Improvement in techniques for understanding the large scale structure of the Universe

by

Praful Gagrani

B. Tech., Indian Institute of Technology, Delhi, India 2015

A THESIS

submitted in partial fulfillment of the
requirements for the degree

MASTER OF SCIENCE

Department of Physics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2018

Approved by:

Major Professor
Lado Samushia

Copyright

Abstract

Part 1: The redshift-space bispectrum (three point statistics) of galaxies can be used to measure key cosmological parameters. In a homogeneous Universe, the bispectrum is a function of five variables and unlike its two point statistics counterpart – the power spectrum, which is a function of only two variables – is difficult to analyse unless the information is somehow reduced. The most commonly considered reduction schemes rely on computing angular integrals over possible orientations of the bispectrum triangle thus reducing it to sets of functions of only three variables describing the triangle shape. We use Fisher information formalism to study the information loss associated with this angular integration. We find that most of the information is in the azimuthal averages of the first three even multipoles. This suggests that the bispectrum of every configuration can be reduced to just three numbers (instead of a 2D function) without significant loss of cosmologically relevant information.

Part 2: One way of enhancing the cosmological information extracted from the clustering of galaxies is by weighting the galaxy field. The most widely used weighting schemes assign weights to galaxies based on the average local density in the region and their bias with respect to the dark matter field. They are designed to minimize the fractional variance of the galaxy power-spectrum. We demonstrate that the currently used bias dependent weighting scheme can be further optimized for specific cosmological parameters.

Part 3: Choice of the box-size of a cosmological simulation involves a crucial trade-off between accuracy and complexity. We use Lagrangian perturbation theory to study the effects of box size on the predicted power spectrum and Baryon Acoustic Oscillation ruler. We find that although the optimal size depends on the final redshift of evolution, in general the 2-point statistics of relevant scales is fairly accurate for a simulation box-size of length greater than 1000 Mpc.

Table of Contents

List of Figures

Acknowledgments

We thank David Pearson, Martin White, Héctor Gil Marin, Florian Beutler, Eichiro Komatsu, Cristiano Porciani and Emiliano Sefusatti for useful discussions in various parts of the project. Some part of the work was supported by SNSF grant SCOPES IZ73Z0 152581, GNSF grant FR/339/6-350/14, and NASA grant 12-EUCLID11-0004. We also used NASA's Astrophysics Data System Bibliographic Service and the arXiv e-print service for bibliography search and http://cosmocalc.icrar.org/ for computing some cosmological parameters.

Many thanks to Ms. Kim Coy, Ms. Peggy Matthews, and the rest of the administration in the department for their extremely tolerant nature towards my constant unintelligent and dense queries. I literally would not be able to graduate without you! I am also very grateful to Professors Larry Weaver and David Auckly for constantly engaging with my curiosities in different areas of physics, maths, philosophy and life in general.

I would like to extend my gratitude towards my family who have constantly supported me, in their own way, in getting here. I also want to thank my friends for the time spent together, fun, love, memories, and realizations we have had, a vast subject to which a thesis can be devoted in its entirety.

Lastly, I want to thank my advisor for his constant support and patience for the past 3 years. I appreciate the freedom I was granted to investigate side projects and subjects, most of which are not mentioned here. I am really grateful for his encouragement towards attending conferences, which exposed me to the bigger picture in the field and at the same time helped me derive a sense of fulfillment from my own work. I am honored and rather lucky be his first graduate student!

Preface

This thesis consists of three parts:

- Part 1: Information Content of the Angular Multipoles of Redshift-Space Galaxy Bispectrum

  Authors: Praful Gagrani, Lado Samushia

  This work appeared in Monthly Notices of the Royal Astronomical Society, Volume 467, Issue 1, 11 May 2017.

- Part 2: Optimal Weights for Measuring Redshift Space Distortions in Multitracer Galaxy Catalogues

  Authors: David W. Pearson Lado Samushia Praful Gagrani

  This work appeared in Monthly Notices of the Royal Astronomical Society, Volume 463, Issue 3, 11 December 2016. I contributed calculations and Mathematica code to the project, the details of which are mentioned in Chap. 3.

- Part 3: Determining Optimal Simulation Box Size using LPT

  This work is a small project, parts of which will be circulated in the DESI Collaboration.

Chapter 1

Overview

One of the biggest goals of cosmology is to understand the large scale structure of the Universe. The current understanding of the energy budget of the Universe is that about 27% and 68% of the total are dark matter and dark energy[1], respectively, both of which remain elusive for direct experimental detection. Thus, we need indirect methods of detection and this is one of the reasons for the excitement in large scale structure (LSS).

Our theoretical understanding and modeling of the LSS has seen some rather dramatic leaps and revisions in the past few decades. Arguably, the major driving force for these advancements is the ever increasing astronomical data made available by galaxy surveys like the Sloan Digital Sky Survey[2] (SDSS), who have probed the local Universe with an unprecedented level of precision. Moreover, with the next generation of surveys like the Dark Energy Spectroscopic Instrument[3;4] (DESI) project and the Large Synoptic Survey Telescope[5] (LSST), we will see a many-fold increase in the amount as well as precision of the available data.

Clearly, the large data sets need to be processed before any meaningful conclusions can be drawn from them. For this, cosmologists employ statistical methods like calculating n-point correlation functions[6] of the data, which can be predicted by theory. The first part of this thesis looks at how the bispectrum, equivalent to a 3-point correlation function, can be calculated more efficiently from large data sets. To optimize the calculations, we employ

several physically motivated approximations and averagings, the robustness of which is also demonstrated in the work[7].

Another crucial aspect of data analysis involves optimizing the already available methods to maximize signal-to-noise ratio. In the second part of this thesis, we look at how different weighting schemes affect the variance of specific cosmological parameters measured and provide an optimal solution for the same. The complete details can be found in Pearson et al.[8].

Computer simulations are a powerful and important way of understanding the Universe. However, owing to the infinite nature of the universe, it is impossible to simulate it beyond a level of precision. In the third part of the thesis, we look at how the simulation box size affects the results one obtains from them and its closeness to the actual Universe. We provide a theoretical analysis and lower bound on the simulation box size one needs to accurately describe the length scales of the order 100 Mpc.

Though seemingly distinct, the nature of our work has been to develop or optimize techniques used in understanding large scale structure of the Universe.

Chapter 2

Part 1: Information Content of the Angular Multipoles of Redshift-Space Galaxy Bispectrum

2.1 Introduction

The statistical properties of the matter distribution in the Universe depend on its expansion and growth history and can be used to measure key cosmological parameters describing the composition of the Universe, the nature of dark energy, and gravity.

The power spectrum (or its Fourier conjugate, the correlation function) is currently the most widely used statistical measurement for the purposes of cosmological analysis of galaxy surveys. The power spectrum of matter is defined as a two point statistic of a Fourier transformed overdensity field $\delta(\boldsymbol{r})$,

$$P(\boldsymbol{k}) \equiv \frac{\langle|\delta(\boldsymbol{k})|^2\rangle}{V_{\rm s}}, \tag{2.1}$$

where

$$\delta(\boldsymbol{k}) = \int \mathrm{d}\boldsymbol{r}\, \delta(\boldsymbol{r}) \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\boldsymbol{r}} \tag{2.2}$$

brackets denote ensemble average, $\delta(\boldsymbol{r})$ is defined in Eq. 2.5, and $V_\mathrm{s} \equiv \int \mathrm{d}\boldsymbol{r}$ is the observed 3-D volume.

For a statistically isotropic field the power spectrum would only depend on the magnitude of the wavevector, $k = |\boldsymbol{k}|$. The observed galaxy field is however anisotropic with respect to the line-of-sight direction to the observer, mainly due to the redshift-space distortions (RSD),[9] and the Alcock-Paczinsky effects (AP)[10]. Because of this anisotropy, in addition to the magnitude of the wavevector k, the power spectrum also depends on its angle with respect to the line-of-sight θ, making it a function of two variables.

To make the cosmological analysis numerically less demanding the power spectrum is usually reduced to the coefficients of the Legendre-Fourier expansion with respect to μ=cos(θ)[11]

$$P_\ell(k) \equiv \frac{2\ell+1}{2} \int_{-1}^{1} \mathrm{d}\mu\, P(k,\mu) \mathcal{L}_\ell(\mu), \tag{2.3}$$

where $\mathcal{L}_\ell$ are Legendre polynomials of order ℓ.

Recent studies showed that the first three even Legendre coefficients contain almost all of the information on key cosmological parameters. This suggests that for the purposes of cosmological analysis the power spectrum at each wavevector can be replaced by just three numbers (instead of a function of μ) without a significant loss of information[12–14].

The bispectrum (or its Fourier conjugate, the three-point correlation function), defined as,

$$B(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3) \equiv \frac{\langle \delta(\boldsymbol{k}_1) \delta(\boldsymbol{k}_2) \delta(\boldsymbol{k}_3) \rangle}{V_\mathrm{s}} \tag{2.4}$$

is more difficult to measure and to model, and is not currently used as frequently as the power spectrum to derive cosmological constraints[15–18]. The bispectrum measurements have mostly been considered as a means of estimating the primordial non-Gaussianity in the

matter field[19;20], but a number of recent studies used them for Baryon-Acoustic Oscillations (BAO) and RSD constraints[21–24].

If the statistical properties of the Universe are homogeneous (a key assumption in the standard model of cosmology) the bispectrum is non-zero only for $\boldsymbol{k}_1 + \boldsymbol{k}_2 + \boldsymbol{k}_3 = 0$ ($\boldsymbol{k}$ vectors must make a triangle) reducing the number of variables from nine to six. From now on we will write $B(\boldsymbol{k}_1, \boldsymbol{k}_2)$ assuming the third vector to be equal to $\boldsymbol{k}_3 = -\boldsymbol{k}_1 - \boldsymbol{k}_2$. The partial isotropy with respect to rotations around the line-of-sight axis removes one more variable, making the bispectrum a five dimensional function. One possible choice of these five variables is a triplet k_1, k_2, k_3 ($k_i \equiv |\boldsymbol{k}_i|$), describing the shape of the bispectrum triangle and two angles describing its orientation, e.g. θ_1 – the angle of $\boldsymbol{k}_1$ vector with respect to the line-of-sight direction, and ξ – azimuthal angle of $\mathbf{k}_2$ around $\mathbf{k}_1$ (see Sec. 2.2.1 for a formal definition).

An obvious extension of the Legendre-Fourier decomposition of the power spectrum is a spherical harmonics decomposition of the bispectrum for angles θ_1 and ξ[25]. Unlike the power spectrum, this double angular multipole expansion of the bispectrum does not truncate at finite order (see Sec. 2.3.2). The main objective of this work is to identify the expansion coefficients that contain the most cosmologically relevant information (see Sec. 2.4).

Galaxies provide a biased, discrete sampling of the underlying matter field and along with the cosmic microwave background experiments currently provide one of the best estimates of the clustering of matter in the Universe[26;27]. Our Fisher information based computations suggest the five dimensional bispectrum with no reduction can deliver up to factor of 1.2 better constraints on the growth rate parameter f compared to the power spectrum, from a sample of emission line galaxies (ELG) expected from future surveys such as the Dark Energy Spectroscopic Instrument survey (DESI;[4]) and Euclid satellite surveys[28] at a redshift of $z \sim 1$ (see Sec. 2.5). For a sample of Luminous Red Galaxies (LRG) at lower redshifts the improvement could be as large as a factor of 3.

We show that most of this information is contained in the first three even multipoles in angle θ_1 averaged over ξ. Constraints on key cosmological parameters from these multipoles are weaker compared to the constraints derived from the full bispectrum by no more than 10

per cent at all redshifts and for all tracer types we studied. This suggests that a bispectrum of each triangular configuration can be replaced by just three numbers (as opposed to a two variable function) for all practical purposes (see Sec. 2.6).

2.2 Review of Power Spectrum and Bispectrum

2.2.1 Leading Order Model

We will start with a standard assumption that galaxies form a Poisson sample of a biased matter density field[6],

$$n(\mathbf{x}) = \bar{n}\left[1 + b_1\delta(\mathbf{x}) + \frac{b_2}{2}\delta(\mathbf{x})^2\right], \tag{2.5}$$

where b_1 and b_2 are the first and second order bias parameters and we ignore higher order bias terms as well as non-local contributions of $\delta(\boldsymbol{x})$ to the number density of galaxies.

To the leading order in δ the power spectrum is given by[9],

$$P(\mathbf{k}) = (b_1 + f\mu^2)^2 P_{\rm m}(k), \tag{2.6}$$

where f is a growth rate and $P_{\rm m}$ is a one dimensional matter power spectrum function that can be numerically computed for any cosmological model*. Also in the leading order of perturbation theory and assuming local bias the bispectrum of galaxies is given by[29],

$$\begin{aligned} B(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) =& 2\mathbf{Z}_1(\mu_1)\mathbf{Z}_1(\mu_2)\mathbf{Z}_2 P(k_1)P(k_2) \\ &+ \text{cyclic terms}, \end{aligned} \tag{2.7}$$

*The bias and the growth rate can not be decoupled from the amplitude parameter σ_8 when using only the galaxy clustering data on linear scales at a single redshift. For brevity, we will continue using b and f to denote parameter combinations $b\sigma_8$ and $f\sigma_8$.

where

$$\mathbf{Z}_1(\mu) = \left(b_1 + f\mu^2\right), \tag{2.8}$$

$$\mathbf{Z}_2 = \left\{ \frac{b_2}{2} + b_1 F_2(\mathbf{k}_1, \mathbf{k}_2) + f\mu_3^2 G_2(\mathbf{k}_1, \mathbf{k}_2) \right.$$
$$\left. - \frac{f\mu_3 k_3}{2}\left[\frac{\mu_1}{k_1}(b_1 + f\mu_2^2) + \frac{\mu_2}{k_2}(b_1 + f\mu_1^2)\right]\right\}, \tag{2.9}$$

$$F_2(\mathbf{k}_1, \mathbf{k}_2) = \frac{5}{7} + \frac{\mathbf{k}_1.\mathbf{k}_2}{2k_1k_2}\left(\frac{k_1}{k_2} + \frac{k_2}{k_1}\right) + \frac{2}{7}\left(\frac{\mathbf{k}_1.\mathbf{k}_2}{k_1k_2}\right)^2, \tag{2.10}$$

$$G_2(\mathbf{k}_1, \mathbf{k}_2) = \frac{3}{7} + \frac{\mathbf{k}_1.\mathbf{k}_2}{2k_1k_2}\left(\frac{k_1}{k_2} + \frac{k_2}{k_1}\right) + \frac{4}{7}\left(\frac{\mathbf{k}_1.\mathbf{k}_2}{k_1k_2}\right)^2, \tag{2.11}$$

and cyclic terms can be derived by replacing indexes 1 and 2 in the first term by 2 and 3, and 1 and 3 respectively.

The AP effect induces distortions in the measured power spectrum and the bispectrum that can be modeled by substituting

$$k \to \frac{k}{\alpha_\perp}\sqrt{1 + \mu^2(A^{-2} - 1)} \tag{2.12}$$

$$\mu \to \frac{\mu}{\sqrt{A^2 + \mu^2(1 - A^2)}} \tag{2.13}$$

and renormalizing the power spectrum by a factor of $1/\alpha_\perp^2\alpha_\parallel$ and the bispectrum by the square of the same factor. $A = \alpha_\parallel/\alpha_\perp$ in the above equations and the α parameters can be linked to properties of dark energy[30–32].

A standard practice when analysing galaxy power spectrum is to assume that the shape of the matter power spectrum is well determined from external cosmological data sets (e.g. the cosmic microwave background experiments) and to treat it as a function of four cosmological parameters $b_1, f, \alpha_\perp, \alpha_\parallel$ The bispectrum in addition will depend on the second order bias parameter b_2. For simplicity we ignore the commonly included $\sigma_{\rm FOG}$[33] parameter here. Its effect is to reduce information content on small scales. Since we are interested only in the

relative constraining power of the power spectrum, the bispectrum, and their multipoles, this omission does not affect our main results. * These parameters then can be estimated from the measured power spectrum and the bispectrum. We will adhere to this standard assumption and will ignore other cosmological parameters that may be relevant (e.g. $f_{\rm NL}$ describing primordial non-Gaussianity, or $N_{\rm eff}$ number of neutrino species).

2.2.2 Variance of the Measurements

If a power spectrum is measured from an observed volume $V_{\rm s}$ using optimal estimators[34] the variance of the measurement is

$$\langle [\Delta P(\mathbf{k})]^2 \rangle = \left(P(\mathbf{k}) + \frac{1}{\bar{n}} \right)^2 , \tag{2.14}$$

where ΔP is the difference between the true power spectrum and the one estimated from finite (and noisy) data and $\overline{n}$ is the average number density of galaxies. In an analogous way, for the bispectrum measured with an optimal estimator the variance is[29;35]

$$\langle [\Delta B(\mathbf{k}_1, \mathbf{k}_2)]^2 \rangle = V_{\rm s} \left(P(\mathbf{k}_1) + \frac{1}{\overline{n}} \right) \left(P(\mathbf{k}_2) + \frac{1}{\overline{n}} \right) \left(P(\mathbf{k}_3) + \frac{1}{\overline{n}} \right) . \tag{2.15}$$

2.3 Bispectrum Multipoles

2.3.1 Parameterization of the Bispectrum

Eq. (2.6) shows that the power spectrum can be expressed as a function of only two variables – k and μ. This results from the azimuthal symmetry of the field and is true even when the linear theory expression in Eq. (2.6) is replaced by its non-linear equivalent.

Similarly, even though the bispectrum in Eq. (2.7) is written in terms of three vectors $\mathbf{k}_1$, $\mathbf{k}_2$ and $\mathbf{k}_3$, as discussed in Sec. 2.1, because of various symmetries, only five variables

*When fitting real data more "nuisance" parameters are required to effectively describe the shortcomings of theoretical modelling. We ignore the effect of these "nuisance" parameters here as well since they depend on the specifics of modelling and do not effect our main results anyway.

are in fact independent. Following[25] we choose these variables to be the lengths of three wavevectors k_1, k_2, k_3 – describing the shape of the bispectrum triangle, and two angles describing its orientation – the angle θ_1 of wavevector $\boldsymbol{k}_1$ with respect to the line-of-sight direction, and the azimuthal angle ξ of vector $\boldsymbol{k}_2$ around $\boldsymbol{k}_1$. The first four variables are trivially obtained from the original wavevectors while the ξ can be computed from

$$\mu_2 = \cos(\theta_1)\cos(\phi_{12}) - \sin(\theta_1)\sin(\phi_{12})\cos(\xi), \tag{2.16}$$

where ϕ_{12} is the angle between $\mathbf{k}_1$ and $\mathbf{k}_2$,

$$\phi_{12} = \cos^{-1}\left(\frac{\mathbf{k}_1\mathbf{k}_2}{k_1 k_2}\right). \tag{2.17}$$

2.3.2 Series Expansion of Bispectrum

The power spectrum can be decomposed into a Legendre-Fourier series in angle μ

$$P(\boldsymbol{k}) = \sum_\ell P_\ell(k)\mathcal{L}_\ell(\mu) \tag{2.18}$$

where $\mathcal{L}_\ell$ are Legendre polynomials of order ℓ and the coefficients of decomposition can be found using Eq. (2.3). In linear theory only the first three even coefficients are nonzero and they contain most of the information on key cosmological parameters.

Since $0 \leq \theta_1 < \pi$ and $0 \leq \xi < 2\pi$, the bispectrum can be decomposed in spherical harmonics of θ_1 and ξ

$$B(k_1, k_2, k_3, \theta_1, \xi) = \sum_\ell \sum_{m=-\ell}^{\ell} B_{\ell m}(k_1, k_2, k_3) Y_\ell^m(\theta_1, \xi). \tag{2.19}$$

Subsequently,

$$B_{\ell m}(k_1, k_2, k_3) = \int_{-1}^{1} \mathrm{d}\cos(\theta) \int_0^{2\pi} d\xi B(k_1, k_2, k_3, \theta_1, \xi)\, Y_\ell^{m*}(\theta_1, \xi). \tag{2.20}$$

Unlike the power spectrum, the bispectrum multipole expansion does not terminate at finite ℓ. Neither does it have zero odd multipoles. Reducing bispectrum to a finite number of its angular multipoles significantly simplifies the cosmological analysis. This reduction however will inevitably result in a loss of information.

From the practical point of view, computing multipoles with $m = 0$ is especially simple[25]. It is therefore interesting to see by how much the information degrades further if we only use $m = 0$ multipoles in the analysis. We will show that the loss of information associated with ignoring m larger than zero is negligible.

We will also show that almost all of the information on key cosmological parameters (compared to using the full bispectrum) is contained within the first three even multipoles ($\ell = 0, 2, 4$ with $m = 0$) of the bispectrum.

2.3.3 Covariance of Bispectrum Multipoles

The bispectrum multipoles from real data can be computed by summing over all triangles with fixed values of k_i and angular weights of Eq. (2.20). This corresponds to

$$\begin{aligned}\overline{B}_{\ell m}(k_1', k_2', k_3') \equiv \\ \frac{1}{2\pi}\int \mathrm{d}\boldsymbol{k}_1\mathrm{d}\boldsymbol{k}_2 \frac{\delta(\boldsymbol{k}_1)\delta(\boldsymbol{k}_2)\delta(\boldsymbol{k}_3)}{V_\mathrm{s}} Y_\ell^{m*}(\theta_1, \xi) \\ \times \quad \frac{\delta^\mathrm{D}(k_1 - k_1')}{k_1}\frac{\delta^\mathrm{D}(k_2 - k_2')}{k_2}\frac{\delta^\mathrm{D}(k_3 - k_3')}{k_3} = \\ \frac{1}{2\pi V_\mathrm{s}}\int \mathrm{d}\theta_1\mathrm{d}\xi\mathrm{d}\phi_1 \delta(\boldsymbol{k}_1')\delta(\boldsymbol{k}_2')\delta(\boldsymbol{k}_3') Y_\ell^{m*}(\theta_1, \xi),\end{aligned} \tag{2.21}$$

where we used the transformation of coordinates

$$\begin{aligned}\mathrm{d}\boldsymbol{k}_1\mathrm{d}\boldsymbol{k}_2 &= k_1^2\mathrm{d}k_1\mathrm{d}\cos(\theta_1)\mathrm{d}\phi_1 k_2^2\mathrm{d}k_2\mathrm{d}\cos(\theta_2)\mathrm{d}\phi_2 \\ &= 2\pi k_1 k_2 k_3 \mathrm{d}k_1\mathrm{d}k_2\mathrm{d}k_3\mathrm{d}\cos\theta_1\mathrm{d}\phi_1\mathrm{d}\xi,\end{aligned} \tag{2.22}$$

and the factor of 2π is to ensure that the expectation value of the estimator matches the definition in Eq. (2.19).

The variance of the bispectrum multipoles is then

$$\begin{aligned}\langle \Delta\overline{B}_{\ell m}(k_1,k_2,k_3)\Delta\overline{B}_{\ell' m'}(k_1,k_2,k_3)\rangle = \\ \frac{V_\mathrm{s}}{2\pi}\int \mathrm{d}\cos(\theta)\,\mathrm{d}\xi Y_\ell^{m*}(\theta,\xi)Y_{\ell'}^{m'*}(\theta,\xi) \\ \times \quad \left[P(\mathbf{k}_1)+\frac{1}{\overline{n}}\right]\left[P(\mathbf{k}_2)+\frac{1}{\overline{n}}\right]\left[P(\mathbf{k}_3)+\frac{1}{\overline{n}}\right]\end{aligned} \tag{2.23}$$

The derivation of this result is analogous to the power spectrum multipole covariance described in Yamamoto et al.[36].

Since we work in the limit of infinitely small k-bins only the multipoles with all $\boldsymbol{k}_i$ identical are correlated, but in general there is a correlation between multipoles with different values of ℓ and m.

2.4 Constraining Cosmological Parameters

For brevity we will use the following notation:

$$\mathrm{VarP}_{\boldsymbol{k}} \equiv \langle [\Delta P(\boldsymbol{k})]^2\rangle \tag{2.24}$$

$$\mathrm{VarB}_{\boldsymbol{k}_1\boldsymbol{k}_2} \equiv \frac{\langle [\Delta B(\boldsymbol{k}_1\boldsymbol{k}_2)]^2\rangle}{V_\mathrm{s}} \tag{2.25}$$

$$\mathrm{VarB}^{\ell m\ell' m'}_{k_1k_2k_3} \equiv \frac{2\pi}{V_\mathrm{s}}\langle \Delta B_{\ell m}(k_1,k_2,k_3)\Delta B_{\ell' m'}(k_1,k_2,k_3)\rangle \tag{2.26}$$

2.4.1 Information Content of the Full Bispectrum

We use a Fisher information formalism[37;38] to derive expected constraints on cosmological parameters $\boldsymbol{\theta} \equiv (b_1, b_2, f, \alpha_\perp, \alpha_\parallel)$.

For the power spectrum we follow the well established procedure of computing

$$F_{ij} = \frac{V_\mathrm{s}}{(2\pi)^3} \int \mathrm{d}\boldsymbol{k} \frac{\partial P(\boldsymbol{k})}{\partial \theta_i} \left(\mathrm{VarP}_{\boldsymbol{k}}\right)^{-1} \frac{\partial P(\boldsymbol{k})}{\partial \theta_i}. \tag{2.27}$$

Since the Fourier transform is computed over a finite volume the $\delta(\boldsymbol{k})$ measurements are independent only at discrete points in $\boldsymbol{k}$ space. The density of these points is $V_\mathrm{s}/(2\pi)^3$. The factor in front of Eq. (2.27) renormalizes the continuous integral over all $\boldsymbol{k}$ which would otherwise overestimate the available information.

We numerically compute the integral

$$F_{ij} = \frac{V_\mathrm{s}}{(2\pi)^2} \int \mathrm{d}\cos(\theta)\, k^2 \frac{\partial P(\boldsymbol{k})}{\partial \theta_i} \left(\mathrm{VarP}_{\boldsymbol{k}}\right)^{-1} \frac{\partial P(\boldsymbol{k})}{\partial \theta_i}, \tag{2.28}$$

where the power spectrum derivatives are obtained by numerically differentiating Eq. (2.6) and the power spectrum variance is given by Eq. (2.14). The integration limits are $0 < k < 0.2\ h/\mathrm{Mpc}$ and $0 \leq \cos(\theta) \leq 1$. The first restriction reflects the fact that the statistical properties of the galaxy field are difficult to model at high wavenumbers because of the effects of nonlinear evolution and baryonic physics and are usually omitted from the analysis. The second restriction reflects the fact that a Fourier transform of a real field obeys $\delta(\boldsymbol{k}) = \delta^*(-\boldsymbol{k})$ symmetry, which implies that the power spectrum estimates (which are proportional to $|\delta(\boldsymbol{k})|^2$) are not independent above and below the z axis. Eq. (2.28) has one less factor of 2π compared to Eq. (2.27) because we integrate over azimuthal angle $0 \leq \phi < 2\pi$ on which neither the power spectrum nor its variance depend.

For the full bispectrum we similarly numerically integrate over all possible triangles (both the shape and the configuration) and propagate the information to the cosmological parameters. The Fisher matrix of cosmological parameters in this case is given by

$$F_{ij} = \frac{V_\mathrm{s}^2}{(2\pi)^6} \int \mathrm{d}\boldsymbol{k}_1 \mathrm{d}\boldsymbol{k}_2 \frac{\partial B(\boldsymbol{k}_1 \boldsymbol{k}_2)}{\partial \theta_i} \left(V_\mathrm{s} \mathrm{VarB}_{\boldsymbol{k}_1, \boldsymbol{k}_2}\right)^{-1} \frac{\partial B(\boldsymbol{k}_1, \boldsymbol{k}_2)}{\partial \theta_j}, \tag{2.29}$$

where the factor of $V_\mathrm{s}^2/(2\pi)^6$ accounts for the density of points on a k-grid due to finite

volume of the survey, as before. The integral can be reduced to five dimensions

$$F_{ij} = \frac{V_{\rm s}}{(2\pi)^5}\int {\rm d}k_1 {\rm d}k_2 {\rm d}k_3 {\rm d}\cos(\theta_1){\rm d}\xi \\ \times \frac{\partial B(\boldsymbol{k}_1\boldsymbol{k}_2)}{\partial\theta_i}\left({\rm VarB}_{\boldsymbol{k}_1,\boldsymbol{k}_2}\right)^{-1}\frac{\partial B(\boldsymbol{k}_1,\boldsymbol{k}_2)}{\partial\theta_j}, \tag{2.30}$$

as the integration over ϕ_1 azimuthal angle is simply 2π.

We use Eq. (2.7) to compute the bispectrum (and its derivatives) and Eq. (2.15) to compute the covariance matrix of the bispectrum. A permutation of vectors $\boldsymbol{k}_i$ corresponds to the same bispectrum measurement. In order to account for this symmetry and not double count the data we impose a condition $k_1 > k_2 > k_3$ on the integration volume in addition to $k_i < 0.2$ h/Mpc restriction on each wavevector. We also impose the triangularity condition $k_1 - k_2 < k_3$.

2.4.2 Information Content of the Multipoles

The Fisher matrix of cosmological parameters from bispectrum multipoles is given a three dimensional integral over a sum

$$F_{ij} = \frac{V_{\rm s}^2}{(2\pi)^6}\int {\rm d}k_1 {\rm d}k_2 {\rm d}k_3\, k_1 k_2 k_3 \\ \times \sum_{\ell\ell' mm'}\frac{\partial B_{\ell m}(k_1,k_2,k_3)}{\partial\theta_i}\left(\frac{V_{\rm s}}{2\pi}{\rm VarB}^{k_1k_2k_3}_{\ell\ell' mm'}\right)^{-1}\frac{\partial B_{\ell' m'}(k_1,k_2,k_3)}{\partial\theta_j}, \tag{2.31}$$

where the integration is over all possible triangle shapes. Similarly to the bispectrum, to avoid double counting, we impose a restriction that $k_1 > k_2 > k_3$ and that the three sides satisfy the triangularity condition $k_1 - k_2 < k_3$. We also restrict ourselves to triangles with $k_1 < 0.2h$/Mpc.

We use Eq. (2.20) to compute numerical derivatives of the multipoles and Eq. (2.23) to compute the variance of the multipoles (and covariance between them). The constraints on parameters can be found using the diagonal entries of the inverse of the Fisher matrix $\sigma_\theta = F^{-1}_{\theta\theta}$. We evaluate the sum for increasing values of $\ell_{\rm max}$. To check the effects of higher

order terms in m we either take all values of $-\ell \leq m \leq \ell$ or only the $m = 0$. We also try only $m = 0$ modes for increasing even values of $\ell_{\rm max}$ as that seemed to have most of the information.

2.5 Results

Results in this section are derived assuming a spatially flat ΛCDM cosmological model with $\Omega_{\rm m} = 0.28$, $\Omega_\Lambda = 0.72$. We consider LRG and ELG samples expected from DESI. For the number density profile and the bias as a function of redshift we use the same numbers as[19] Tellarini et al.

Fig. 2.1 shows the expected cosmological constraints on $\boldsymbol{\theta}$ from the bispectrum multipoles for increasing values of $\ell_{\rm max}$. These results are for the LRG sample in the redshift range $0.6 < z < 0.7$, We compute this for all ℓ and m values, all ℓ values with only $m = 0$, and for only even ℓ modes with $m = 0$. We show expected constraints from the power spectrum and the bispectrum on the same plots for comparison.

Fig. 2.1 shows that the full (unreduced) bispectrum is capable of providing better constraints compared to the power spectrum if we use all information from scales up to $k_{\rm max} = 0.2\ h/{\rm Mpc}$. This is especially true for the growth rate parameter f where the improvement is almost a factor of 2 in the statistical errors. For the α parameters the constraints derived from the full bispectrum are still a factor of about 1.5 better compared to the power spectrum, but become slightly worse for the multipoles. In all cases the information in the multipoles seems to be mostly in the first three even ℓ modes with $m = 0$.

The behaviour seems to be qualitatively similar for other redshifts and tracers. Fig. 2.2 shows similar results over a wider redshift range. This means that the first even multipoles averaged over azimuthal angle are as good as the full bispectrum for the purposes of deriving cosmological constraints.

The bispectrum provides significantly larger improvement over the power spectrum at low redshifts. This is due to a high number density of galaxies and the higher amplitude of fluctuations.

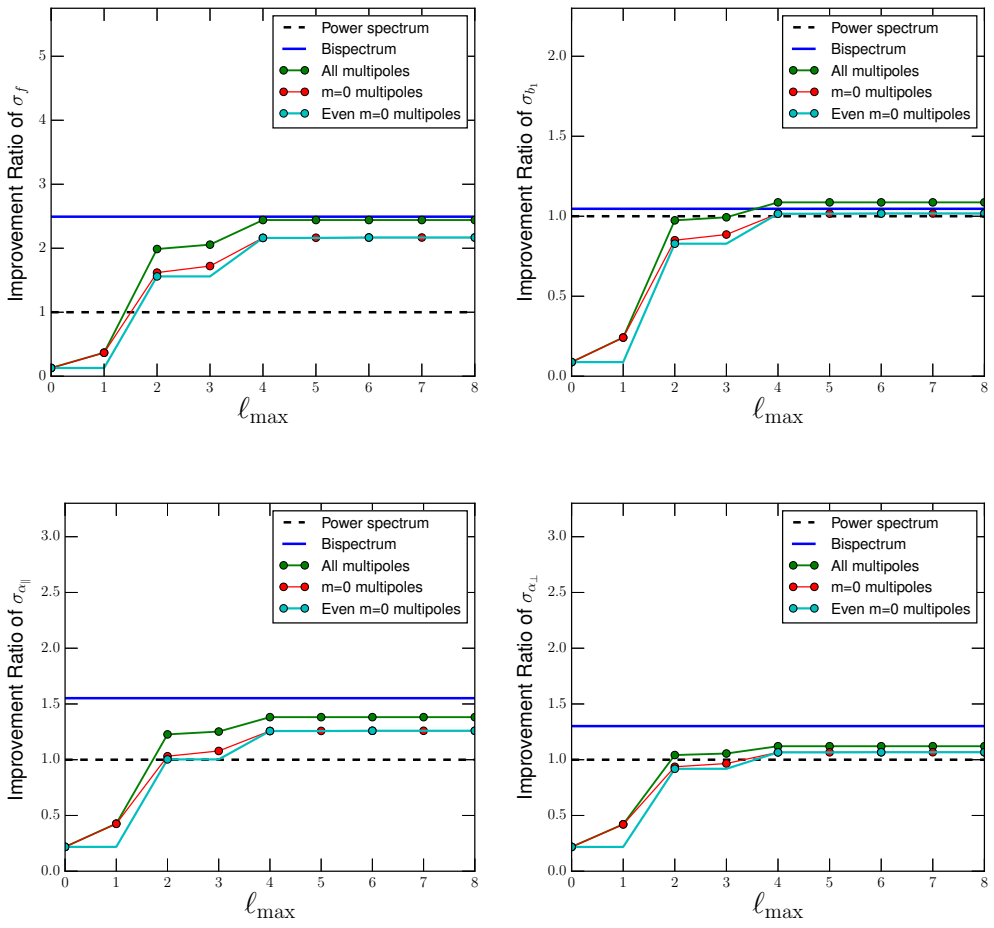

Figure 2.1: Cosmological constraints expected from the bispectrum multipoles as a function of maximum ℓ used in the analysis for a sample of DESI LRGs in $0.6 < z < 0.7$. The constraints from power spectrum and the full bispectrum for $k_{\rm max} = 0.2\ h/{\rm Mpc}$ are also displayed for comparison. The results are normalized to the expected power spectrum constraints so that the ordinate axis is an improvement factor over the power spectrum. The multipole constraints can never be stronger than the full bispectrum constraints. Our top right panel is consistent with this within the numerical error associated with monte carlo integration.

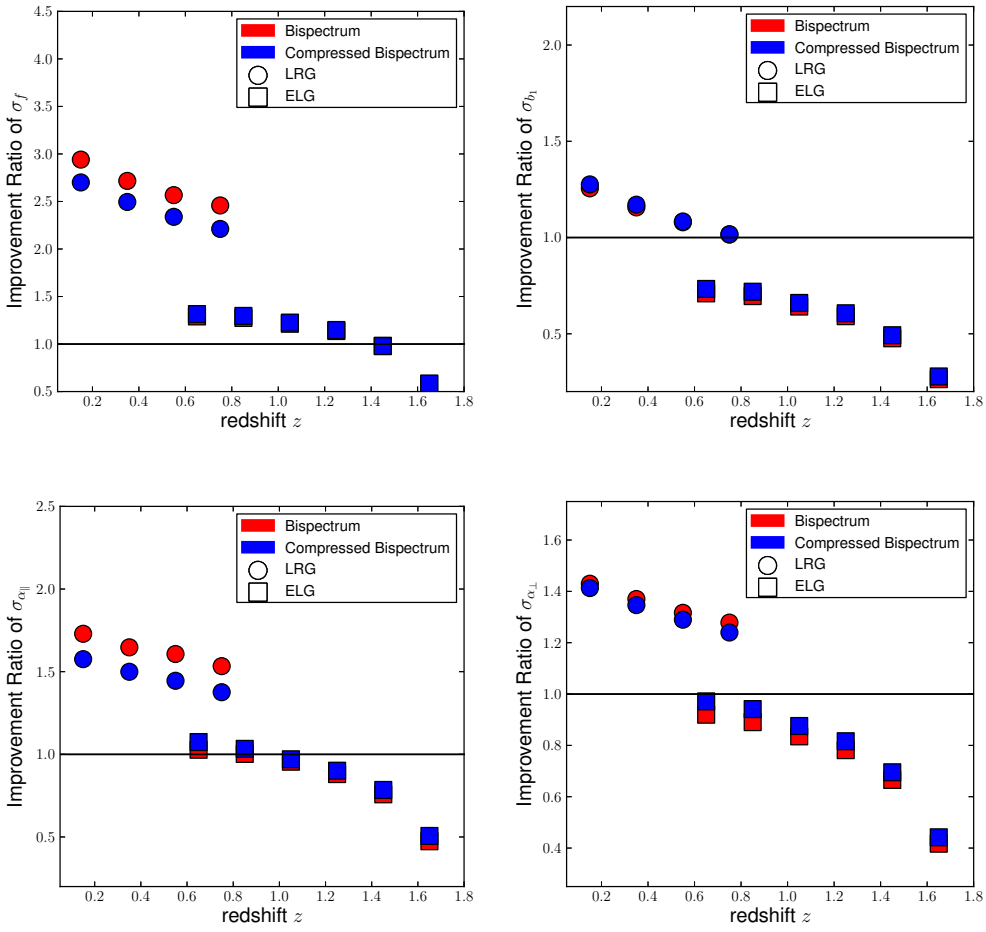

Figure 2.2: Improvement on derived errors of cosmological parameters compared to the power spectrum for different redshifts and tracer types if we consider all the modes up to 0.2 h/Mpc. Red symbols (on top) represent the constraints derivable from the full bispectrum, while the blue symbols (on the bottom) represent constraints from first three even multipoles with $m = 0$. For some redshifts the multipole constraints are slightly better than the full bispectrum constraints, but they are consistent within the numerical errors associated with the monte carlo integration.

2.6 Conclusions

We developed a Fisher information matrix based method of computing the expected constraints on cosmological parameters from the bispectrum and the angular multipoles of the bispectrum of a given galaxy sample. Since the full bispectrum is difficult to analyse, some kind of data reduction will inevitably have to be applied to the measurements. We computed the information loss associated with the commonly proposed reduction schemes that rely on angular integration of the bispectrum.

We find that the full bispectrum alone can deliver cosmological constraints that are a factor of upto 3 better than the ones derivable from the power spectrum at low z. This improvement steeply scales with $k_{\rm max}$ considered in the analysis. For $k_{\rm max} = 0.1\ h/{\rm Mpc}$ the information content of the Bispectrum is already comparable with the power spectrum (as shown in App. A), while for $k_{\rm max} = 0.2\ h/{\rm Mpc}$ it exceeds the power spectrum by a factor of 2 to 3. The improvement is especially large for the growth rate parameter f where the improvement on the measurement error is almost a factor of 3. The improvement is the largest at lower redshifts where the number density of galaxies in the sample is the highest. Most of the information is in the first three even multipoles with $m = 0$, which means that just three numbers per bispectrum shape are enough for the purposes of obtaining cosmological constraints.

Our results at first may seem to contradict previously published results that claim a more modest improvement when adding the bispectrum to the power spectrum[18;39–41]. This is due to a number of reasons. Many previous works have looked at the monopole of the bispectrum which will obviously contain much less information on f. The bispectrum information increases more steeply compared to the power spectrum with the number density of galaxies, therefore this large improvement will only result in future dense surveys and will not necessarily show in current and past surveys that have a lower galaxy number density. Finally, many past claims refer to "amplitude like" parameters (e.g. primordial amplitude of fluctuations) for isotropic fields. The f parameter is not really "amplitude like" since it describes the angular variation in the statistics, and the 5D shape of the bispectrum turns

out to be more sensitive to this parameter than it would be to a mere change in amplitude.

Our results are consistent with the ones reported in[15] if we only consider strictly linear scales of $k_i < 0.1$ Mpc/h. This is expected since the bispectrum signal to noise scales better with increasing $k_{\rm max}$ compared to the power spectrum. Their model includes the Finger of God effects and therefore the forecasts are more conservative and realistic. Since our main goal was not to produce accurate forecasts but rather to study the effects of the multipole reduction we decided to sacrifice the realism of constraints for clarity. We explicitly checked that our main conclusions are robust with respect to the choice of $k_{\rm max}$ and do not change when we include $\sigma_{\rm FOG}$.

In this work we do not consider a cross correlation between the power spectrum and the bispectrum measurements and it is difficult to say how big the overall improvement in the errors is when the two are properly combined (see[15] Song et al. for correlated full bispectrum DESI forecasts). We know however that the improvement will be at least as big as the improvement from the bispectrum (or the bispectrum multipoles) alone. Recent studies indicated that the cosmological constraints from power spectrum and bispectrum are not very strongly correlated[21;22;24], so the improvement may actually be much larger.

The main conclusions from our work are as follows:

- The bispectrum measurements from future surveys have a potential of improving the growth rate measurements by at least a factor of 2.5 at low redshifts (this is a very conservative estimate assuming that the bispectrum information is perfectly correlated with the power spectrum).

- When expanding the bispectrum in angular multipoles, the three numbers corresponding to the first three even terms with $m = 0$ in the multipole expansion contain most of the information relevant for the derivation of cosmological constraints.

Chapter 3

Part 2: Optimal Weights For Measuring Redshift Space Distortions in Multi-tracer Galaxy Catalogues

Note: The text and figures have been used with permission from David Pearson. These appear in Pearson, Samushia, Gagrani, 2016[8].

3.1 Introduction

Future galaxy redshift surveys, such as the Dark Energy Spectroscopic Instrument survey DESI;[4], the Extended Baryon Oscillation Spectroscopic survey eBOSS;[42], the *Euclid* satellite surveys[28], and the Wide Field Infrared Survey Telescope surveys WFIRST;[43] will cover vast cosmological volumes with a high number density of galaxies. Since the available cosmic volume is fundamentally limited a lot of effort is going into developing optimal ways of analysing galaxy clustering data.

One way of improving the variance of measured 2-point statistics is to weight the galaxy field to achieve the optimal signal-to-noise. The most commonly used weighting scheme is the one developed by[34] Feldman et al. (FKP), which is used in all analyses employing

2-point statistics. The FKP weights,

$$w_{\rm FKP}(\mathbf{r}) \propto \frac{1}{1+\overline{n}(\mathbf{r})P(\mathbf{k})}, \tag{3.1}$$

where $\overline{n}(\mathbf{r})$ is the average number density of galaxies at a position $\mathbf{r}$ and $P(\mathbf{k})$ is the power-spectrum at a wavelength of interest $\mathbf{k}$, are straightforward to apply and reduce the variance of the measured power-spectrum when the completeness of the galaxy sample is significantly non-uniform.

Percival, Verde, Peacock[44] (PVP) further optimized the FKP scheme for samples that include galaxies with a range of biases with respect to the dark matter. If the number density is uniform the PVP weights are

$$w_{\rm PVP} \propto b, \tag{3.2}$$

where b is the bias with respect to the dark matter, and will minimize the fractional variance in the measured power-spectrum.

In this work we generalize the PVP weighting scheme to minimize the variance of specific cosmological parameters measured from the power-spectrum (section 3.2).

3.2 Optimal Weighting

For simplicity, we will assume that galaxies of two types with densities $n_1(\mathbf{r})$ and $n_2(\mathbf{r})$ are present in an overlapping volume and the average number densities $\overline{n}_1$ and $\overline{n}_2$ do not vary significantly within the volume. The formalism is easy to generalize for more than two tracers and varying number densities. If we assign weights w_1 and w_2 to these galaxies, then the number density of the combined field is

$$n(\mathbf{r}) = w_1 n_1(\mathbf{r}) + w_2 n_2(\mathbf{r}) \tag{3.3}$$

and the overdensity field is

$$\delta(\mathbf{r}) \equiv \frac{n(\mathbf{r}) - \overline{n}}{\overline{n}} = A_1\delta_1(\mathbf{r}) + A_2\delta_2(\mathbf{r}), \tag{3.4}$$

where the overdensities are defined by

$$\delta_i(\mathbf{r}) = \frac{n_i(\mathbf{r}) - \overline{n}_i}{\overline{n}_i}, \tag{3.5}$$

and

$$A_i = \frac{w_i\overline{n}_i}{w_1\overline{n}_1 + w_2\overline{n}_2} \tag{3.6}$$

is the weighted fractional density. We will assume the weights to be normalised by $w_1 + w_2 = 1$. This will shorten some of our formulas, although in practice only the ratio of weights is relevant. The power-spectrum of the overdensity field,

$$P(\mathbf{k}) \equiv \left|\widetilde{\delta}(\mathbf{k})\right|^2, \tag{3.7}$$

can be estimated from the squared modulus of the Fourier transform,

$$\widetilde{\delta}(\mathbf{k}) = \int \mathrm{d}\mathbf{r}\, \mathrm{e}^{-i\mathbf{k}\mathbf{r}}\delta(\mathbf{r}), \tag{3.8}$$

We will assume that the overdensity fields are Gaussian (this is a common assumption when deriving optimal weights) with

$$\left\langle \widetilde{\delta}_i(\mathbf{k})\widetilde{\delta}^\star_j(\mathbf{k}) \right\rangle = \left[\left(b_i + \mu^2 f\right)\left(b_j + \mu^2 f\right) P_\mathrm{m}(k) + \frac{\delta^\mathrm{C}_{ij}}{\overline{n}_i} \right] V_\mathrm{s}, \tag{3.9}$$

where the angular brackets denote the expectation value, δ^C is a Kronecker delta function, V_s is the survey volume, and $P_\mathrm{m}(k)$ is the matter power-spectrum that can be computed in any given cosmological model[9]. The last term in equation (3.9) is the *shot noise term* due to the sampling of the overdensity field with a finite number of galaxies[34].* For the weighted

*The power-spectrum estimators are usually defined after subtracting the shot noise term, but this is

field in equation (3.4) this results in

$$P(\mathbf{k}) = \left[\left(b_{\rm eff}^{w_1 w_2} + \mu^2 f\right)^2 P_{\rm m}(k) + S^{w_1 w_2}\right] V_{\rm s}, \tag{3.10}$$

with the weighting dependent effective bias,

$$b_{\rm eff}^{w_1 w_2} = A_1 b_1 + A_2 b_2, \tag{3.11}$$

and the shot noise term,

$$S^{w_1 w_2} = \frac{A_1^2}{\overline{n}_1} + \frac{A_2^2}{\overline{n}_2}. \tag{3.12}$$

Since we assumed the overdensity field to be Gaussian, the variance of the galaxy power-spectrum estimator is simple to compute and is

$$\mathrm{Var}\left[P(\mathbf{k})\right] \propto \left[\left(b_{\rm eff}^{w_1 w_2} + \mu^2 f\right)^2 P_{\rm m}(k) + S^{w_1 w_2}\right]^2 \tag{3.13}$$

[34] FKP;[44] PVP. The fractional variance in the galaxy power-spectrum is then

$$\frac{P(\mathbf{k})}{\mathrm{Var}\left[P(\mathbf{k})\right]} \propto \frac{\left(b_{\rm eff}^{w_1 w_2} + \mu^2 f\right)^2 P_{\rm m}(k)}{\left[\left(b_{\rm eff}^{w_1 w_2} + \mu^2 f\right)^2 P_{\rm m}(k) + S^{w_1 w_2}\right]^2}. \tag{3.14}$$

This expression is minimized[†] by

$$\begin{aligned} w_1 &= \frac{b_1}{b_1 + b_2}, \\ w_2 &= \frac{b_2}{b_1 + b_2}, \end{aligned} \tag{3.15}$$

which, for constant number densities, is equivalent to PVP weighting.[‡]

The minimum fractional variance in the power-spectrum, however, does not necessarily correspond to the minimum variance in the cosmological parameters derived from the power-spectrum. The power-spectrum is most sensitive to the bias – $b_{\rm eff}$, growth rate – f, and

irrelevant for our results.

[†]This can be verified by simply equating the partial derivatives of equation (3.14) with respect to the weights to zero along with the Legendre multipliers to enforce the condition $w_1 + w_2 = 1$.

[‡]For a more rigorous derivation also accounting for the number density variations see[44].

the Alcock-Paczinsky parameters $\alpha_\parallel$ and $\alpha_\perp$ [10]. The dependence on b and f is already in equation (3.10), and the dependence on $\alpha_\perp$ and $\alpha_\parallel$ can be introduced by replacing

$$k \longrightarrow \frac{k}{\alpha_\perp}\left[1+\mu^2\left(\frac{\alpha_\perp^2}{\alpha_\parallel^2}-1\right)\right]^{1/2}, \tag{3.16}$$

$$\mu \longrightarrow \frac{\mu\alpha_\perp}{\alpha_\parallel}\left[1+\mu^2\left(\frac{\alpha_\perp^2}{\alpha_\parallel^2}-1\right)\right]^{-1/2}, \tag{3.17}$$

and dividing the power-spectrum by a factor of $\alpha_\parallel\alpha_\perp^2$ [32]. The Fisher information matrix of these parameters is

$$F_{ij} = \int \mathrm{d}\mathbf{k}\, \frac{\partial P(\mathbf{k})}{\partial\theta_i}\frac{1}{\mathrm{Var}\left[P(\mathbf{k})\right]}\frac{\partial P(\mathbf{k})}{\partial\theta_j}, \tag{3.18}$$

where $\theta = (b_{\rm eff}, f, \alpha_\parallel, \alpha_\perp)$ is a parameter vector, and the integration is over all wavevectors, the power-spectrum measurements of which were used in the analysis. The inverse of the Fisher matrix gives a covariance matrix

$$\mathsf{C} = \mathsf{F}^{-1} \tag{3.19}$$

and the diagonal elements of the covariance matrix correspond to the expected variance of the parameters measured from the power-spectrum. Because of the presence of the derivative terms (that also depend on the weights) in equation (3.18) the weighting scheme that minimizes the variance of the power-spectrum does not necessarily minimize the diagonal elements of the covariance matrix in equation (3.19).

A simple analytic solution for the optimal weights in this case does not exist, but they are relatively straightforward to find numerically. To find the optimal weights we numerically compute the variance and the derivatives in equation (3.18) and take the integral over the wavevectors of interest. We then numerically find the weights that minimize the diagonal elements of the inverse Fisher matrix of equation (3.19).[§] These weights will in general be different for each parameter. In the tradition of FKP and PVP, we refer to these as PSG

[§]Since we adopt the normalization $w_1 + w_2 = 1$ this turns into a simple one parameter minimization procedure.

weights in the work.

3.3 Conclusion

As can be seen in fig. 3.1, PSG weights are more optimal for 'f' and 'b' than PVP and FKP weights. For $\alpha_\perp$ and $\alpha_\parallel$, the variance is flat for a broad range of weights leading all weighting schemes to perform equally well. This shows advantage of using the PSG weighting scheme in practice.

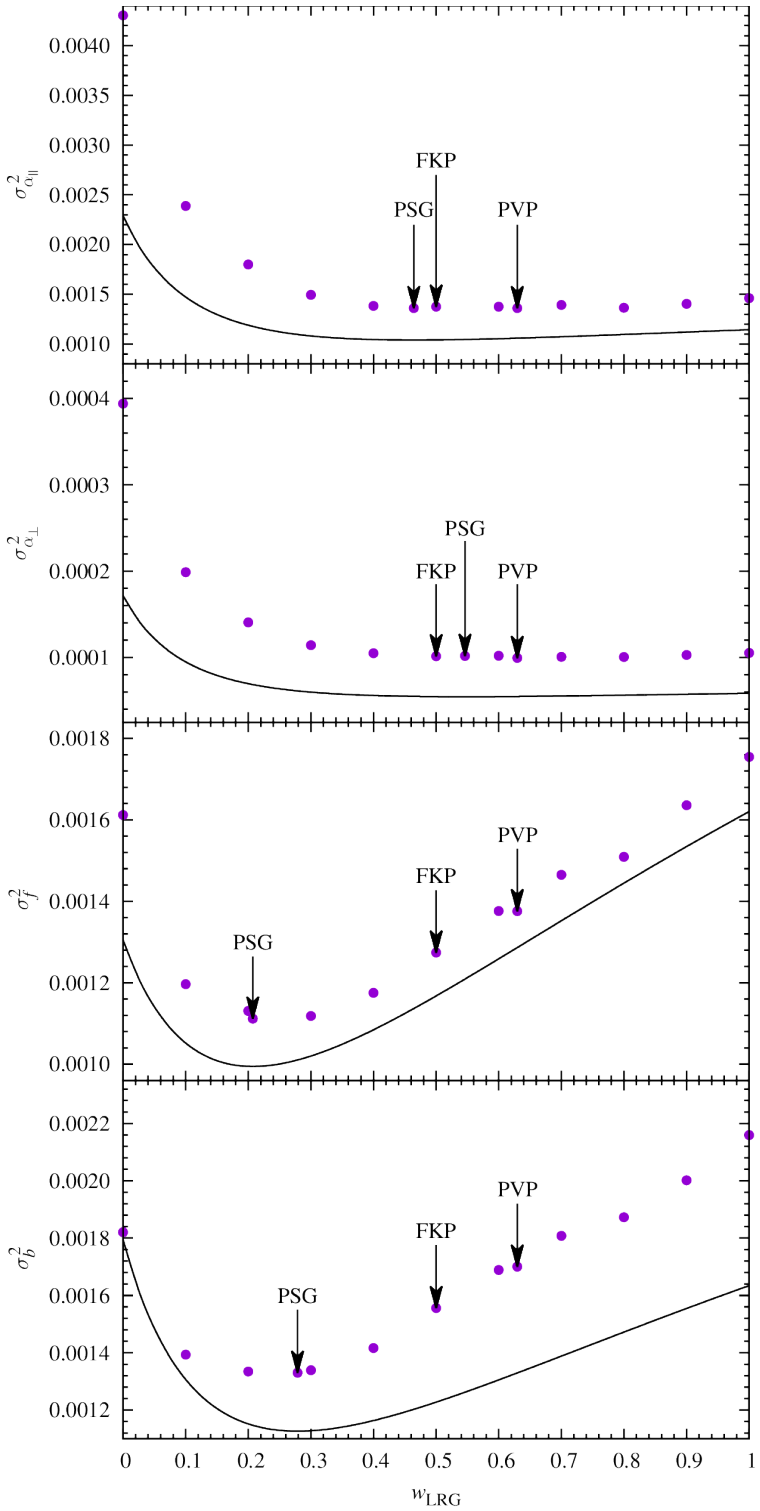

Figure 3.1: Variance of each parameter versus the relative weight of the LRGs for the first redshift bin of the DESI mocks. The points associated with the FKP, PVP, and PSG weights have been labelled. The solid lines are the theoretical predictions for the variance. In general, the points follow the shapes of the theoretical curves. For b and f, the PSG weights are clearly optimal, while for $\alpha_\perp$ and $\alpha_\parallel$ the variance is flat for a broad range of weights leading all weighting schemes to perform equally well.

Chapter 4

Part 3: Optimal Simulation Box Size

4.1 Introduction

In cosmology, simulations are a very useful tool to understand the dynamics of evolution in a given theoretical framework. In principle, cosmological simulations have the impossible task of simulating an infinite universe. In practice however, one fixes a size of the simulation box and imposes periodic boundary conditions. The following question then arises- for a given scale of interest, what should be the minimum simulation box-size that one needs to consider so as to leave the predictions from the simulation consistent with that from observation? This is the question that we will try to address in this part of the thesis.

By fixing a box-size, we naturally fix the lowest mode in the Fourier space for the simulation. Let us refer to this Fourier mode as k_{min} where $k_{min} = \frac{2\pi}{L_{box}}$. Consequently, to investigate the effects of fixing a box-size, it is sufficient to look at how the choice of k_{min} affects the predictions of the simulation. We use Lagrangian Perturbation Theory (LPT), which is easily expressed in Fourier space, to model the results of the simulation and probe the consequences of using a particular box-size. In particular, we look at the effects of changing k_{min} on the scale dilation factor which essentially captures the relative position of the BAO peak in the simulation compared to the true peak.

4.2 Methodology

In fluid mechanics, Eulerian and Lagrangian specifications are related by the material derivative, which is interpreted as total rate of change of flow as experienced by a specific flow parcel. Analogously in structure formation, the Lagrangian description relates the Eulerian position to its initial, or Lagrangian, position through a displacement vector field.

As shown in[45] Padmanabhan et al., in LPT the power spectrum is of the form

$$P(k) = P_{lin}(k) + P_{mc}(k) \tag{4.1}$$

$$P_{obs}(k) = e^{-k^2\Sigma^2/2} P_{lin}(k) + P_{mc}(k) + ... \tag{4.2}$$

where $\Sigma = [\int_0^\infty dk P(k)/(6\pi^2)]^{1/2}$ and P_{mc} are the mode coupling terms[45;46] the explicit form of which is relegated to App. B.

The exponential damping of the linear power spectrum by Σ or equivalently the smoothing of the correlation function reduces the precision with which the size of BAO ruler may be measured. Thus if we incorrectly calculate Σ by raising k_{min}, it will affect our measurements of the standard ruler. Σ with a raised k_{min} is defined as Σ_{min} as can be seen in Eq 4.3.

Our power spectrum model is given by:

$$\Sigma^2_{min} = \int_{k_{min}}^\infty dk P(k)/(6\pi^2) \tag{4.3}$$

$$P_{true}(k) = e^{-k^2\Sigma^2/2} P_{lin}(k) \tag{4.4}$$

$$P^{fit}_{min}(\alpha, k) = e^{-(\alpha k)^2\Sigma^2_{min}} P_{lin}(\alpha k) \tag{4.5}$$

$$\chi^2(\alpha, k_{min}) = \sum_{i=k_{min}} \frac{(P_{true}(k_i) - \bar{P}^{fit}_{min}(\alpha k_i))^2}{\sigma^2(k_i)} \tag{4.6}$$

where $\sigma^2(k) \propto k^2\, dk$ and $P_{true}(k) = P_{obs}(k_{min} = 0, k)$.

The scale dilation factor α, similar to the one used in Anderson et al.[47], in Eq 4.5 captures the k-mode distance constraints. It measures the relative position of the BAO peak as predicted by P_{true} versus the model, thereby characterizing any observed shift.

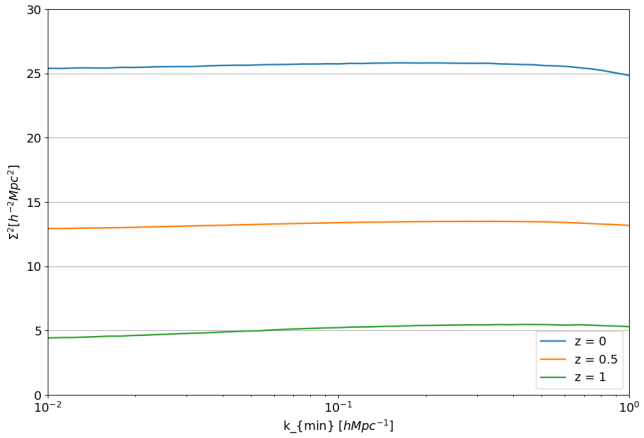

Figure 4.1: The plot shows the effect of increasing k_{min} on Σ^2 which is referred to as Σ^2_{min}. The reasons for deviations in Σ^2_{min} become obvious once one observes the contribution to power from mode-coupling terms in App B.

We obtain the best fit value of α by computing χ^2 goodness-of-fit indicator as defined in Eq 4.6 for $0.9 < \alpha < 1.10$ and define the best-fit value α_{opt} as the value that minimizes χ^2. The deviation of α_{opt} from 1 clearly indicates a significant loss of information from the low k-modes.

4.3 Results and Conclusion

Results in this section are derived assuming a spatially flat ΛCDM cosmological model with $\Omega_{\rm m} = 0.28$, $\Omega_\Lambda = 0.72$. We get the linear power spectra at various redshifts using Code for Anisotropies in the Microwave Background (CAMB) created by Antony Lewis and Anthony Challinor.[48] For the analysis we look at 3 different redshifts, $z = 0, 0.5, 1$.

We notice that Σ^2_{min} (Fig. 4.1) varies slightly as a function of k_{min}. However, since the exponential damping term, as in Eq. 4.1 consists of $k^2\Sigma^2$, even slight changes can make a considerable impact on large k-modes or equivalently small physical length scales.

These changes are evident in Fig. 4.2 where one can notice the deviation in the χ^2 minimizing scale factor α as a function of k_{min}. The trend that seems to emerge is that the deviation starts at higher k_{min} or equivalently smaller box-sizes for higher redshifts with

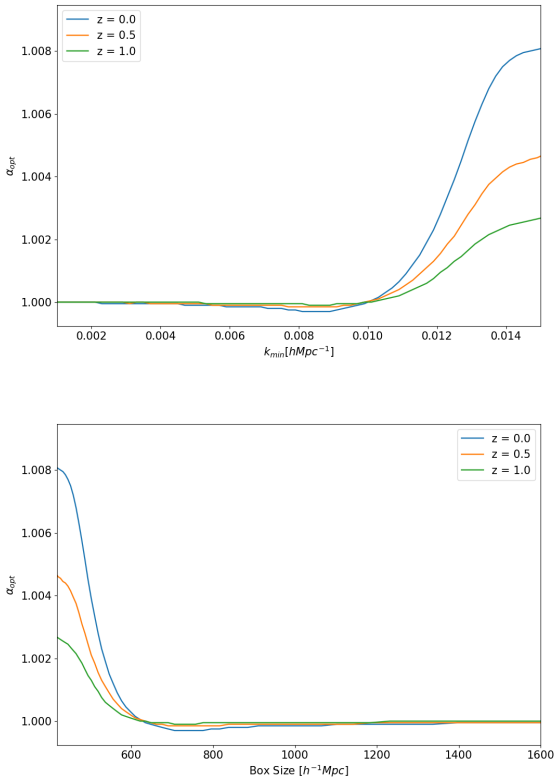

Figure 4.2: Dependence of χ^2 minimizing α on k_{min} and equivalently box size for various redshifts.

the biggest box size needed to acccurately describe redshift of 0. This result can be easily explained by studying the structure formation of the Universe and reminding oneself that a significant portion of the non-linear evolution actually happens at lower redshifts. In other words, at lower redshifts the small scale (physical distance) modes actually affect the large scale modes (hence non-linearity), thus if one excludes the large scale modes (by increasing k_{min}), one cannot expect to get accurate description of evolution of the small scale modes.

Thus, fig. 4.2 is consistent with our understanding of structure formation. From the plots, we conclude that if the redshift of interest is 0, a box size of at least 1 Gpc^3 should be sufficient to capture the features of interest in a simulation. If one intends to study higher redshifts, a smaller box size may be more optimal with each side 900 Mpc and 750 Mpc for redshift 0.5 and 1 respectively.

Chapter 5

Conclusion

In this thesis, we looked at different tools and techniques used in understanding the large scale structure of the Universe.

- Bispectrum: Tool for constraining cosmological parameters
- Weighting Scheme: Technique for optimizing the analysis of data
- Cosmological Simulations: Useful for simulating and understanding the cosmological models

In the thesis, we firstly develop a theory of angular multipoles of the bispectrum and propose a significant optimization which makes it practical to use.

Secondly, we look at how different weighting schemes can provide better constraints on cosmological parameters. Subsequently, we provide an optimal weighting scheme for minimizing the constraints on any parameter of interest.

Lastly, we determine the minimum box length of a cosmological simulation required to for the evolution up to the redshift of interest.

Thus, we looked at various aspects of understanding the LSS and inferring useful information from vast amounts of cosmological data available.

Bibliography

[1] Peter AR Ade, N Aghanim, M Arnaud, M Ashdown, J Aumont, C Baccigalupi, AJ Banday, RB Barreiro, JG Bartlett, N Bartolo, et al. Planck 2015 results-xiii. cosmological parameters. *Astronomy & Astrophysics*, 594:A13, 2016.

[2] Bela Abolfathi, DS Aguado, Gabriela Aguilar, Carlos Allende Prieto, Andres Almeida, Tonima Tasnim Ananna, Friedrich Anders, Scott F Anderson, Brett H Andrews, Borja Anguiano, et al. The fourteenth data release of the sloan digital sky survey: first spectroscopic data from the extended baryon oscillation spectroscopic survey and from the second phase of the apache point observatory galactic evolution experiment. *The Astrophysical Journal Supplement Series*, 235(2):42, 2018.

[3] DESI Collaboration. Science Final Design Report, US Department of Energy Office of Science, 2016.

[4] Michael Levi, Chris Bebek, Timothy Beers, Robert Blum, Robert Cahn, Daniel Eisenstein, Brenna Flaugher, Klaus Honscheid, Richard Kron, Ofer Lahav, et al. The desi experiment, a whitepaper for snowmass 2013. *arXiv preprint arXiv:1308.0847*, 2013.

[5] Paul A Abell, David L Burke, Mario Hamuy, Martin Nordby, Tim S Axelrod, David Monet, Bojan Vrsnak, Paul Thorman, DR Ballantyne, Joshua D Simon, et al. Lsst science book, version 2.0. Technical report, 2009.

[6] Phillip James Edwin Peebles. *The large-scale structure of the universe*. Princeton university press, 1980.

[7] Praful Gagrani and Lado Samushia. Information content of the angular multipoles of redshift-space galaxy bispectrum. *Monthly Notices of the Royal Astronomical Society*, 467(1):928–935, 2017.

[8] David W Pearson, Lado Samushia, and Praful Gagrani. Optimal weights for measuring redshift space distortions in multitracer galaxy catalogues. *Monthly Notices of the Royal Astronomical Society*, 463(3):2708–2715, 2016.

[9] Nick Kaiser. Clustering in real space and in redshift space. *Monthly Notices of the Royal Astronomical Society*, 227(1):1–21, 1987.

[10] Charles Alcock and Bohdan Paczynski. An evolution free test for non-zero cosmological constant. *Nature*, 281:358, 1979.

[11] AN Taylor and AJS Hamilton. Non-linear cosmological power spectra in real and redshift space. *Monthly Notices of the Royal Astronomical Society*, 282(3):767–778, 1996.

[12] Atsushi Taruya, Shun Saito, and Takahiro Nishimichi. Forecasting the cosmological constraints with anisotropic baryon acoustic oscillations from multipole expansion. *Physical Review D*, 83(10):103527, 2011.

[13] Eyal A Kazin, Michael R Blanton, Román Scoccimarro, Cameron K McBride, Andreas A Berlind, Neta A Bahcall, Jon Brinkmann, Paul Czarapata, Joshua A Frieman, Stephen M Kent, et al. The baryonic acoustic feature and large-scale clustering in the sloan digital sky survey luminous red galaxy sample. *The Astrophysical Journal*, 710 (2):1444, 2010.

[14] Florian Beutler, Shun Saito, Hee-Jong Seo, Jon Brinkmann, Kyle S Dawson, Daniel J Eisenstein, Andreu Font-Ribera, Shirley Ho, Cameron K McBride, Francesco Montesano, et al. The clustering of galaxies in the sdss-iii baryon oscillation spectroscopic survey: testing gravity with redshift space distortions using the power spectrum multipoles. *Monthly Notices of the Royal Astronomical Society*, 443(2):1065–1089, 2014.

[15] Yong-Seon Song, Atsushi Taruya, and Akira Oka. Cosmology with anisotropic galaxy clustering from the combination of power spectrum and bispectrum. *Journal of Cosmology and Astroparticle Physics*, 2015(08):007, 2015.

[16] Bradley Greig, Eiichiro Komatsu, and J Stuart B Wyithe. Cosmology from clustering of lyα galaxies: breaking non-gravitational lyα radiative transfer degeneracies using the bispectrum. *Monthly Notices of the Royal Astronomical Society*, 431(2):1777–1794, 2013.

[17] Roman Scoccimarro, HMP Couchman, and Joshua A Frieman. The bispectrum as a signature of gravitational instability in redshift space. *The Astrophysical Journal*, 517 (2):531, 1999.

[18] E. Sefusatti and E. Komatsu. Bispectrum of galaxies from high-redshift galaxy surveys: Primordial non-Gaussianity and nonlinear galaxy bias. , 76(8):083004, October 2007. doi: 10.1103/PhysRevD.76.083004.

[19] M. Tellarini, A. J. Ross, G. Tasinato, and D. Wands. Galaxy bispectrum, primordial non-Gaussianity and redshift space distortions. , 6:014, June 2016. doi: 10.1088/1475-7516/2016/06/014.

[20] Emiliano Sefusatti, Martın Crocce, and Vincent Desjacques. The halo bispectrum in n-body simulations with non-gaussian initial conditions. *Monthly Notices of the Royal Astronomical Society*, 425(4):2903–2930, 2012.

[21] Zachary Slepian and Daniel J Eisenstein. Modeling the large-scale redshift-space 3-point correlation function of galaxies. *arXiv preprint arXiv:1607.03109*, 2016.

[22] Zachary Slepian, Daniel J Eisenstein, Joel R Brownstein, Chia-Hsun Chuang, Héctor Gil-Marín, Shirley Ho, Francisco-Shu Kitaura, Will J Percival, Ashley J Ross, Graziano Rossi, et al. Detection of baryon acoustic oscillation features in the large-scale 3-point correlation function of sdss boss dr12 cmass galaxies. *arXiv preprint arXiv:1607.06097*, 2016.

[23] Héctor Gil-Marín, Jorge Noreña, Licia Verde, Will J Percival, Christian Wagner, Marc Manera, and Donald P Schneider. The power spectrum and bispectrum of sdss dr11

boss galaxies–i. bias and gravity. *Monthly Notices of the Royal Astronomical Society*, 451(1):5058–5099, 2015.

[24] Héctor Gil-Marín, Will J Percival, Licia Verde, Joel R Brownstein, Chia-Hsun Chuang, Francisco-Shu Kitaura, Sergio A Rodríguez-Torres, and Matthew D Olmstead. The clustering of galaxies in the sdss-iii baryon oscillation spectroscopic survey: Rsd measurement from the power spectrum and bispectrum of the dr12 boss galaxies. *arXiv preprint arXiv:1606.00439*, 2016.

[25] Román Scoccimarro. Fast estimators for redshift-space clustering. *Physical Review D*, 92(8):083532, 2015.

[26] PAR Ade, N Aghanim, C Armitage-Caplan, M Arnaud, M Ashdown, F Atrio-Barandela, J Aumont, C Baccigalupi, Anthony J Banday, RB Barreiro, et al. Planck 2013 results. xvi. cosmological parameters. *Astronomy & Astrophysics*, 571:A16, 2014.

[27] David J Schlegel, Chris Bebek, Henry Heetderks, Shirley Ho, Michael Lampton, Michael Levi, Nick Mostek, Nikhil Padmanabhan, Saul Perlmutter, Natalie Roe, et al. Bigboss: the ground-based stage iv dark energy experiment. *arXiv preprint arXiv:0904.0468*, 2009.

[28] R Laureijs, J Amiaux, S Arduini, J-L Augueres, J Brinchmann, R Cole, M Cropper, C Dabin, L Duvet, A Ealet, et al. Euclid definition study report. *arXiv preprint arXiv:1110.3193*, 2011.

[29] Román Scoccimarro. The bispectrum: from theory to observations. *The Astrophysical Journal*, 544(2):597, 2000.

[30] WE Ballinger, JA Peacock, and AF Heavens. Measuring the cosmological constant with redshift surveys. *arXiv preprint astro-ph/9605017*, 1996.

[31] Fergus Simpson and John A Peacock. Difficulties distinguishing dark energy from modified gravity via redshift distortions. *Physical Review D*, 81(4):043512, 2010.

[32] Lado Samushia, Will J Percival, Luigi Guzzo, Yun Wang, Andrea Cimatti, Carlton Baugh, James E Geach, Cedric Lacey, Elisabetta Majerotto, Pia Mukherjee, et al. Effects of cosmological model assumptions on galaxy redshift survey measurements. *Monthly Notices of the Royal Astronomical Society*, 410(3):1993–2002, 2011.

[33] JC Jackson. A critique of rees's theory of primordial gravitational radiation. *Monthly Notices of the Royal Astronomical Society*, 156(1):1P–5P, 1972.

[34] Hume A Feldman, Nick Kaiser, and John A Peacock. Power spectrum analysis of three-dimensional redshift surveys. *arXiv preprint astro-ph/9304022*, 1993.

[35] Emiliano Sefusatti, Martín Crocce, Sebastián Pueblas, and Román Scoccimarro. Cosmology and the bispectrum. *Physical Review D*, 74(2):023522, 2006.

[36] Kazuhiro Yamamoto, Masashi Nakamichi, Akinari Kamino, Bruce A Bassett, and Hiroaki Nishioka. A measurement of the quadrupole power spectrum in the clustering of the 2df qso survey. *Publications of the Astronomical Society of Japan*, 58(1):93–102, 2006.

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

[38] Andreas Albrecht, Gary Bernstein, Robert Cahn, Wendy L Freedman, Jacqueline Hewitt, Wayne Hu, John Huth, Marc Kamionkowski, Edward W Kolb, Lloyd Knox, et al. Report of the dark energy task force. *arXiv preprint astro-ph/0609591*, 2006.

[39] I. Szapudi. Introduction to Higher Order Spatial Statistics in Cosmology. In V. J. Martínez, E. Saar, E. Martínez-González, and M.-J. Pons-Bordería, editors, *Data Analysis in Cosmology*, volume 665 of *Lecture Notes in Physics, Berlin Springer Verlag*, pages 457–492, 2009. doi: 10.1007/978-3-540-44767-2_14.

[40] J. Carron and M. C. Neyrinck. On the Inadequacy of N-point Correlation Functions to Describe Nonlinear Cosmological Fields: Explicit Examples and Connection to Simulations. , 750:28, May 2012. doi: 10.1088/0004-637X/750/1/28.

[41] J. Carron and I. Szapudi. Sufficient observables for large-scale structure in galaxy surveys. , 439:L11–L15, March 2014. doi: 10.1093/mnrasl/slt167.

[42] Will J Percival, Beth A Reid, Daniel J Eisenstein, Neta A Bahcall, Tamas Budavari, Joshua A Frieman, Masataka Fukugita, James E Gunn, Željko Ivezić, Gillian R Knapp, et al. Baryon acoustic oscillations in the sloan digital sky survey data release 7 galaxy sample. *Monthly Notices of the Royal Astronomical Society*, 401(4):2148–2168, 2010.

[43] D Spergel, N Gehrels, J Breckinridge, M Donahue, A Dressler, BS Gaudi, T Greene, O Guyon, C Hirata, J Kalirai, et al. Wide-field infrared survey telescope-astrophysics focused telescope assets wfirst-afta final report. *arXiv preprint arXiv:1305.5422*, 2013.

[44] Will J Percival, Licia Verde, and John A Peacock. Fourier analysis of luminosity-dependent galaxy clustering. *Monthly Notices of the Royal Astronomical Society*, 347 (2):645–653, 2004.

[45] Nikhil Padmanabhan, Martin White, and JD Cohn. Reconstructing baryon oscillations: A lagrangian theory perspective. *Physical Review D*, 79(6):063523, 2009.

[46] Martin White. Reconstruction within the zeldovich approximation. *Monthly Notices of the Royal Astronomical Society*, 450(4):3822–3828, 2015.

[47] Lauren Anderson, Eric Aubourg, Stephen Bailey, Dmitry Bizyaev, Michael Blanton, Adam S Bolton, Jon Brinkmann, Joel R Brownstein, Angela Burden, Antonio J Cuesta, et al. The clustering of galaxies in the sdss-iii baryon oscillation spectroscopic survey: baryon acoustic oscillations in the data release 9 spectroscopic galaxy sample. *Monthly Notices of the Royal Astronomical Society*, 427(4):3435–3467, 2012.

[48] Antony Lewis and Anthony Challinor. Camb: Code for anisotropies in the microwave background. *Astrophysics Source Code Library*, 2011.

[49] Benjamin Grinstein and Mark B Wise. On the validity of the zel'dovich approximation. *The Astrophysical Journal*, 320:448–453, 1987.

Appendix A

Robustness of Multipole Decomposition

We find our main conclusion – that the first three even ℓ modes of the bispectrum contain most of the cosmological information – to be robust with respect to various assumptions. To show that this assumption is robust with respect to the choice of $k_{\rm max}$ we repeat the computations of Sec. 2.5 for $k_{\rm max} = 0.1\ h/\text{Mpc}$. These results are presented on Fig. A.1 which is virtually indistinguishable from Fig. 2.1. The only thing that changes is the relative constraining power of the bispectrum compared to the power spectrum which scales steeply with the value of $k_{\rm max}$. Even for $k_{\rm max} = 0.1\ h/\text{Mpc}$ however, the bispectrum constraints on f are as good as the ones resulting from the power spectrum.

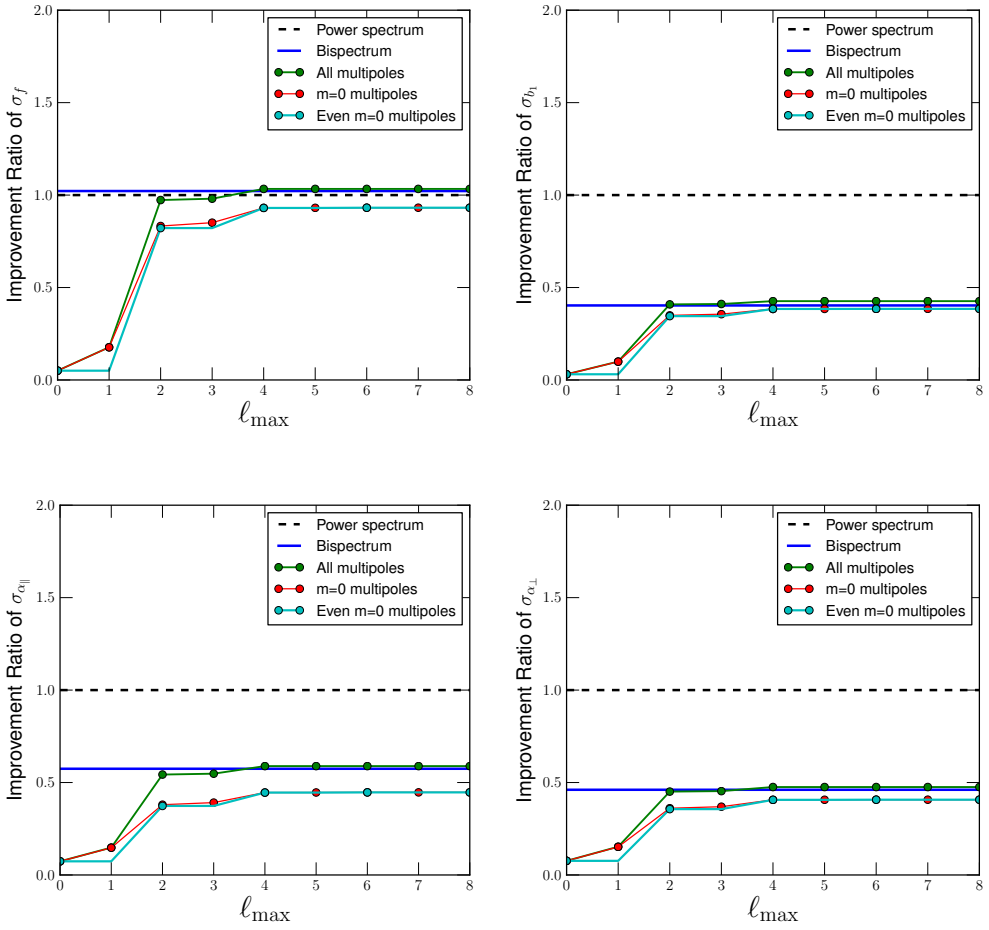

Figure A.1: Cosmological constraints expected from the bispectrum multipoles as a function of maximum ℓ used in the analysis for a sample of DESI LRGs in $0.6 < z < 0.7$ considering strictly linear scales of $k < 0.1$ Mpc/h. The constraints from power spectrum and the full bispectrum are also displayed for comparison.

Appendix B

Mode Coupling Terms

In the Zeldovich approximation, assuming azimuthal symmetry, the second order contribution to $P(k)$ i.e. $P_{mc}(k)$ takes the following form:

$$P^{(2,2)}(k) = \frac{k^3}{2\pi^2} \int_0^\infty dr P(kr) \int_{-1}^{+1} d\mu P(k\sqrt{1+r^2-2r\mu}) \frac{\mu^2(\mu-r)^2}{4(1-2\mu r+r^2)^2} \tag{B.1}$$

$$P^{(1,3)}(k) = -k^2 P_{Lin}(k) \int_0^\infty \frac{dp}{6\pi^2} P_{Lin}(p) \tag{B.2}$$

$$P_{mc}(k) = P^{(2,2)}(k) + P^{(1,3)}(k) \tag{B.3}$$

These equations are the same as in[46] White 2015. The detailed derivation can be found in[49] Grinstein et al.

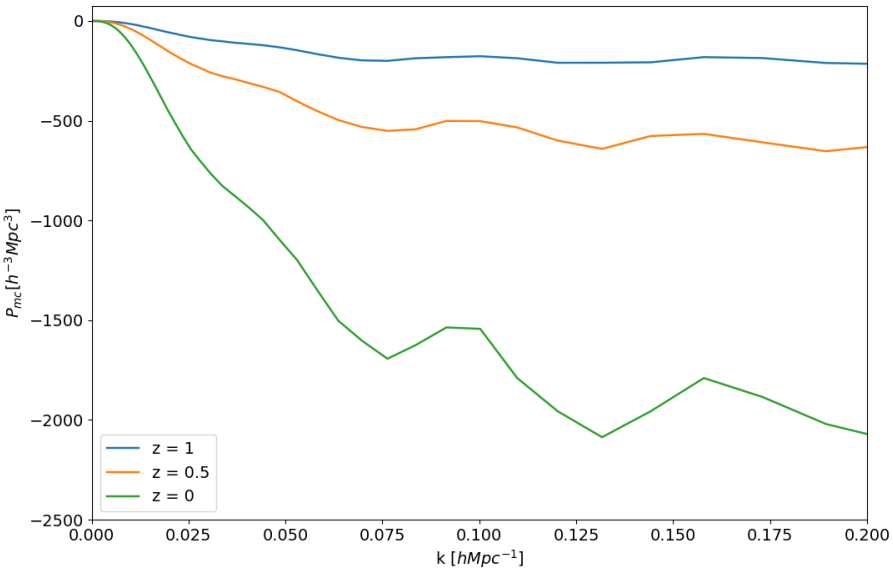

Figure B.1: 2nd order non linear contributions to the power spectrum