error method with respect to the scale cut – namely, that one wouldn't need to adjust the choice of k_{max} by running a long stream of MCMC analyses. We will still argue, however, that at that same cost one can obtain increased benefits – i.e., tighter constraints on cosmological parameters resulting from a closer-to-optimal inclusion of modes in the analysis.

As regards practical implementation, we have used Velocileptors [12], in

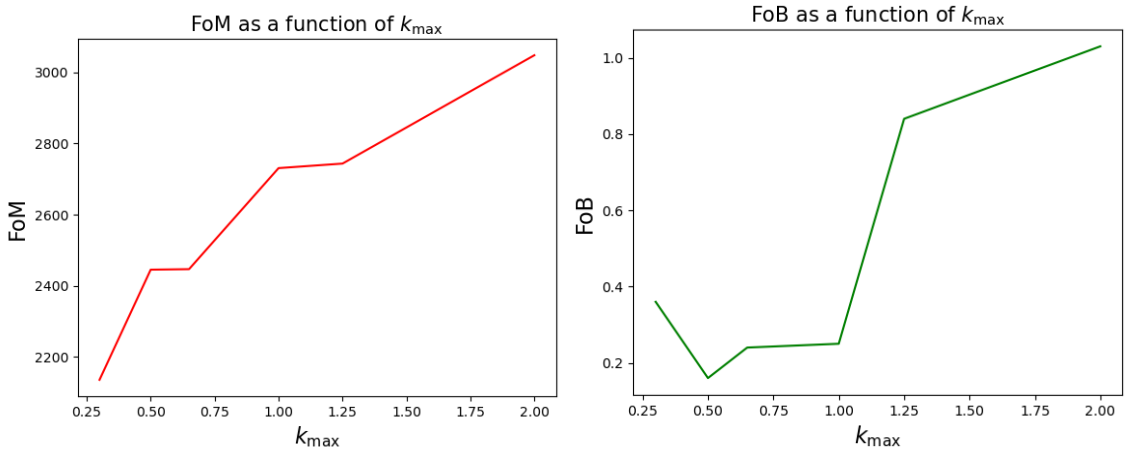

Figure 4.9: Figure of merit and figure of bias of an estimation with theoretical error as functions of k_{max}. The curves show no sign of saturation as k_{max} increases towards $2h\,\mathrm{Mpc}^{-1}$, which indicates that one can't count on the envelope alone to suppress large k-modes. The case above is of Envelope 1 and the typical-bias data vector.

its RKECLEFT formulation, to compute the biased power spectra $P(k)$ at one-loop order in the Lagrangian description. These were projected on angular spectra through the DESC Core Cosmology Library [13]. The evaluation of likelihoods was based on the Monte Carlo Markov Chain sampler of Cobaya [28, 27]. It was enabled by resources supplied by the Center for Scientific Computing (NCC/GridUNESP) of Sao Paulo State University (UNESP).

4.5 Results

Results of estimations with Envelope 1 are shown in figure 4.10. These were obtained with mock data vectors representing the three cases of nonlinear bias – "typical", "high", and "low" – defined in section 4.4. The mean-error vector entering equation (4.22) is, in this case, computed as the actual average of the theoretical error when nonlinear bias parameters satisfy the prescribed distributions, as noted in section 4.3.

To obtain comparable tensions on the estimated parameters, the analyses with theoretical error (TE) and the scale cut were made with k_{max} values of $0.3h\,\mathrm{Mpc}^{-1}$ and $0.08h\,\mathrm{Mpc}^{-1}$, respectively. Small tensions were then found with both methods: 0.11 with TE and 0.09 with the scale cut. With the "typical bias" data vector, one obtains a considerable improvement of TE upon the scale cut, with a change of

Data vector	Theoretical error			Scale cut		
	FoM	FoB	Tension	FoM	FoB	Tension
Typical bias	2007	0.38	0.11	743	0.49	0.09
High bias	1955	1.28	0.71	703	0.85	0.39
Low bias	2046	0.64	0.28	778	0.31	0.02

Table 4.4: Results with Envelope 1 and the scale cut. Estimations were made with $k_{\max}$ equal to $0.3h\,\mathrm{Mpc}^{-1}$ and $0.08h\,\mathrm{Mpc}^{-1}$, respectively. TE performs significantly better in terms of FoM for all test vectors, and yields a tension equivalent to the scale cut's in the analysis with "typical bias". As for robustness – expressed in the results with "high bias" and "low bias" –, both methods can recover the fiducial cosmology with less than 1-σ tension, but the scale cut has better performance.

170% in the FoM. This is due mainly to an increased precision on σ_8.

To test robustness, we performed analyses with other two vectors – "large" and "low bias" – representing deviations from "typical bias". (We recall that "high" and "low" refer to the signed value (not the modulus) of bias parameters.) In all cases, both methods were able to recover the fiducial parameters with less than 1σ tension. TE had slightly worse performance in this regard, as its tension rose from 0.11 to 0.71 when the "high bias" vector was given as input, whereas the tension yielded by the scale cut rose only from 0.09 to 0.39. These results are summarized on table 4.4.

Results obtained with Envelope 2 are shown in figure 4.11. Here, the mean-error vector is equal to the error under typical bias conditions – that is, when nonlinear bias parameters assume the central values (4.24) and (4.27). To obtain comparable tensions, we have used $k_{\max} = 0.2h\,\mathrm{Mpc}^{-1}$ for TE and $k_{\max} = 0.1h\,\mathrm{Mpc}^{-1}$ for the scale cut. We note that, due to the smaller size of Envelope 2 with respect to Envelope 1, the estimation could not be extended to the previous limit of $0.3h\,\mathrm{Mpc}^{-1}$ without introducing considerable bias in the results. This illustrates the need of choosing a $k_{\max}$ with some care even in estimations with theoretical error.

In terms of relative effectiveness, the situation is similar to that seen with Envelope 1. In the analysis with "typical bias", the inclusion of theoretical error led to considerable improvement in precision, with a gain of 52% in the FoM and a smaller tension.

As for robustness, both methods – TE with Envelope 2 and the scale cut – perform badly under "high bias", yielding tensions higher than 1σ. In the "low bias" case, they could recover the fiducial parameters with less than 1-σ tension,

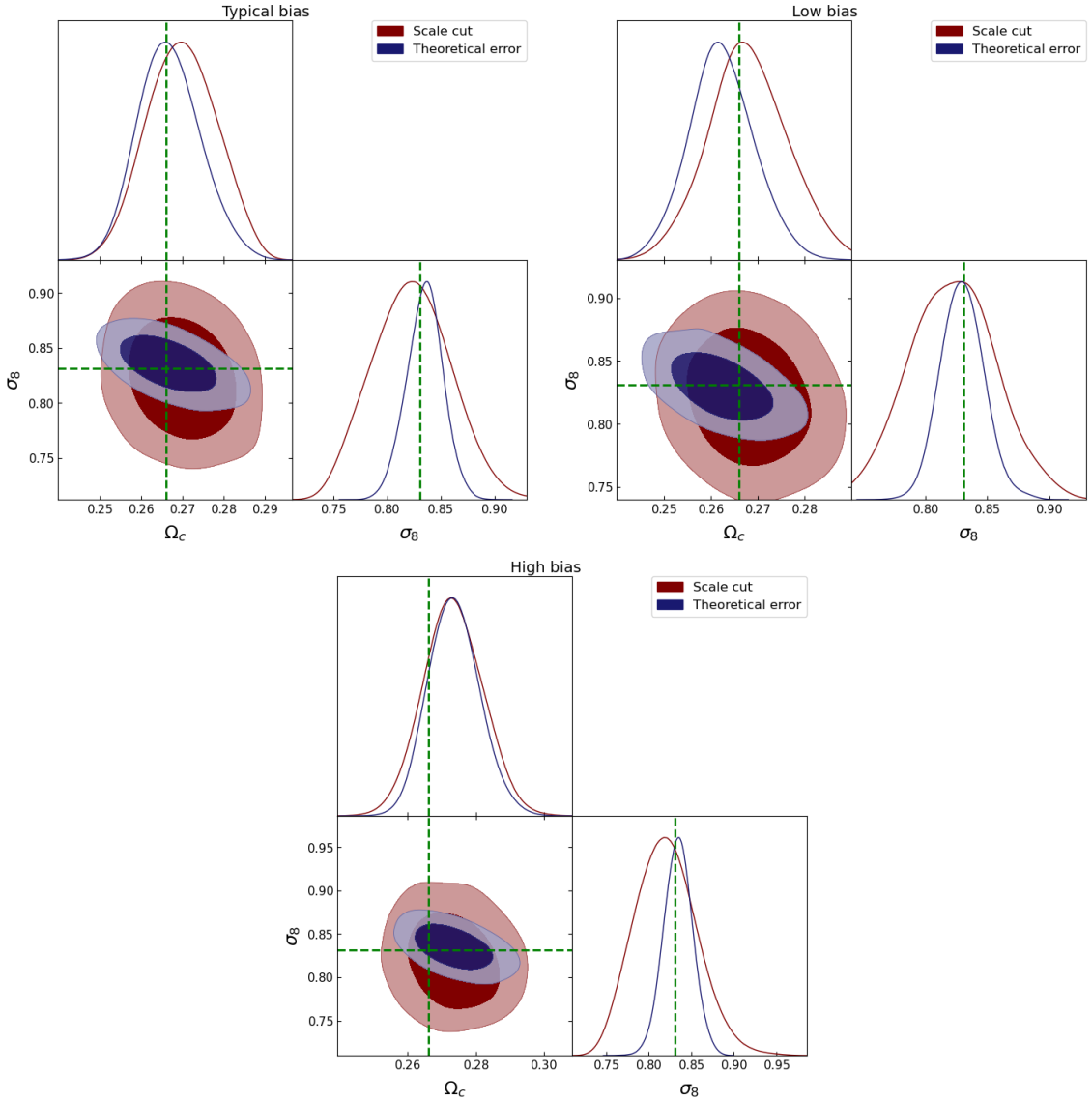

Figure 4.10: Results of the estimation with Envelope 1. Upper left, upper right, and lower panels show results obtained with data vectors representing typical, low, and high galaxy bias, respectively. Details are provided on table 4.4.

Data vector	Theoretical error			Scale cut		
	FoM	FoB	Tension	FoM	FoB	Tension
Typical bias	2090	0.73	0.26	1379	1.15	0.68
High bias	1857	2.35	1.85	1344	1.94	1.37
Low bias	2077	1.03	0.58	1304	0.56	0.16

Table 4.5: Results with Envelope 2 and the scale cut. Estimations were made with $k_{\max}$ equal to $0.2h\,\mathrm{Mpc}^{-1}$ and $0.1h\,\mathrm{Mpc}^{-1}$, respectively. As with Envelope 1, TE performs significantly better in terms of FoM for all data vectors, and yields a tension equivalent to the scale cut's in the analysis with the "typical bias" data vector. In the test with the "high bias" data vector, neither of the methods can recover the fiducial parameters with less than 1σ tension. In the "low bias" case they yield comparable tensions, but the scale cut shows better performance.

with the scale cut having a slightly better performance: its tension fell from 0.68 to 0.16, whereas TE's tension rose from 0.26 to 0.58.

4.6 Conclusion

We have tested a method of handling data for which the available models are unreliable. Two probes dependent on the galaxy overdensity have been considered: the power spectra of galaxy clustering and galaxy-galaxy lensing. The angular spectra have been computed in two redshift bins with the galaxy distributions prescribed for the year 10 of LSST [16]. We have adopted the simplest model of bias – linear bias, containing only one parameter per redshift bin – and considered the true spectra to be given exactly by the nonlinear bias expansion at second order and in the Lagrangian frame.

Our approach is based on the inclusion of theoretical uncertainty in the covariance matrix. The task of specifying such a matrix has been simplified through a few assumptions: that the theoretical error is Gaussian; that its non-diagonal elements decrease according to a Gaussian law; and that there is no correlation between theoretical errors at different redshift bins. The matrix is then determined by two functions: the so-called "envelope", which defines its diagonal elements, and the correlation coefficients. Through the assumption of a Gaussian law for their decrease, the latter are determined from a characteristic length setting the width of the Gaussian function, which we took as equal to the BAO scale at the given redshift bin.

We have tested two ways of specifying the envelope. "Envelope 1" is, by

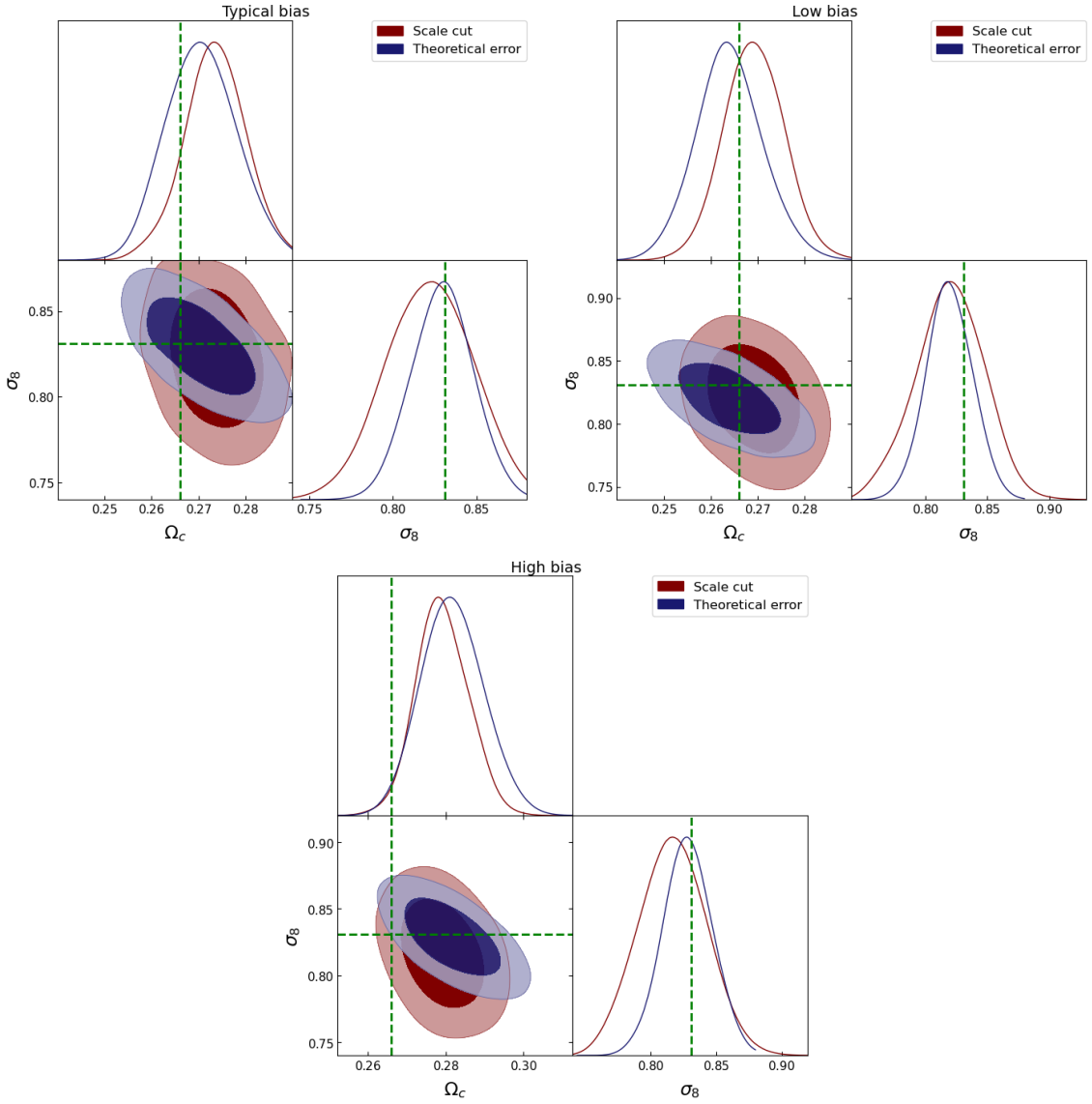

Figure 4.11: Results of the estimation with Envelope 2. Upper left, upper right, and lower panels show results obtained with data vectors representing typical, low, and high galaxy bias, respectively. Details are provided on table 4.5.

definition, equal to the standard deviation of the theoretical error on power spectra. By "theoretical error", we mean the difference between the best-fit, linear-bias prediction and the nonlinear bias spectrum. The standard deviation is then computed from a set of nonlinear-bias spectra with parameter ranges consistent with what is found in hydrodynamic simulations [30, 7].

As a second definition, we made the envelope equal to the theoretical error in a typical scenario of galaxy bias. We took the latter to be determined by the best-fit values of nonlinear bias parameters appearing in hydrodynamic simulations. This is similar to recent approaches to modelling the theoretical error in the case of baryonic feedback [23] or matter power spectra [15]. Envelope 2 was then defined as the difference between that typical, non-linear bias spectrum and its best-fit, linear-bias counterpart. We note that specifying a $k_{\rm max}$ is still required even in the analysis with theoretical error, although that value is, in general, considerably larger (2 to 3 times) than the corresponding value in the scale cut.

To compare this method with the scale cut, we performed likelihood analyses on a set of three data vectors. The main vector, named "typical bias", is similar to what is used in the Dark Energy Survey, year 3, to validate scale cuts [19]. To test robustness – i.e., the ability of one method to perform well under various realistic conditions –, we included two other test vectors obtained from "typical bias" by shifting the nonlinear bias parameters, b_2, b_{s^2}, and b_{∇^2}, by 70% of their standard deviations. This produced a "high bias" data vector, when all shifts were positive, and a "low bias" vector, when they were all negative.

Aimed at a proof of concept, our analysis is simplified to some extent: other systematic effects, such as intrinsic alignments, haven't been included; the estimation is made with only two redshift bins and two free cosmological parameters, Ω_c and σ_8. We recommend further studies to be performed under conditions more similar to those found in actual surveys.

We have found both envelopes to perform significantly better than the scale cut with respect to precision on the estimated parameters. With the "typical bias" data vector, indeed, the gains in the FoM with Envelopes 1 and 2 have been of 170% and 52%, respectively. As for robustness, although TE and the scale cut yield comparable results, the latter's performance is marginally better. That is, a change in the data vector causes a larger shift in the σ-tension obtained with TE than with the scale cut. In general, however, both methods can recover fiducial parameters with a tension smaller than 1σ.

We have found the approach considered in this work to give satisfactory results

when handling data for which one's models are unreliable. Indeed, we note that significant gains in precision have been attained without any change in the baseline model – that is, still with linear galaxy bias. If robustness is deemed a priority, then the performance with TE can be ameliorated by means of other definitions of the envelope – e.g., as either Envelope 1 or 2, defined in this work, multiplied by some scaling factor larger than 1. Further tests are thus recommended in this regard.

Chapter 5

Conclusion

We have covered some of the effects of theoretical uncertainty on analyses of large-scale structure. Our focus has been on probes dependent on the galaxy density contrast – namely, the spectra of galaxy clustering and galaxy-galaxy lensing. Measurements of the galaxy field yield information on the underlying distribution of dark matter, upon which cosmological models can be constrained.

The inference from one spectrum to the other can be made through the galaxy-bias expansion, which provides a connection between the galaxy and matter fields. We have thus, before presenting such applications, reviewed the solution to fluid equations in standard perturbation theory, as well as the corrections leading to an effective field theory of structure growth.

At one-loop order, the galaxy-bias expansion is known to be accurate on scales as small as $k \sim 0.3h\,\mathrm{Mpc}^{-1}$. It involves, however, at least four bias parameters. Alternatively, analyses can be based on a simpler, one-parameter model and restrict themselves to larger scales, with k-modes only as high as $k \sim 0.1h\,\mathrm{Mpc}^{-1}$. This has the advantages of a smaller parameter space, and it is the approach adopted, for example, in the Dark Energy Survey [19].

It is clear that, irrespective of the model informing one's analyses, there will be a range of scales for which it won't accurately represent the phenomenon in question – i.e., fluctuations of the galaxy number density. One conservative way to deal with such limitations, the so-called *scale cut*, consists of simply neglecting that part of the spectrum for which the model is unreliable. This means discarding data on k-modes modes above a certain k_{max}, and it is the preferred method in several present-day observational programs, such as the Dark Energy Survey and the LSST [19, 16]. Although this approach guarantees the model's accuracy with respect to the data used in the analysis, it may unnecessarily reduce the estimation's precision.

In this work, we tested an alternative approach. It can be viewed as a gradual scale cut: the k-modes included in the analysis are weighted down according to the model's credibility in that regime. That credibility, or the level of theoretical

uncertainty on a certain scale, can be assessed in various ways. It is generally dependent on the model's ability to predict the power spectrum on given scales.

A comparison can be made, for example, between the theoretical power spectrum based on the linear-bias model, $P_{\rm t}^{\rm (gg)}$, and a "true" power spectrum taken to be given, for example, by the nonlinear bias model at one-loop order, $P_{\rm d}^{\rm (gg)}$. When the error $P_{\rm d}^{\rm (gg)}(k) - P_{\rm t}^{\rm (gg)}(k)$ is large, so is deemed the theoretical uncertainty at k. That information can be included in the covariance matrix of the analysis. As a result, the estimation can be extended to a $k_{\rm max}$ significantly higher than the corresponding limit in the scale cut. Indeed, even if the additional modes have reduced statistical significance – in proportion to the model's credibility in the regime in question –, one is still able to achieve higher precision for the same level of estimation bias.

One inconvenient aspect of this approach is its greater complexity. Besides a maximum k-mode, one must also specify a covariance matrix for the theoretical error, a quantity with a large number of degrees of freedom. That problem may be mitigated by the adoption of a few simple rules for the assessment of theoretical error, from which the corresponding covariance matrix can be readily computed. For some such options – namely, one based on the standard deviation of the error and one on the error in a typical scenario –, we have compared the results yielded by this technique with those of the more-conservative scale cut. We have been able to achieve, in general, a higher performance in terms of figure of merit and estimation bias. We expect these studies may be of relevance to analyses implemented in future surveys.

Bibliography

[1] Muntazir Mehdi Abidi and Tobias Baldauf. Cubic halo bias in eulerian and lagrangian space. *Journal of Cosmology and Astroparticle Physics*, 2018(07):029–029, jul 2018.

[2] and N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, S. Basak, R. Battye, K. Benabed, J.-P. Bernard, M. Bersanelli, P. Bielewicz, J. J. Bock, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, J.-F. Cardoso, J. Carron, A. Challinor, H. C. Chiang, J. Chluba, L. P. L. Colombo, C. Combet, D. Contreras, B. P. Crill, F. Cuttaia, P. de Bernardis, G. de Zotti, J. Delabrouille, J.-M. Delouis, E. Di Valentino, J. M. Diego, O. Doré , M. Douspis, A. Ducout, X. Dupac, S. Dusini, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, Y. Fantaye, M. Farhang, J. Fergusson, R. Fernandez-Cobos, F. Finelli, F. Forastieri, M. Frailis, A. A. Fraisse, E. Franceschi, A. Frolov, S. Galeotta, S. Galli, K. Ganga, R. T. Génova-Santos, M. Gerbino, T. Ghosh, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gruppuso, J. E. Gudmundsson, J. Hamann, W. Handley, F. K. Hansen, D. Herranz, S. R. Hildebrandt, E. Hivon, Z. Huang, A. H. Jaffe, W. C. Jones, A. Karakci, E. Keihänen, R. Keskitalo, K. Kiiveri, J. Kim, T. S. Kisner, L. Knox, N. Krachmalnicoff, M. Kunz, H. Kurki-Suonio, G. Lagache, J.-M. Lamarre, A. Lasenby, M. Lattanzi, C. R. Lawrence, M. Le Jeune, P. Lemos, J. Lesgourgues, F. Levrier, A. Lewis, M. Liguori, P. B. Lilje, M. Lilley, V. Lindholm, M. López-Caniego, P. M. Lubin, Y.-Z. Ma, J. F. Macías-Pérez, G. Maggio, D. Maino, N. Mandolesi, A. Mangilli, A. Marcos-Caballero, M. Maris, P. G. Martin, M. Martinelli, E. Martínez-González, S. Matarrese, N. Mauri, J. D. McEwen, P. R. Meinhold, A. Melchiorri, A. Mennella, M. Migliaccio, M. Millea, S. Mitra, M.-A. Miville-Deschênes, D. Molinari, L. Montier, G. Morgante, A. Moss, P. Natoli, H. U. Nørgaard-Nielsen, L. Pagano, D. Paoletti, B. Partridge, G. Patanchon, H. V. Peiris, F. Perrotta, V. Pettorino, F. Piacentini, L. Polastri, G. Polenta, J.-L. Puget, J. P. Rachen, M. Reinecke, M. Remazeilles, A. Renzi, G. Rocha, C. Rosset, G. Roudier, J. A. Rubiño-Martín, B. Ruiz-Granados, L. Salvati, M. Sandri, M. Savelainen,

D. Scott, E. P. S. Shellard, C. Sirignano, G. Sirri, L. D. Spencer, R. Sunyaev, A.-S. Suur-Uski, J. A. Tauber, D. Tavagnacco, M. Tenti, L. Toffolatti, M. Tomasi, T. Trombetti, L. Valenziano, J. Valiviita, B. Van Tent, L. Vibert, P. Vielva, F. Villa, N. Vittorio, B. D. Wandelt, I. K. Wehus, M. White, S. D. M. White, A. Zacchei, and A. Zonca. iplanck/i2018 results. *Astronomy &amp Astrophysics*, 641:A6, sep 2020.

[3] Raul E Angulo, Matteo Zennaro, Sergio Contreras, Giovanni Aricò, Marcos Pellejero-Ibañez, and Jens Stücker. The BACCO simulation project: exploiting the full power of large-scale structure for cosmology. *Monthly Notices of the Royal Astronomical Society*, 507(4):5869–5881, 07 2021.

[4] G.B. Arfken, G.B. Arfken, H.J. Weber, and F.E. Harris. *Mathematical Methods for Physicists: A Comprehensive Guide*. Elsevier Science, 2013.

[5] Tobias Baldauf. Lecture notes: Field theory in cosmology – statistics, non-linearity and the cmb, March 2021.

[6] Tobias Baldauf, Mehrdad Mirbabayi, Marko Simonović, and Matias Zaldarriaga. Lss constraints with controlled theoretical uncertainties, 2016.

[7] Alexandre Barreira, Titouan Lazeyras, and Fabian Schmidt. Galaxy bias from forward models: linear and second-order bias of IllustrisTNG galaxies. *Journal of Cosmology and Astroparticle Physics*, 2021(08):029, aug 2021.

[8] D. Baumann. *Cosmology*. Cambridge University Press, 2022.

[9] Daniel Baumann, Alberto Nicolis, Leonardo Senatore, and Matias Zaldarriaga. Cosmological non-linearities as an effective fluid. *Journal of Cosmology and Astroparticle Physics*, 2012(07):051–051, jul 2012.

[10] F. R. Bouchet, S. Colombi, E. Hivon, and R. Juszkiewicz. Perturbative lagrangian approach to gravitational instability, 1994.

[11] John Joseph M. Carrasco, Mark P. Hertzberg, and Leonardo Senatore. The effective field theory of cosmological large scale structures. *Journal of High Energy Physics*, 2012(9), sep 2012.

[12] Shi-Fan Chen, Zvonimir Vlah, and Martin White. Consistent modeling of velocity statistics and redshift-space distortions in one-loop perturbation

theory. *Journal of Cosmology and Astroparticle Physics*, 2020(07):062–062, jul 2020.

[13] Nora Elisa Chisari, David Alonso, Elisabeth Krause, C. Danielle Leonard, Philip Bull, Jé rémy Neveu, Antonio Villarreal, Sukhdeep Singh, Thomas McClintock, John Ellison, Zilong Du, Joe Zuntz, Alexander Mead, Shahab Joudaki, Christiane S. Lorenz, Tilman Tröster, Javier Sanchez, Francois Lanusse, Mustapha Ishak, Renée Hlozek, Jonathan Blazek, Jean-Eric Campagne, Husni Almoubayyed, Tim Eifler, Matthew Kirby, David Kirkby, Stéphane Plaszczynski, Anže Slosar, Michal Vrastil, and Erika L. Wagoner and. Core cosmology library: Precision cosmological predictions for LSST. *The Astrophysical Journal Supplement Series*, 242(1):2, may 2019.

[14] Anton Chudaykin, Mikhail M. Ivanov, Oliver H. E. Philcox, and Marko Simonović. Nonlinear perturbation theory extension of the boltzmann code CLASS. *Physical Review D*, 102(6), sep 2020.

[15] Anton Chudaykin, Mikhail M. Ivanov, and Marko Simonović . Optimizing large-scale structure data analysis with the theoretical error likelihood. *Physical Review D*, 103(4), feb 2021.

[16] The LSST Dark Energy Science Collaboration, Rachel Mandelbaum, Tim Eifler, Renée Hložek, Thomas Collett, Eric Gawiser, Daniel Scolnic, David Alonso, Humna Awan, Rahul Biswas, Jonathan Blazek, Patricia Burchat, Nora Elisa Chisari, Ian Dell'Antonio, Seth Digel, Josh Frieman, Daniel A. Goldstein, Isobel Hook, Željko Ivezić, Steven M. Kahn, Sowmya Kamath, David Kirkby, Thomas Kitching, Elisabeth Krause, Pierre-François Leget, Philip J. Marshall, Joshua Meyers, Hironao Miyatake, Jeffrey A. Newman, Robert Nichol, Eli Rykoff, F. Javier Sanchez, Anže Slosar, Mark Sullivan, and M. A. Troxel. The lsst dark energy science collaboration (desc) science requirements document, 2021.

[17] S. Dodelson and F. Schmidt. *Modern Cosmology*. Elsevier Science, 2020.

[18] Wayne Hu and Bhuvnesh Jain. Joint galaxy-lensing observables and the dark energy. *Physical Review D*, 70(4), aug 2004.

[19] E. Krause, X. Fang, S. Pandey, L. F. Secco, O. Alves, H. Huang, J. Blazek, J. Prat, J. Zuntz, T. F. Eifler, N. MacCrann, J. DeRose, M. Crocce, A. Porredon,

B. Jain, M. A. Troxel, S. Dodelson, D. Huterer, A. R. Liddle, C. D. Leonard, A. Amon, A. Chen, J. Elvin-Poole, A. Ferté, J. Muir, Y. Park, S. Samuroff, A. Brandao-Souza, N. Weaverdyck, G. Zacharegkas, R. Rosenfeld, A. Campos, P. Chintalapati, A. Choi, E. Di Valentino, C. Doux, K. Herner, P. Lemos, J. Mena-Fernández, Y. Omori, M. Paterno, M. Rodriguez-Monroy, P. Rogozenski, R. P. Rollins, A. Troja, I. Tutusaus, R. H. Wechsler, T. M. C. Abbott, M. Aguena, S. Allam, F. Andrade-Oliveira, J. Annis, D. Bacon, E. Baxter, K. Bechtol, G. M. Bernstein, D. Brooks, E. Buckley-Geer, D. L. Burke, A. Carnero Rosell, M. Carrasco Kind, J. Carretero, F. J. Castander, R. Cawthon, C. Chang, M. Costanzi, L. N. da Costa, M. E. S. Pereira, J. De Vicente, S. Desai, H. T. Diehl, P. Doel, S. Everett, A. E. Evrard, I. Ferrero, B. Flaugher, P. Fosalba, J. Frieman, J. García-Bellido, E. Gaztanaga, D. W. Gerdes, T. Giannantonio, D. Gruen, R. A. Gruendl, J. Gschwend, G. Gutierrez, W. G. Hartley, S. R. Hinton, D. L. Hollowood, K. Honscheid, B. Hoyle, E. M. Huff, D. J. James, K. Kuehn, N. Kuropatkin, O. Lahav, M. Lima, M. A. G. Maia, J. L. Marshall, P. Martini, P. Melchior, F. Menanteau, R. Miquel, J. J. Mohr, R. Morgan, J. Myles, A. Palmese, F. Paz-Chinchón, D. Petravick, A. Pieres, A. A. Plazas Malagón, E. Sanchez, V. Scarpine, M. Schubnell, S. Serrano, I. Sevilla-Noarbe, M. Smith, M. Soares-Santos, E. Suchyta, G. Tarle, D. Thomas, C. To, T. N. Varga, and J. Weller. Dark energy survey year 3 results: Multi-probe modeling strategy and validation, 2021.

[20] Elisabeth Krause and Tim Eifler. cosmolike – cosmological likelihood analyses for photometric galaxy surveys. *Monthly Notices of the Royal Astronomical Society*, 470(2):2100–2112, may 2017.

[21] Marilena LoVerde and Niayesh Afshordi. Extended limber approximation. *Phys. Rev. D*, 78:123506, Dec 2008.

[22] Cameron McBride, Andreas Berlind, R. Scoccimarro, Risa Wechsler, M. Busha, Jeffrey Gardner, and Frank van den Bosch. Las damas mock galaxy catalogs for sdss. 41, 01 2009.

[23] Maria G Moreira, Felipe Andrade-Oliveira, Xiao Fang, Hung-Jin Huang, Elisabeth Krause, Vivian Miranda, Rogerio Rosenfeld, and Marko Simonović . Mitigating baryonic effects with a theoretical error covariance. *Monthly Notices of the Royal Astronomical Society*, 507(4):5592–5601, sep 2021.

[24] Oliver Philcox. An introduction to the EFT of LSS. https://oliverphilcox.github.io/files/eft_intro.pdf. [Online; accessed 14-August-2023].

[25] E. Rozo, E. S. Rykoff, A. Abate, C. Bonnett, M. Crocce, C. Davis, B. Hoyle, B. Leistedt, H. V. Peiris, R. H. Wechsler, T. Abbott, F. B. Abdalla, M. Banerji, A. H. Bauer, A. Benoit-Lé vy, G. M. Bernstein, E. Bertin, D. Brooks, E. Buckley-Geer, D. L. Burke, D. Capozzi, A. Carnero Rosell, D. Carollo, M. Carrasco Kind, J. Carretero, F. J. Castander, M. J. Childress, C. E. Cunha, C. B. D'Andrea, T. Davis, D. L. DePoy, S. Desai, H. T. Diehl, J. P. Dietrich, P. Doel, T. F. Eifler, A. E. Evrard, A. Fausti Neto, B. Flaugher, P. Fosalba, J. Frieman, E. Gaztanaga, D. W. Gerdes, K. Glazebrook, D. Gruen, R. A. Gruendl, K. Honscheid, D. J. James, M. Jarvis, A. G. Kim, K. Kuehn, N. Kuropatkin, O. Lahav, C. Lidman, M. Lima, M. A. G. Maia, M. March, P. Martini, P. Melchior, C. J. Miller, R. Miquel, J. J. Mohr, R. C. Nichol, B. Nord, C. R. O'Neill, R. Ogando, A. A. Plazas, A. K. Romer, A. Roodman, M. Sako, E. Sanchez, B. Santiago, M. Schubnell, I. Sevilla-Noarbe, R. C. Smith, M. Soares-Santos, F. Sobreira, E. Suchyta, M. E. C. Swanson, J. Thaler, D. Thomas, S. Uddin, V. Vikram, A. R. Walker, W. Wester, Y. Zhang, and L. N. da Costa. redMaGiC: selecting luminous red galaxies from the DES science verification data. *Monthly Notices of the Royal Astronomical Society*, 461(2):1431–1450, may 2016.

[26] Max Tegmark. Measuring cosmological parameters with galaxy surveys. *Physical Review Letters*, 79(20):3806–3809, nov 1997.

[27] Jesús Torrado and Antony Lewis. Cobaya: Bayesian analysis in cosmology. Astrophysics Source Code Library, record ascl:1910.019, October 2019.

[28] Jesú s Torrado and Antony Lewis. Cobaya: code for bayesian analysis of hierarchical physical models. *Journal of Cosmology and Astroparticle Physics*, 2021(05):057, may 2021.

[29] Francisco Villaescusa-Navarro, ChangHoon Hahn, Elena Massara, Arka Banerjee, Ana Maria Delgado, Doogesh Kodi Ramanah, Tom Charnock, Elena Giusarma, Yin Li, Erwan Allys, Antoine Brochard, Cora Uhlemann, Chi-Ting Chiang, Siyu He, Alice Pisani, Andrej Obuljen, Yu Feng, Emanuele Castorina, Gabriella Contardo, Christina D. Kreisch, Andrina Nicola, Justin Alsing, Roman Scoccimarro, Licia Verde, Matteo Viel, Shirley Ho, Stephane

Mallat, Benjamin Wandelt, and David N. Spergel. The quijote simulations. *The Astrophysical Journal Supplement Series*, 250(1):2, aug 2020.

[30] Matteo Zennaro, Raul E Angulo, Sergio Contreras, Marcos Pellejero-Ibá ñez, and Francisco Maion. Priors on lagrangian bias parameters from galaxy formation modelling. *Monthly Notices of the Royal Astronomical Society*, 514(4):5443–5456, jun 2022.