



**UCL**

**Cosmology from  
compressed high-order statistics  
in galaxy surveys**

**Davide Gualdi**

*Supervised by:*

Prof. Ofer Lahav and Dr. Benjamin Joachimi

University College London  
Department of Physics and Astronomy

*A thesis submitted for the degree of  
Doctor of Philosophy*

July, 2018

## Statement of Originality

I, Davide Gualdi, confirm that the work presented in this PhD. thesis is my own. Where scientific results and information have been derived from other sources, I confirm that the proper references have been included and indicated in the thesis.

Davide Gualdi

---

*Signature*

---

*Date*

## Acknowledgements

My sincere gratitude and utmost respect go to my two supervisors Prof. Ofer Lahav and Dr. Benjamin Joachimi. Prof. Lahav, beside being an expert mentor in terms of cosmological knowledge, he has always created through his kind behaviour a very positive and supporting environment in which also not an exactly self-confident person like me could grow both personally and scientifically. Dr. Joachimi never stopped to astonish me for his logical brightness and ability to get to the core of every topic. It would have been difficult to find a better example of what a scientist should be. Moreover, in every discussion regarding the projects, even if 98 % of the times I was wrong, he always patiently listened to me. He limited himself to ask questions targeted at making myself realise the faults in my arguments, from which I could improve.

Even if officially he has not been one of my supervisors, a good part of the work contained in this thesis would have not been possible without the help and insights given to me by Dr. Marc Manera. Together with being very smart, he has always been generous with his time and altruistic with his work by helping many PhD. students here at UCL.

An heartfelt thank you to my friends and colleagues from the very beginning: Keir Rogers, William Jennings and John Yue Han Soo. Thanks to their friendship I felt at home in the U.K. and I had the chance to learn and to appreciate (and hopefully also a bit assimilate) the best aspects of the British and Malaysian values and customs.

During these years at UCL I would have much more often noticed the absence of the sun from the English sky if it wouldn't have been for the many friends and colleagues encountered along the road, and whose company I have enjoyed. Being all very bright people I also had the chance to learn something from everyone of them and I hope there will be the chance to keep doing so in the future. Thanks to all of you: Niall Jeffrey, Felix Presley, Antonella Palmese, Luisa Lucie Smith, Robert Schuhmann, Will Hartley, Daniela Sadeeh, Giuseppe Morello,

---

Tiziano Zingales, Giuliana Cosentino, Mario Damiano, Martin Rey, Gioacchino Accurso, Fabian Krause, Henrique Xavier, Arthur Loureiro, Arianna Sorba, Edward Scott, Bruno Moraes, Lorne Whiteway, Roger Wesson, Jonathan Braden.

I am really grateful to Héctor Gil-Marin for allowing me to use his measurements on the BOSS data which were essential in order to produce the results presented in Chapter 4.

If I followed the physics path, it's thanks to my inspiring and kind high school teacher Prof. Claudio Borghi.

If I continued on the physics path, it's thanks to the friends I met while studying it at the University of Trento: Dante Bonolis, Sunny Vagnozzi, Lorenzo Andreoli, Lorenzo Festa and Paolo Mori.

If I reached this goal, it's because of the example and hard work of my grandparents. My grandfather Enea taught me to be patient. My grandfather Emilio has shown me with his life that worthy achievements rarely come without sacrifices.

I have been able to make so many valuable friendships and connections along the way thanks to the example of my two cousins: Dario with his unlimited kindness and Beatrice with her amazing wisdom.

I don't even know where to start to thank my wonderful parents. Their support and love and how incredibly lucky and honoured I feel to be their son cannot be quantified by words. My gratitude will never be enough for all the efforts and sacrifices they made in order to give me all the chances that they never had and which they deserved much more than I do.

Finally thanks to 我心中的快乐Liu Qun, her smile and love are the center of my Universe.

# Abstract

The work presented in this thesis focuses on developing compression techniques to exploit fully the constraining power of high-order statistics when applied to the cosmological observable of interest. We present four different methods in the three-point (3pt) case. The mathematical theoretical framework is first developed and then followed, for all the methods, by application on real data. In particular we use data from the CMASS sample of the Sloan Digital Sky Survey III BOSS Data Releases 11 and 12. Our compression results are compared to those obtained via standard analysis, for example Markov chain Monte Carlo (MCMC) sampling.

First, we consider the three-point auto-correlation function as an integrated compressed version of the standard correlation one. We derive analytic expressions including corrections for the Primordial non-Gaussianity. We then test the model on data to constrain cosmological parameters.

Secondly, we explore two methods of compressing the redshift-space galaxy power spectrum and bispectrum with respect to a chosen set of cosmological parameters. Both methods transform the original data-vector into a compressed one with dimension equal to the number of model parameters considered using the Multiple Optimised Parameter Estimation and Data compression algorithm (MOPED) algorithm. Analytic expressions for the covariance matrix are derived in order both to compress the data-vector and to test the compression performance by comparing with standard MCMC sampling on the full data-vector.

Finally, we apply our compression methods to the galaxy power spectrum monopole, quadrupole and bispectrum monopole measurements from the BOSS DR12 CMASS sample. We derive an analytic expression for the covariance matrix of the new data-vector. We show that compression allows a much longer data-vector to be used, returning tighter constraints on the cosmological parameters of interest.

# Impact Statement

The work presented in this thesis has the potential to become the standard procedure to analyse three-point (and higher-order) statistics in future data sets like those from the "Dark Energy Spectroscopic Instrument" (DESI), "Euclid" and the "Subaru Prime Focus Spectroscopic" (PFS). The probe on which the statistics used in this work focus is galaxy clustering. The surveys mentioned above will create much larger and richer catalogues of galaxies than previous generation surveys. This will require us to use the full potential of statistical tools available in order to extract all the information contained in new data sets. In particular, this can be interpreted as using all the possible elements of data-vectors representing the statistics employed for the analysis of the data. Limitations imposed by the high cost of computational power needed to analyse the full data-vector can be overcome using the compression methods presented here.

Moreover, these techniques can be applied to higher-order statistics applied for other cosmological probes, for example weak lensing, cosmic microwave background anisotropies, 21cm emission lines and Ly- $\alpha$  forest.

The analytic work presented here also lays the foundation for a systematic approach in the derivation of similar expressions for higher-order statistics, or the same three-point statistics in other fields.

Outside of the astronomical environment, the compression methods can also be applied to any situation in which it is possible to model the covariance matrix of the data-vector of interest. In several areas outside academia it can be interesting to study the deviations from Gaussianity of statistical fields.

In this era of "Big Data", one of the most pressing active research areas is indeed the compression of large data-sets.

"Good Morning!" said Bilbo, and he meant it. The sun was shining, and the grass was very green. But Gandalf looked at him from under long bushy eyebrows that stuck out further than the brim of his shady hat.

"What do you mean?" he said. "Do you wish me a good morning, or mean that it is a good morning whether I want it or not; or that you feel good this morning; or that it is a morning to be good on?"

"All of them at once," said Bilbo. "And a very fine morning for a pipe of tobacco out of doors, into the bargain.

...

"Good morning!" he said at last. "We don't want any adventures here, thank you! You might try over The Hill or across The Water." By this he meant that the conversation was at an end. "What a lot of things you do use Good morning for!" said Gandalf. "Now you mean that you want to get rid of me, and that it won't be good till I move off."

— J.R.R. Tolkien, *The Hobbit*

# Thesis roadmap

Current and future surveys like the Baryon Oscillation Spectroscopic Survey (BOSS), the Dark Energy Survey (DES), the Dark Energy Spectroscopic Instrument (DESI), Euclid, the Subaru Prime Focus Spectroscopy (PFS), will produce extremely large and rich of information data-sets which require an improvement of the current analysis techniques. In particular, beyond the level of two points (2pt) correlation functions (or power spectra in Fourier space), 3pt statistics represent the simplest higher-order alternative which can be used to study the non-Gaussian components of the data. 3pt statistics are therefore fundamental in better constraining the cosmological parameter space (e.g. how much dark matter/energy are there in the Universe?) in order to discriminate between different models of the Universe. However, 3pt statistics are both theoretically and numerically more difficult to measure and to model than the 2pt ones, suffering also of the limited number of computationally expensive simulations needed to estimate the errors on their measurements. Efficiently compressing (minimal information loss) 3pt statistics (in ways easily extendable to higher-order statistics like 4-5pt ones, etc.) provides an effective solution to their drawbacks and enhances their application to the cosmological data-sets.

In Chapter 2 we derive analytical expressions for the 2pt and 3pt auto-correlation functions and for their ratio, the skewness. All quantities have been derived considering also primordial non Gaussianity contributions. These are integrated statistics, namely of the 2pt and 3pt correlation functions. By integrating, the length of the data-vector is shortened. Nevertheless, as the measurement on BOSS DR11 reveals, part of the information is lost.

In Chapter 3 are presented two compression techniques for the redshift space galaxy power spectrum and bispectrum, both described in [Gualdi et al., 2018b](#). The first method consists in running an MCMC sampling on the compressed data vector. In the second method, the parameter space is first orthogonalised using a principal component analysis transformation. In this way, the transformed parameters are uncorrelated. It is then possible to reconstruct the multidimensional posterior distribution by randomly sampling the 1D posterior distributions of the transformed parameter set.

---

Chapter 4 describes the application of the compression methods presented in Chapter 3 to the measurements from the DR12 BOSS CMASS sample of the power spectrum (monopole and quadrupole) and the bispectrum (monopole). This work has been reported in [Galdi et al., 2018a](#). In particular we studied what happens to the 1 and 2D posterior distributions of the model parameters when many more triangle configurations are included in the bispectrum part of the data-vector.

Finally, we conclude in Chapter 5 where it is also presented an alternative way to compress in general 3pt statistics using the geometrical properties of a triangle configuration. This method performance is then applied to the measurement from the DR12 BOSS CMASS sample of the bispectrum monopole, and compared to the performances of the methods described in Chapters 3 and 4.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Cosmology . . . . .	1
1.1.1	Friedmann equation and cosmological evolution . . . . .	3
1.1.2	Inflation . . . . .	6
1.1.3	Cosmic Microwave Background . . . . .	7
1.1.4	Dark matter . . . . .	9
1.1.5	Dark energy . . . . .	12
1.1.6	Cosmological probes . . . . .	15
1.1.7	Best-fit cosmology . . . . .	19
1.2	Perturbation theory . . . . .	21
1.2.1	Static Universe . . . . .	22
1.2.2	Expanding Universe . . . . .	24
1.3	Statistics of the density field . . . . .	27
1.3.1	From primordial to late-time matter power spectrum . . . . .	30
1.3.2	Three-point statistics . . . . .	32
1.4	Galaxy bias . . . . .	33
1.5	Redshift-space distortions . . . . .	35
1.6	Data and simulations . . . . .	38
1.6.1	SDSS-III BOSS . . . . .	38
1.6.2	Galaxy mocks . . . . .	39
<b>2</b>	<b>Skewness: 3pt auto-correlation function</b>	<b>41</b>
2.1	Second-order perturbation theory . . . . .	42
2.2	Skewness derivation . . . . .	44
2.3	Skewness for a $\Omega_m = 0$ and $\Omega_\Lambda = 1$ Universe . . . . .	47
2.4	Primordial non-Gaussianity contribution . . . . .	50
2.4.1	Three-point contribution . . . . .	50
2.4.2	Two-point contribution . . . . .	56
2.5	$\mathbf{S}_3$ from data . . . . .	60
2.5.1	BOSS DR11 data and mocks . . . . .	62
2.5.2	Statistical estimators . . . . .	62

2.6	Treatment of noise sources . . . . .	64
2.7	Redshift-space distortions effect . . . . .	64
2.8	RSD-corrected expressions for $\langle\delta^2\rangle$ and $\langle\delta^3\rangle$ . . . . .	64
2.9	RSD-corrected expressions for $\langle\delta^2\rangle$ and $\langle\delta^3\rangle$ : $\mathbf{f}_{\text{NL}}$ terms . . . . .	67
2.10	Model parameters constraints from the joint data-vector $[\langle\delta^2\rangle; \langle\delta^3\rangle]$ . . . . .	69
2.11	Chapter recap . . . . .	71

## Appendices 73

2.A	Transfer function from the matter power spectrum . . . . .	73
-----	--	----

## 3 Maximal compression of the redshift space galaxy

### power spectrum and bispectrum 74

3.1	Abstract . . . . .	74
3.2	Introduction . . . . .	75
3.3	Perturbation theory with redshift-space distortions . . . . .	79
	3.3.1 Bias model . . . . .	79
	3.3.2 Redshift space formalism . . . . .	80
3.4	Covariance . . . . .	82
	3.4.1 $\mathbf{C}^{\text{PP}}$ : power spectrum covariance matrix . . . . .	82
	3.4.2 $\mathbf{C}^{\text{BB}}$ : bispectrum covariance matrix . . . . .	83
	3.4.3 $\mathbf{C}^{\text{BP}}$ : cross-variance matrix . . . . .	83
	3.4.4 Shot noise contribution . . . . .	84
3.5	Analysis setup . . . . .	84
3.6	Compression Formalism . . . . .	85
	3.6.1 Fisher information matrix . . . . .	85
	3.6.2 Compression Algorithm . . . . .	86
3.7	MCMC of Compressed data-vectors . . . . .	93
3.8	Posterior distributions directly from compressed data-vectors . . . . .	93
	3.8.1 Parameter space orthogonalisation - PCA . . . . .	95
	3.8.2 Comparison with MCMC sampling . . . . .	97
	3.8.3 Limitations of the compression . . . . .	98
3.9	Joint data-vector added value . . . . .	102
3.10	Conclusions . . . . .	103

## Appendices 106

3.A	Estimators definition and unbiasedness check . . . . .	106
	3.A.1 Power spectrum . . . . .	107
	3.A.2 Bispectrum estimator . . . . .	107
	3.A.3 Trispectrum definition . . . . .	109

3.A.4	Tetraspectrum definition . . . . .	112
3.A.5	Unconnected part of the five points correlation function . . . . .	114
3.B	Covariance terms derivation . . . . .	116
3.B.1	Covariance term: $\mathbf{C}^{\text{PP}}$ . . . . .	116
3.B.1.1	$\mathbf{C}_\text{p}^{\text{PP}}$ term: . . . . .	116
3.B.1.2	$\mathbf{C}_\text{T}^{\text{PP}}$ term: . . . . .	117
3.B.2	Covariance term: $\mathbf{C}^{\text{BB}}$ . . . . .	118
3.B.3	Cross - variance term: $\mathbf{C}^{\text{BP}}$ . . . . .	120
3.B.3.1	$\mathbf{C}_{\text{ml}}^{\text{BP}}$ term: . . . . .	120
3.B.3.2	$\mathbf{C}_{\text{Te}}^{\text{BP}}$ term: . . . . .	121
3.C	Compressed covariance matrix . . . . .	122
3.D	Weights orthogonalisation . . . . .	123
<b>4</b>	<b>Enhancing BOSS bispectrum</b>	
	<b>cosmological constraints</b>	
	<b>with maximal compression</b>	<b>126</b>
4.1	Abstract . . . . .	126
4.2	Introduction . . . . .	127
4.3	Data, mocks and analysis . . . . .	129
4.3.1	DR12 BOSS data and mocks catalogues . . . . .	129
4.3.2	Analysis settings . . . . .	129
4.4	Data-vector and covariance matrix . . . . .	131
4.4.1	Power spectrum monopole and quadrupole . . . . .	131
4.4.2	Analytical expression for $\text{P}_\text{g}^{(0,2)}$ covariance matrices . . . . .	132
4.4.3	Bispectrum monopole . . . . .	132
4.4.4	Analytical expression for $\text{B}_\text{g}^{(0)}$ covariance matrix . . . . .	133
4.4.5	Analytical expression for $[\text{P}_\text{g}^{(0,2)}, \text{B}_\text{g}^{(0)}]$ cross-covariance matrix	134
4.5	Compression methods . . . . .	135
4.5.1	PCA + MP . . . . .	135
4.5.2	Gaussianisation pre-step . . . . .	138
4.5.3	Analytical covariance matrix: usage . . . . .	138
4.6	Recover MCMC-derived posterior distribution . . . . .	139
4.7	Information content and number of triangles . . . . .	142
4.8	Consistency check . . . . .	144
4.8.1	Comparison with BOSS DR12 bias constraints . . . . .	145
4.8.2	Difference in time and computer resources needed . . . . .	146
4.9	Conclusions . . . . .	147

<b>Appendices</b>	<b>151</b>
4.A Estimators and covariance terms . . . . .	151
4.A.1 Power spectrum monopole/quadrupole and bispectrum monopole estimators . . . . .	151
4.A.2 Power spectrum monopole and quadrupole covariance ma- trix: Gaussian term . . . . .	154
4.A.3 Bispectrum monopole covariance matrix: Gaussian term .	155
4.A.4 Bispectrum monopole covariance matrix: non-Gaussian term . . . . .	156
4.A.5 Cross-covariance term . . . . .	157
4.B Validation tests . . . . .	159
<b>5 Conclusions and future work</b>	<b>163</b>
5.1 Chapter 2: skewness and 3pt auto-correlation function . . . . .	163
5.1.1 Chapter 2: future work . . . . .	164
5.2 Chapter 3: maximal compression . . . . .	164
5.2.1 Chapter 3: future work . . . . .	165
5.3 Chapter 4: Enhancing BOSS bispectrum . . . . .	166
5.3.1 Chapter 4: future work . . . . .	166
5.4 Geometrical compression . . . . .	167
5.4.1 New triangles parametrisation . . . . .	167
5.4.2 Number of bins: optimal choice . . . . .	169
5.4.3 MC-GC vs. MCMC vs. MC-MP vs. G-PCA . . . . .	170
<b>Bibliography</b>	<b>174</b>

# List of Figures

1.1	Cosmic Microwave Background . . . . .	10
1.2	Hubble diagram and residuals . . . . .	14
1.3	Matter power spectrum . . . . .	29
1.4	BOSS CMASS completeness map . . . . .	40
2.1	$\sigma^2$ , $\delta^3$ and $S_3$ computed in the case of $f_{NL} = 0$ and for $f_{NL} \Rightarrow$ . . .	57
2.2	Plot from Juszkiewicz (2013) showing the skewness BAO feature . . .	61
2.3	Auto-correlation functions and skewness for true and observed redshift . . . . .	65
2.4	Joint data vector parameter constraints . . . . .	71
3.1	Methods diagram . . . . .	78
3.2	Bispectrum case: 1D and 2D posterior distributions MCMC vs MP . . . . .	87
3.3	Joint data-vector $[P_g^s, B_g^s]$ posteriors . . . . .	88
3.4	Logarithm of the absolute value of the weights for all the triangle configurations used in the bispectrum data-vector . . . . .	90
3.5	Comparison between the MCMC-derived posteriors and the ones obtained using only the compression for $B_g^s$ . . . . .	91
3.6	1D and 2D posterior distributions using as data-vector $[P_g^s, B_g^s]$ MCMC vs. PCA + MP vs MP + ORT . . . . .	92
3.7	compression performance checks . . . . .	94
3.8	Forecasted posteriors for the redshift bin of a DESI-like survey . . . . .	99
4.1	Joint data-vector posteriors: MC-MP four-parameter case . . . . .	136
4.2	Joint data-vector posteriors: G-PCA four-parameters case . . . . .	137
4.3	Constraints improvement as a function of the number of triangles used. Check for shifted fiducial cosmology. . . . .	143
4.4	Reduced $\chi^2$ and $p$ -values for the best-fit models obtained using the MCMC, MC-MP and G-PCA compression methods . . . . .	144
4.5	MCMC vs. MC-MP vs. G-PCA . . . . .	148
4.A.1	Bispectrum monopole integration volume . . . . .	153

---

4.B.1 Joint data-vector posteriors: MC-MP and G-PCA four-parameter $\Delta k_6$ case . . . . .	160
4.B.2 Joint data-vector posteriors: MC-MP and G-PCA four-parameter $\Delta k_5$ case . . . . .	160
4.B.3 Reduced $\chi^2$ and $p$ -values for the best-fit parameters with varying $\sigma_8$	162
5.1 MCMC vs. MC-MP vs. G-PCA vs. MC-GC . . . . .	172

# List of Tables

- 1.1 cosmological surveys . . . . . 16
- 1.2 best-fit  $\Lambda$ CDM parameters from *Planck* . . . . . 20
- 1.3 Specifications for the DR11 and DR12 BOSS-SDSS CMASS samples 39
  
- 3.1 The 68% confidence intervals of the 1D posteriors for the bispectrum and the joint data-vector . . . . . 101
  
- 4.1 Four parameter-case, check consistency . . . . . 140
- 4.2 Four-parameter case, constraints improvement . . . . . 141
- 4.B.1 Four parameter-case, checking consistency for shifted fiducial cosmology . . . . . 161
  
- 5.1 Four parameter-case, best-fit parameters . . . . . 171
- 5.2 Four-parameter case, constraints improvement . . . . . 171

*“It’s a dangerous business, Frodo, going out your door... You step into the Road, and if you don’t keep your feet, there is no knowing where you might be swept off to.”*

- J.R.R. Tolkien, The Fellowship of the Ring

# 1 | Introduction

## 1.1 Cosmology

The theory describing the evolution of the Universe on large-scales is General Relativity which was proposed by Einstein in 1915. One of the greatest achievements and breakthrough of this theory is to connect the shape of space-time with what is embedded in it. In other words, the Universe and its evolution are shaped by its content. The equation

$$G_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}, \quad (1.1)$$

relates the Einstein’s tensor  $G_{\mu\nu}$ , which encodes the geometrical description of space-time, with the stress-energy tensor  $T_{\mu\nu}$ , which describes the nature of the matter or energy embedded in space-time. The matter content characterises the geometry of space-time and the geometry of space-time defines the dynamics of the matter content. In the above equation,  $G$  is Newton’s constant and  $c$  is the speed of light (hereafter, set equal to one).

It is fairly logical to assume that the Earth, the point from which we observe the Universe, is not located in a special position. It is also natural to expect that the observations done by astrophysicists on our planet would be on average equivalent to the ones done from any other planet in any other galaxy in the Universe. This is the Cosmological Principle: from all possible spatial positions the observed statistical properties of the Universe are the same. This idea is supported by cosmological observations, which during the last century have increasingly implied that our Universe is homogeneous and isotropic when observed at large enough scales ( $\geq 100$  Mpc, [Hajian and Souradeep, 2003](#); [Zunckel et al., 2011](#), recently it has been proved by [Saadeh et al., 2016](#) that anisotropic expansion of the Universe is strongly disfavored, with odds of 121 000:1 against). Isotropy means that the Universe appears the same in all directions. Homogeneity means that there is no preferred location: the properties of the Universe and the physics describing them are everywhere the same. If these two properties are assumed to be valid then a solution can be found to Einstein’s equations. The solution is given by the Friedmann-Leimatre-Robertson-Walker (FLRW) metric:

$$ds^2 = c^2 dt^2 - a(t)^2 \left[ \frac{dr^2}{1 - Kr^2} + r^2 (d\theta^2 + \sin^2 \theta d\phi^2) \right], \quad (1.2)$$

where  $ds$  is the infinitesimal line element of space-time,  $t$  is the coordinate time and  $r, \theta, \phi$  are the classical polar coordinates. Two parameters characterise the nature and dynamics of space-time described by the FLRW metric:  $a$  and  $K$ .  $a$  is the scale factor and is a function of time: if it increases (decreases) with time then the Universe is expanding (contracting).  $K$  is the curvature constant and parameterises the spatial geometry. If one defines a *curvature radius*  $r_k$  such that  $K = k/r_k^2$  then the unitless constant  $k$  can take a discrete set of values:  $+1, 0$  and  $-1$  corresponding to a spatial geometry that is closed, flat and open respectively.

In cosmology it is usual to substitute coordinate time  $t$  and the physical distance  $r$  with the *conformal time*  $\eta$  and *conformal radius*  $\chi$  defined by the transformations

$$\eta(t) = \int_0^t \frac{dt'}{a(t')} \text{ and } \chi(r) = r_k S_k^{-1}(r/r_k) \text{ where } S_k(x) := \begin{cases} \sin(x) & \text{if } k = +1 \\ x & \text{if } k = 0 \\ \sinh(x) & \text{if } k = -1. \end{cases} \quad (1.3)$$

The FLRW metric can then be rewritten in the elegant form

$$ds^2 = a^2(\eta) r_k^2 \left[ d\tilde{\eta}^2 - d\tilde{\chi}^2 - S_k^2(\tilde{\chi}) d\Omega \right], \quad (1.4)$$

where  $\tilde{\eta} = \eta/r_k$ ,  $\tilde{\chi} = \chi/r_k$  and  $d\Omega = (d\theta^2 + \sin^2 \theta d\phi^2)$ .

In the 1920s and 1930s the consensus among physicists was for the Universe to be static. Einstein, sharing this belief, introduced a cosmological constant  $\Lambda$  in his equations in order for the Universe to be static. Soon after Hubble discovered that galaxies were moving away from our Milky Way with a velocity proportional to their distance from us ([Hubble and Humason, 1931](#)). He measured the distance of far-away galaxies by exploiting the Cepheid stars property of having luminosity varying with know periods. By observing the luminosity of Cepheid stars in our Galaxy it was possible to infer the distance of other galaxies by measuring their apparent luminosity. When light is emitted by a source that is not at rest with respect to the receiver, it is observed at shifted frequencies according to the Doppler effect. Spectral lines characterising the light emitted by a star are shifted to longer wavelengths ("redshifted") if the star is moving away, and shifted to shorter wavelengths ("blueshifted") if the star is moving towards the observer. The Doppler effect states that the frequency shift is connected to the velocity of the emitting object with respect to the receiver and is given by

$$\frac{\nu_{\text{emi}} - \nu_{\text{obs}}}{\nu_{\text{emi}}} \simeq \frac{v_{\text{obs}}}{c}, \quad (1.5)$$

where  $\nu_{\text{emi}}$  is the frequency of the light at the emission and  $\nu_{\text{obs}}$  is the light frequency measured by the receiver. Hubble realised that the great majority of galaxies was moving away from us with velocity proportional to their distance from the Milky Way,  $v_{\text{obs}} = H d$ . The relation in Equation 1.5 can be rewritten using the wavelengths to define the redshift  $z$  of an emitting object moving away from the observer

$$z = \frac{\lambda_{\text{obs}} - \lambda_{\text{emi}}}{\lambda_{\text{emi}}}, \quad (1.6)$$

where  $\lambda$  is the wavelength. The implied equivalence between recessional velocity  $v_{\text{obs}}$  and redshift is valid only at small cosmological scales. However from Equation 1.2, imposing for the light  $ds^2 = 0$ , at all scales one obtains

$$1 + z = \frac{a_0}{a}, \quad (1.7)$$

where  $a_0$  is the scale factor today, usually set to be equal to unity. Equation 1.7 implies that the redshift can be used to parameterise the expansion of the Universe as well as the scale factor. Therefore, the measured redshift of an astronomical object can tell us how far both in space and in time (in terms of how long ago was the light signal emitted) the object is from us.

### 1.1.1 Friedmann equation and cosmological evolution

In order to discover whether, during its history, the Universe has been expanding, contracting, or remaining static, we need to study the evolution of the scale factor as a function of time. For this reason it is necessary to solve Einstein's equations for the FLRW metric shown in Equation 1.2. Rewriting Einstein's equations including also the cosmological constant gives

$$G_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu} + \Lambda g_{\mu\nu}, \quad (1.8)$$

where we follow the modern interpretation by [Gliner \(1966\)](#) and [Zel'dovich \(1968\)](#) of the cosmological constant being a contribution to the energy-momentum tensor for the quantum vacuum. For the energy-momentum tensor describing the content of the Universe, a common assumption is that it behaves as a perfect fluid with

$$T_{\mu\nu} = -p g_{\mu\nu} + (p + \rho c^2) u_\mu u_\nu, \quad (1.9)$$

where  $p$  and  $\rho$  are respectively the fluid's isotropic pressure and its energy density, while  $u$  is the four velocity which, in a (co-moving) system of reference at rest with the fluid, is  $(1, 0, 0, 0)$ . From Einstein's equations it is possible to derive the Friedmann equations

$$\begin{aligned} H^2 &= \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G\rho}{3} - \frac{kc^2}{a^2} + \frac{\Lambda c^2}{3}, \\ \frac{\ddot{a}}{a} &= -\frac{4\pi G}{3} \left(\rho + \frac{3p}{c^2}\right) + \frac{\Lambda c^2}{3}, \end{aligned} \quad (1.10)$$

where  $H$  is the Hubble parameter and the over-dots indicate coordinate time derivatives of the scale factor. From Equations 1.10, it is evident that for a non-empty Universe ( $\rho \neq 0$ ), the space-time is either contracting or expanding. Depending on the values of  $\rho$ ,  $p$  and  $\Lambda$  on the right hand side, the second equation tell us whether the Universe expansion or contraction is accelerating or decelerating.

Combining the two Friedmann equations or by imposing the energy conservation condition  $\nabla_\mu T^{\mu\nu} = 0$ , where  $\nabla_\mu$  is the covariant derivative, it is possible to obtain the continuity equation

$$\dot{\rho} + 3\frac{\dot{a}}{a} \left(\rho + \frac{p}{c^2}\right) = 0. \quad (1.11)$$

In Equation 1.10 if one temporarily sets the cosmological constant  $\Lambda = 0$ , then  $H^2$  can be thought as the kinetic energy,  $\frac{8\pi G\rho}{3}$  as the potential energy, and  $-\frac{k}{a^2}$  as the conserved total energy. In order to better understand the physical meaning of Equations 1.10, considering the  $\Lambda = 0$  case, it is useful to introduce a characteristic energy density for which the kinetic energy is equal to the potential energy (giving flat space,  $k = 0$ ). This is called the critical energy density

$$\rho_c = \frac{3H^2}{8\pi G}. \quad (1.12)$$

If the total energy density is smaller than  $\rho_c$  then the Universe will be open ( $k = -1$ ) while if it is greater the Universe will be spatially closed ( $k = 1$ ). From this quantity it is also then possible to define density parameters for all the energy components present in the Universe such as

$$\begin{aligned}
 \Omega_{\text{m}} &\equiv \frac{\rho_{\text{m}}}{\rho_{\text{c}}} = \frac{8\pi G\rho_{\text{m}}}{3H^2}, \\
 \Omega_{\text{r}} &\equiv \frac{\rho_{\text{r}}}{\rho_{\text{c}}} = \frac{8\pi G\rho_{\text{r}}}{3H^2}, \\
 \Omega_{\Lambda} &\equiv \frac{\rho_{\Lambda}}{\rho_{\text{c}}} = \frac{\Lambda c^2}{3H^2}, \\
 \Omega_{\text{K}} &\equiv \frac{\rho_{\text{k}}}{\rho_{\text{c}}} = \frac{-kc^2}{H^2},
 \end{aligned} \tag{1.13}$$

where  $\Omega_{\text{m}}$ ,  $\Omega_{\text{r}}$ ,  $\Omega_{\Lambda}$  and  $\Omega_{\text{k}}$  are respectively the density parameters for the total non-relativistic matter, the relativistic radiation, the cosmological constant and curvature contributions. By integrating the continuity equation 1.11 and by labeling the quantities observed at the present day as  $(\rho_0, a_0)$  we find

$$\rho = \rho_0 a^{-3(1+w)} \implies a(t) = a_0 t^{\frac{2}{3(1+w)}} \quad \text{for } w \neq -1, \tag{1.14}$$

where  $w = p/\rho$  is the equation of state parameter characterising each component. The second equation for  $a(t)$  has been obtained by substituting the first into the Friedmann equations. For pressureless matter  $w = 0$ ; for relativistic species  $w = 1/3$ ; for a cosmological constant  $w = -1$ . The first Friedmann equation (Equation 1.10) can be rewritten in terms of the density parameters

$$\frac{H^2}{H_0^2} = \Omega_{\text{r},0} (1+z)^4 + \Omega_{\text{m},0} (1+z)^3 + \Omega_{\text{K}} (1+z)^2 + \Omega_{\Lambda}, \tag{1.15}$$

where Equation 1.7 has also been used to convert between scale factor and redshift. From Equation 1.15 it is clear that different species evolve differently with redshift. For example, the radiation density decreases more quickly, as  $(1+z)^4$ , than the matter density, as  $(1+z)^3$ . For a spatially flat Universe, Equation 1.15 tell us that the cosmological history can be divided in three phases dominated by the energy components:

- **radiation:**  $a(t) \propto t^{\frac{1}{2}}$  with  $H = \frac{1}{2t}$ ;
- **matter:**  $a(t) \propto t^{\frac{2}{3}}$  with  $H = \frac{2}{3t}$ ;
- **cosmological constant:**  $a(t) \propto e^{H_0 t}$  with  $H_0 = \sqrt{\frac{\Lambda}{3}}$ .

All three phases describe an expanding Universe; however by checking the second time derivative of the scale factor, it can be seen that expansion decelerates during the radiation- and matter-dominated epoch, and accelerates during the

$\Lambda$ -dominated epoch. Extrapolating backwards in time, the Friedmann equations indicate that the Universe started expanding from a single point. This idea has been given the well-known name of the “Big Bang”. Nevertheless, this conclusion brings up three important questions:

- why is the Universe observed today so near to spatial flatness? Looking at Equation 1.15 and considering  $z \rightarrow \infty$  if the sum of all the current day energy density components is close to the critical density, then the difference between spatially curved and flat space must have been much smaller at the beginning of the cosmological history. This is the so-called "flatness problem".
- Since light signals have a finite speed, how can it be that different apparently causally disconnected patches of the Universe observed today look so similar (isotropic and homogeneous) and are well thermalised between each others? (horizon paradox)
- How did structure form?

### 1.1.2 Inflation

In order to answer the above questions Guth (1981) proposed a mechanism known as Cosmological Inflation. Several other physicists developed alternative versions (Linde, 1982; Albrecht and Steinhardt, 1982). This theory describes an extremely fast expansion of the early Universe (between  $\sim 10^{-36}$  s and  $\sim 10^{-32}$  s after the initial singularity). The first models theorised that the cause of the fast accelerated expansion was a scalar field called inflaton with equation of state parameter  $w < -1/3$ . This requirement can be understood by computing the deceleration parameter from the Friedmann equations 1.10 for a generic species with energy density and pressure related by  $p/\rho = w$

$$q = -\frac{a\ddot{a}}{\dot{a}^2} = \frac{\rho + \frac{3p}{c^2}}{\rho}. \quad (1.16)$$

For the flatness problem, manipulating the Friedmann equations it follows that:

$$\left(\Omega^{-1} - 1\right) \rho a^2 = -\frac{3kc^2}{8\pi G}. \quad (1.17)$$

The right hand side of the above equation is constant with time, while the left hand side is not. In particular the term  $\rho a^2$  increases during time, since the scale factor grows exponentially during inflation. Therefore, the initial difference  $|\left(\Omega^{-1} - 1\right)|$  is irrelevant since by the end of a large enough period of inflation

it can be made arbitrarily close to zero. This means that an extremely brief inflation can flatten the spatial geometry of the very early Universe to such a degree that, even with the departure from flatness described by Equation 1.15, it is not unexpected nowadays to observe a spatial geometry very close to be flat.

The observation that apparently causally-disconnected regions of the Universe present extremely similar properties can also be explained through the inflationary mechanism. During the very short time of accelerated expansion, a small region in equilibrium with itself has expanded into several causally disconnected regions which preserved the same conditions of when they were in equilibrium. Different regions of the Universe in particular at large-scales, moving far away from each other at a speed greater than the speed of light ("causally disconnected"), can look statistically homogeneous and isotropic.

Finally, since inflation expands a tiny region by several orders of magnitude, quantum fluctuations can be stretched to cosmological scales. After inflation is over and the scalar field has decayed into matter and radiation in a process called "reheating" (Shtanov et al., 1995; Linde, 1996; Kofman, 1996), these fluctuations grow into under/over-densities of the matter and radiation fields. The imprinted oscillations in the radiation and, in particular, the matter field act as seeds for the formation of structure due to gravitational collapse.

Following inflation, the next fundamental step in cosmic history was baryogenesis. During this period, the first standard matter (as we know it) formed. One particular aspect that several models have been proposed in the literature to explain is the asymmetry between matter and antimatter. The first models in the literature propose to explain the decay of super massive particles (Weinberg, 1979; Toussaint and Wilczek, 1979), and more recent models are based on supersymmetry (Affleck and Dine, 1985).

Light element isotopes like D,  $^3\text{He}$ ,  $^4\text{He}$  and  $^7\text{Li}$  were formed during the Big-Bang Nucleosynthesis which took place when the average temperature of the Universe was around 1 MeV.

### 1.1.3 Cosmic Microwave Background

Due to the cooling expansion of the Universe, at redshift  $z \simeq 1100$  ( $\sim 379000$  years after the Big Bang) the average temperature dropped below 3000 K. The average kinetic energy of electrons became low enough for the hydrogen nuclei to capture them and form neutral hydrogen atoms. As a consequence of this process, labelled as "recombination", the photons were then free to propagate. The cosmic microwave background radiation (CMB) is made of the photons from this epoch also called "last scattering" (between photons and electrons).

This blackbody radiation was first discovered by [Wilson and Penzias \(1965\)](#) and afterwards investigated by dedicated surveys like the Cosmic Background Explorer *COBE* ([Smoot et al., 1992](#)), the Wilkinson Microwave Anisotropy Probe *WMAP* ([Bennett et al., 2003b](#)) and *Planck* ([Planck Collaboration et al., 2014](#)). The CMB radiation is extremely well described by a black body with average temperature  $T = 2.7255$  K, with fluctuations that are five orders of magnitude smaller. The surveys listed above had as their primary goal the mapping of these temperature fluctuations. The resulting 2D maps can be described using spherical harmonics, expanding the temperature field as a function of two angular variables  $\theta$  and  $\phi$ ,

$$T(\theta, \phi) = \sum_{\ell m} a_{\ell m} Y_{\ell m}(\theta, \phi), \quad (1.18)$$

where  $\ell$  and  $m$  are the orders of the multipole expansion and  $a_{\ell m}$  are the spherical harmonic coefficients. The multipole  $\ell$  is related to features in the sky of size  $\theta \approx 180^\circ/\ell$ . Since the temperature fluctuations field is to a very good approximation independent of direction (statistically isotropic), most of the information can be described using simply the variance of angular separation, parameterised by the  $C_\ell = \langle |a_{\ell m}|^2 \rangle$  coefficients which are estimated by

$$\hat{C}_\ell = \frac{1}{2\ell + 1} \sum_{m=-\ell}^{\ell} |a_{\ell m}|^2. \quad (1.19)$$

These  $C_\ell$  values are shown as a function of the multipole  $\ell$  in [Figure 1.1](#) and form the angular power spectrum of the CMB temperature fluctuations. It is possible to divide the  $\ell$ -range in the regions listed below, which are dominated by different physical processes modifying the primordial temperature perturbation field:

- **Integrated Sachs-Wolfe effect**,  $\ell \leq 10$  : low multipole anisotropies correspond to the largest angular separations in the sky. These are affected by the variation in time of the gravitational potential in the regions of space crossed by the CMB photons. At these low multipoles and at low redshifts the cosmological constant energy component dominates and photons find themselves climbing out from a lower potential step than the one they fell into, due to the repulsive action of Dark Energy. This changes the spatial energy distribution of CMB photons, i.e. their anisotropies.
- **Sachs-Wolfe Plateau**,  $10 \leq \ell \leq 100$  : the angular separation corresponding to the horizon dimension at the time of the last scattering between electrons and photons is to  $\ell \simeq 100$ . If the curvature perturbations

had a spectrum close to be featureless at the time of last scattering, then it would imply a nearly constant value for the quantity shown in Figure 1.1, namely  $\ell(\ell + 1)C_\ell$ . The effect that characterises this  $\ell$ -range is called Sachs-Wolfe effect, the combination of intrinsic temperature fluctuations and gravitational redshift.

- **Acoustic Peaks**,  $100 \leq \ell \leq 1000$  : for perturbations inside the horizon at the time of last scattering, oscillations of the baryon-photon plasma before recombination can be observed. Before the formation of neutral atoms, photons were tightly coupled with electrons via Compton scattering. The photons followed the baryon oscillations which were formed by the balance between the gravitational collapse of matter and the radiation pressure of the tightly-coupled plasma. Later, when the efficiency of the Compton scattering became low enough due to cooling from the Universe's expansion, the mean-free path of the photons became effectively infinite since they decoupled from the electrons. Therefore, where at the last scattering surface there was an over-density of the photon-baryon plasma, after the decoupling more energetic photons were freed to propagate from that region than for example from another region where there was an under-density. These patterns in the photons reflect the distribution of matter at the time of the last scattering and the amplitude of the soundwaves propagating in the photon-baryon plasma. The shape and patterns of these oscillations are strongly dependent on the cosmological model and its parameters. Theoretically, the existence of these peaks was predicted since 1970 (Peebles and Yu, 1970; Sunyaev and Zeldovich, 1970) and their detection and study began in the early 1990s (Smoot et al., 1992; Scott et al., 1995).
- **Silk Damping**,  $\ell \geq 1000$  : since recombination was not an instantaneous process, the surface of last scattering had a finite thickness. The perturbations corresponding to angular separations smaller than this thickness were smoothed out by the residual recombination interaction inside the shell. In other words, during the interval of time in which recombination took place, the photons were still partially dragged by the baryons during their oscillations. This is similar to what happens to a sand castle's tower after a wave passes back and forth over it).

#### 1.1.4 Dark matter

A first indication of the need to introduce an exotic component of matter comes from the expansion history of the Universe and in particular from the time at

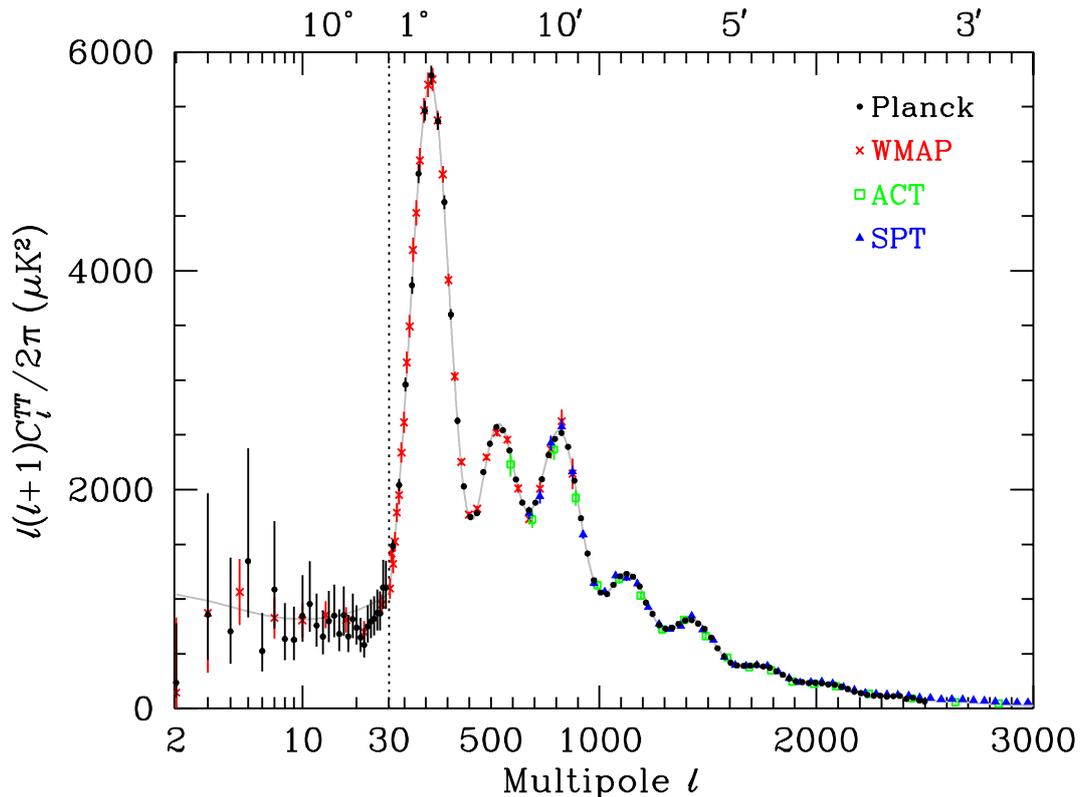


Figure 1.1: Plot of the angular power spectrum presented in the Cosmic Microwave Background section of the Particle Physics Review 2017 (Patrignani et al., 2016). The data shown are from *Planck*, *WMAP*, Atacama Cosmology Telescope (ACT Swetz et al., 2011) and the South Pole Telescope (SPT Schaffer et al., 2011) surveys. In order to show the Sachs-Wolfe plateau, the  $x$ -axis scale is logarithmic for the lowest multipoles, and it is linear for higher multipoles. The acoustic peaks and damping region are well above the the errorbars limit and therefore well detected. The black line show the best fit given by a  $\Lambda$ CDM cosmology.

which the radiation-matter equivalence occurred. If only baryonic matter were present in the Universe, then the equivalence time would have occurred much later than what is necessary in order to allow the gravitational instability to create the large-scale structures that we observe now. The inclusion of a dark matter component increases the value of the redshift at which matter started dominating, allowing enough time for structure formation.

In the last decades it was possible to determine through spectroscopic and photometric surveys the total amount of baryonic matter in the Universe in the form of either collapsed objects like stars or gas clouds. The density parameter of baryonic matter  $\Omega_b$  has been estimated to be between  $\Omega_b \simeq 0.04$  and  $0.05$  (Fukugita et al., 1998; Percival et al., 2010) with the most precise and recent measurement being  $\Omega_b h^2 = 0.02242 \pm 0.00014$  (Planck Collaboration et al., 2018)

for  $\Omega_b$  multiplied by  $h^2$  where  $H = 100h$  (km/s)Mpc<sup>-1</sup>. Since the Universe has been observed to be very close to spatial flatness, as tested recently by combining of CMB and BAO data to constrain the curvature density parameter to be  $\Omega_K = 0.0007 \pm 0.0019$  (Planck Collaboration et al., 2018), then only approximately 4% of the required energy density needed to reach the critical density is given by baryons. Therefore the most straightforward hypothesis is to assume that there is some kind of exotic matter which does not interact electromagnetically.

Moreover, the largest objects in the Universe (clusters of galaxies) require dark matter to explain their dynamics, as understood first by Zwicky (1933).

Based on the dynamics of astrophysical objects, probably the strongest indirect evidence in support of the existence of this "dark" form of matter came from observations of star rotation curves inside galaxies (Rubin and Ford, 1970). Stars have been observed to move around the center of the galaxy much faster than expected from gravitational potentials only taking into account visible matter. In order to explain such higher velocities, especially at the largest radii, the concept of a dark matter "halo" was introduced, where a cloud of dark matter embeds the galaxy. One example supporting the halo-model is the Bullet Cluster (Clowe et al., 2006). Including the haloes of dark matter in the energy balance, the total matter density parameter reached the value of  $\Omega_m \simeq 0.3$  (Burkert and Silk, 1997). The most recent constraint on  $\Omega_m$  was also given by Planck Collaboration et al. (2018):  $\Omega_m = 0.3111 \pm 0.0056$ .

Alternative theories have been proposed that do not require the introduction of another matter component in order to account also for the star rotation curves in a galaxy. Among these the most known is the Modified Newtonian Dynamics (MOND Milgrom, 2015). MOND however at the moment still has a relativistic formulation (TeVeS, Bekenstein, 2004) but seems to have problem in independently recovering the formation of large-scale structure as are observed nowadays (Bernard and Blanchet, 2015).

General Relativity tells us that the energy/matter content shapes the space time in which it is embedded. A dark type of matter would have this effect. When light passes through an area where a halo of dark matter is present, its path deviates from a straight line, it is bent. The distortions of the light path propagating in the Universe due to the matter encountered along its path is known as "gravitational lensing". Even if in practice this effect is very small (weak-lensing regime) for each lens (beside more rare cases of "strong" gravitational lensing), it is on average statistically detectable by photometric and spectroscopic galaxy surveys (Benjamin et al., 2007) and can be used to constrain the total amount of matter in the Universe (Falco et al., 1998). Gravitational lensing indeed does not distinguish between visible and dark matter since both have the same gravitational

effect on photon trajectories. Recent results from weak-lensing analysis have been given in terms of the parameter  $S_8 = \sigma_8 \sqrt{\Omega_m/0.3}$  where  $\sigma_8$  describes the matter density fluctuations amplitude averaged over a sphere of 8 Mpc/ $h$ . These are  $S_8 = 0.745 \pm 0.039$  (Hildebrandt et al., 2017) and  $S_8 = 0.67 \pm 0.03$  (Alsing et al., 2017).

Of course the CMB described in the previous section also allows us to determine constraints on the total amount of matter present at the epoch of the last scattering. Indeed the shape and amplitude of fluctuations of the photon-baryon fluid imprinted onto the photon distributions after decoupling strongly depend on the total quantity of matter.

### 1.1.5 Dark energy

The Friedmann Equations 1.10 predict that a Universe containing pressureless matter would undergo a decelerated expansion. In the second half of the 90s scientists wanted to quantify the deceleration of the expansion of the Universe by using type Ia supernovae measurements. The fundamental quantity used in these analyses is the luminosity distance which compares the apparent luminosity (measured flux  $S$  by an observer) with its known intrinsic one  $L$ . The flux measured by an observer at a distance  $r$  from the source is given by the inverse square relation:

$$S = \frac{L}{4\pi r^2}. \quad (1.20)$$

However this relation needs to be corrected in the case of an expanding Universe by a factor of  $a$  because the radiation travelling to the observer gets redshifted. Moreover another factor of  $a$  is necessary to take into account the time dilatation due to the expansion. Finally, generalizing to non-spatially flat cases by substituting  $r$  with the normalised conformal radius defined in Eq. 1.4, the luminosity distance  $d_L$  is defined as:

$$d_L = (1 + z) S_k [\tilde{\chi}(r)]. \quad (1.21)$$

Considering a null geodesic the normalised conformal radius can be computed as:

$$\tilde{\chi}(z) = \int_0^z \frac{dz'}{H(z')}. \quad (1.22)$$

Therefore by measuring the luminosity distance and comparing it with the prediction made by Eq. 1.22 using the Friedmann equation, for example expressed as in Eq. 1.15, the relation between  $\Omega_m$  and  $\Omega_\Lambda$  can be constrained.

Type Ia supernovae can be used to measure astronomical distances since they all have approximately the same luminosity when exploding. By measuring their apparent brightness and redshift, it is possible to use them as "standard candles". Independently, two groups found out that the supernovae brightness was lower than expected for their redshifts and that therefore they were further away than as predicted by a decelerating model of the expansion. In other words they found out that the Universe's expansion was accelerating (Riess et al., 1998; Perlmutter et al., 1999). This contribution to the energetic balance of the Universe was soon called "Dark Energy" where the adjective reflects the fact that still up to now it is an unknown form of energy.

One of the first hypotheses was to link this unknown form of energy to the cosmological constant  $\Lambda$ , the vacuum energy density that Einstein introduced in 1917 in order to produce a static model of the Universe. However, it is still an open problem to reconcile the measured value of  $\Lambda$  with what is theoretically predicted from quantum field theory ( $\sim 120$  orders of magnitude of difference).

Figure 1.2 shows representative plots from Riess' fundamental work. In particular it can be seen that introducing a cosmological constant-like component better fits the data in particular between redshifts  $0.3 \leq z \leq 1$ . The surprising aspect of the discovery was also given by the fact that this new component needed to be  $\sim 70\%$  of the total energy balance of the Universe. CMB later confirmed this subdivision of the total energy which in order to match the observations needs to be close to the critical density (Komatsu et al., 2009). Previously to the supernovae luminosity measurements, galaxy cluster evolution studies, for example on the APM survey data-set (Maddox et al., 1990), already considered the possibility of introducing a cosmological constant in order to better fit the data (Baugh and Efstathiou, 1993; Gaztanaga and Frieman, 1994).

From the deceleration parameter in Equation 1.16, it follows that in order to have an accelerated expansion the equation of state must satisfy  $p < -\frac{1}{3}\rho$ . In the particular case of the cosmological constant the equation of state parameter is  $w = p/\rho = -1$  and from the continuity Equation 1.11 it is clear that sooner or later it will become the leading component of the energetic balance of the Universe since  $\dot{\rho}_\Lambda = 0$ . However it is possible to imagine different alternatives for this substance characterised by a negative pressure with equation of state  $w \neq -1$  and even not constant in time. Refer to the several reviews available in the literature for a description of these models (Joyce et al., 2016; Bamba et al., 2012; Frieman et al., 2008b; Copeland et al., 2006; Peebles and Ratra, 2003). What is important in observational cosmology is to be able to translate all these models into a common parametrisation which can be used to discriminate among them. However, constraining a general evolution of the equation of state

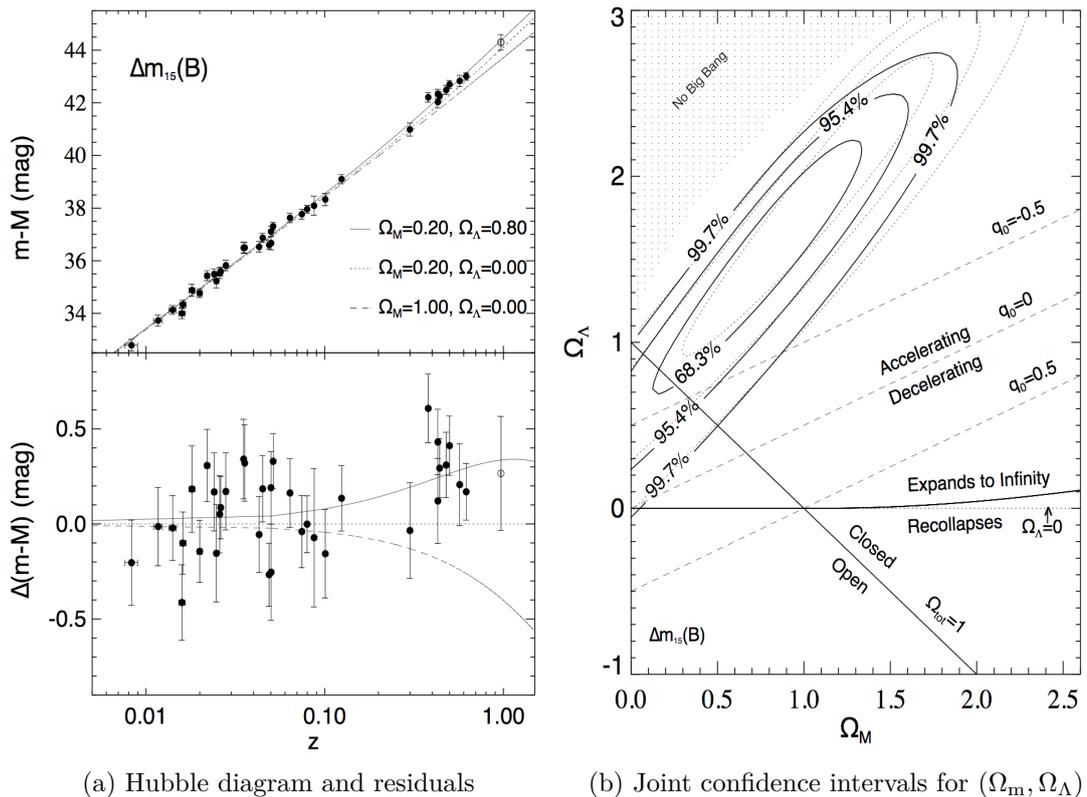


Figure 1.2: Both left and right panels show plots contained in Riess' seminal paper (Riess et al., 1998). **a)** In the upper half is reported the Hubble diagram with the magnitude (function of the observed brightness) as a function of redshift for all the supernovae considered in the analysis. The difference between the three different models is more clear in the lower half where the residuals between the data and the models are plotted with as reference the model with parameters ( $\Omega_m = 0.2, \Omega_\Lambda = 0.0$ ) which was the favoured one before the Dark Energy discovery.

**b)** The 2D contours for the parameters ( $\Omega_m, \Omega_\Lambda$ ) derived from the supernovae analysis are shown. It can be seen that both  $1\sigma$  and  $2\sigma$  contours are well inside the accelerated expansion area denoted by the decelerating parameter being  $q_0 < 0$ .

parameter as a function of the redshift  $w(z)$  is extremely difficult, in particular because of the strong degeneracy with the Hubble parameter  $H(z)$ . For this reason the following linear approximation is usually implemented:

$$w(a) = w_0 + w_a(1 - a) = w_p + w_a(a_p - a), \quad (1.23)$$

where  $a$  is the scale factor which can be easily converted into the redshift using Equation 1.6,  $w_0 = w(z = 0)$ ,  $w_p$  is the value of  $w$  at a pivot redshift  $z_p$  chosen such that  $w$  is well constrained at the epoch for the considered survey.

### 1.1.6 Cosmological probes

Below is a brief description of the main cosmological probes used to constrain the parameters of a  $\Lambda$ CDM cosmology (and its extensions):

- **Direct measurements of  $H$**  : the study started by Hubble of the relation between the recession velocity of galaxies and their distance from us has become more and more precise during the last decades. In particular the Hubble Space Telescope Key Project (Freedman et al., 2001) used the known empirical period-luminosity relation of Cepheid variable stars to obtain a value of  $H_0 = 73.2 \pm 1.7 \text{ km s}^{-1} \text{ Mpc}^{-1}$ . However this measurement is influenced by systematics like the high abundance of heavy elements in the Cepheids stars and the uncertainties in the distance of the Large Magellanic Cloud galaxy, which is used to normalise the Cepheids stars distances.
- **Supernovae** : the most widely accepted theoretical model for supernovae are thermonuclear explosions occurring at the end of white dwarfs's lifetime. These explosions are generated by the same elements, so the peak-luminosity shows little variance. Therefore supernovae can be used as "standard candles" to measure cosmological distances via the relation between apparent brightness and measured luminosity. This allows us to constrain cosmological models by tracking the expansion of the Universe and hopefully future experiments will reach such a precision to be able to constrain the Dark Energy equation of state parameter's evolution,  $w(z)$ . Key projects include: High-Z Supernova Search Team (Riess et al., 1998), Supernova Cosmology Project (Goldhaber, 2009), Hubble Space Telescope (Freedman et al., 2001), Supernova Legacy Survey (Astier et al., 2006) and Sloan Digital Sky Survey (Frieman et al., 2008a)
- **CMB** : as mentioned in the previous Section 1.1.3, the cosmic microwave background contains a picture showing us the primary anisotropies present at the time of recombination. The photons radiated from the last scattering surface also carry information regarding the spacetime through which they propagated. For example they are affected by the crossed gravitational potential wells (Integrated Sachs-Wolfe effect) and their trajectories are deviated by the presence of mass clumps (lensing). These anisotropies create particular and well defined features, for example the acoustic peaks in the angular power spectrum ( $C_\ell$ ). These depend on the particular cosmological model and allow us to put tighter constraints on the cosmological parameters. Among the main surveys dedicated to the CMB there are *COBE*

(Wright et al., 1992), *WMAP* (Bennett et al., 2003a) and *Planck* (Lamarre et al., 2003).

Table 1.1: Partial list of recent and future cosmological surveys. Table extended from the one presented in Dark Energy section of the Particle Physics Review 2017-update (Patrignani et al., 2016), based on the review by Weinberg et al. (2013). Abbreviations in the "Data" column refer to optical (Opt) or near-infrared (NIR) imaging (I) or spectroscopy (S). For spectroscopic experiments, the " $z$ -range" column lists the primary redshift range for galaxies (gal.s), quasars (QSOs), or the Lyman- $\alpha$  forest (Ly $\alpha$ F). Abbreviations in the "Techniques" column are weak lensing (WL), gravitational lensing, clusters (CL), supernovae (SN), baryon acoustic oscillations (BAO), and redshift-space distortions (RSD), cosmic microwave (CMB).

Survey	Years	Area [deg <sup>2</sup> ]	Data	$z$ -range	Techniques
BOSS	2008-14	10000	Opt-S	0.3-0.7 (gal.s) 2.0-3.5 (Ly $\alpha$ F)	BAO/RSD
DES	2013-18	5000	Opt-I		WL/CL SN/BAO
eBOSS	2014-20	7500	Opt-S	0.6-2.0 (gal.s/QSO) 2.0-3.5 (Ly $\alpha$ F)	BAO/RSD
DESI	2019-24	14000	Opt-S	0.0 - 1.7 (gal.s) 2.0-3.5 (Ly $\alpha$ F)	BAO/RSD
LSST	2020-30	20000	Opt-I		WL/CL SN/BAO
<i>Euclid</i>	2020-26	15000	Opt-I NIR-S	0.7-2.2 (gal.s)	WL/CL BAO/RSD
WFIRST	2024-30	2200	NIR-I NIR-S	1.0-3.0 (gal.s)	WL/CL/SN BAO/RSD
<i>Planck</i>	2009-13	full-sky			CMB/ISW/GL
HSC	2018-	1400	Opt-I	0.0-7.0	WL/GL/SN
KiDS	2011-19	1500	Opt-I	0.1 - 1.2	WL/GL

- **Baryonic Acoustic Oscillations** : the oscillations of the photon-baryon fluid imprinted on the photon distribution at the last scattering surface had a characteristic scale ( $\sim 150$  Mpc). In the following evolution of these overdensities the corresponding comoving value of the oscillations particular length was preserved, meaning that there was a higher probability than

average of finding two overdensities of baryonic matter separated by this scale. Due to the cosmic expansion, this physical distance has also evolved. By measuring it at different redshifts and by relating it to the initial value deduced from the CMB it is possible to build a "standard ruler". This can impose tighter constraints on the expansion of the Universe. Surveys performing these studies include the 2-degree Field Galaxy Redshift Survey (2dF [Cole et al., 2005](#)), WiggleZ ([Parkinson et al., 2012](#)), Sloan Digital Sky Survey (SDSS [Eisenstein et al., 2005](#)) and in particular its extension, the Baryon Oscillation Spectroscopic Survey (BOSS [Anderson, 2014](#)).

- **redshift-space distortions and Alcock-Paczynski effect**: the matter density perturbation growth rate  $f = d \ln \delta / d \ln a \simeq \Omega_m^\gamma(z)$  parameterises the distortions in the galaxy distributions due to the measurement of their redshift.  $f$  allows us to place tight constraints on alternative gravity theories (for General Relativity and  $\Lambda$ CDM models,  $\gamma \simeq 0.55$  [Kaiser, 1987](#); [Guzzo et al., 2008](#); [Nusser and Davis, 2011](#)). For objects in the sky expanding with the cosmological flow, there are two directly observable measures of its size: the angular width (orthogonal to the line of sight) and the radial extent in redshift (parallel to the line of sight). If there is no preferred alignment then on average the ratio between the two inferred sizes should be equal to unity. One can then check the conversion factors from redshift interval and angular width to physical distances. This procedure is called the Alcock-Paczynski test ([Alcock and Paczynski, 1979](#)). Both WiggleZ and BOSS quantified both RSD and AP effects and the degeneracy existing between them ([Blake et al., 2012](#); [Gil-Marín, 2017](#)).
- **Galaxy clustering** : the matter distribution in the Universe at late-times can impose additional constraints on the cosmological models and in particular to the consensus one,  $\Lambda$ CDM. By measuring the position of galaxies in the sky via spectroscopic (3D mapping) and photometric (tomographic bins projected 2D mapping) imaging, it is possible to use several statistics like the galaxy power spectrum (which will be introduced later on) to infer cosmological parameter values from the data. However the problem of observing only the luminous form of matter is that the light-emitting objects like galaxies are biased tracers of the total matter density distribution. This biased relation between baryonic matter and dark matter needs to be taken into account in the modelling ([Kaiser, 1984a](#)). More complex models like scale dependent or stochastic bias which (taking into account the partially random nature of the baryons and dark matter fields) may be also needed to be taken into consideration ([Dekel and Lahav, 1999a](#)).

Another systematic present in this type of analysis is connected to the fact that we measure the redshifts of galaxies and not their actual distance. Galaxies have a proper motion due to the local gravitational field which gives them a peculiar velocity. These peculiar velocities affect the galaxies' redshift via the Doppler mechanism, introducing uncertainties. For example at scales smaller than  $\sim 10\text{Mpc}$ , galaxies tend to fall towards the center of their cluster which their distribution squeezes along the line of sight. This is because the average peculiar velocity of closer galaxies will be towards the center, increasing the redshift of the galaxies via Doppler effect. On the contrary further galaxies will have average peculiar velocities directed towards the cluster core away from the observer, decreasing the redshift of the galaxies. Finally at Megaparsec scales, galaxy velocity dispersions create an elongation along the line of sight usually referred to as "Finger of God" effect, since long thin filaments of structure in redshift-space point directly back at observer. This arises since the Doppler effect from the peculiar velocities of galaxies is only observed radially. Key projects include BOSS (White, 2011), DES (Dark Energy Survey Collaboration et al., 2016) and in the future DESI (Levi et al., 2013) and EUCLID (Laureijs et al., 2011).

- **Gravitational lensing** : as was mentioned earlier in Section 1.1.4, GR predicts the bending of light trajectories in proximity of a strong gravitational field. The light that the telescopes observe coming from far away galaxies has followed a non-linear path. If a background population of galaxies is considered, their image observed by us will have been distorted by the interposing mass distribution. The magnitude of this distortion is the discriminant between "strong" and "weak" gravitational lensing. Strong lensing happens when the image of a background galaxy is so distorted that it appears as a full or segments of an arc, or even multiple images centered around a mass clump. Smaller perturbations of the galaxy images are referred to as weak lensing phenomena. Even if they are very small effects, for a large enough sample they are statistically significant and have the advantage with respect to galaxy clustering of tracing the distribution of the total mass, both luminous and dark. As for the galaxy clustering case, the same statistics (projected on a 2D map orthogonal to the line of sight) can be used to constrain cosmological parameters (reviews Bartelmann and Schneider, 2001; Refregier, 2003; Kilbinger, 2015). Current surveys include the Kilo-Degree Survey (KiDS de Jong et al., 2013b), the Dark Energy

Survey (DES [The Dark Energy Survey Collaboration, 2005](#)) and the Hyper Suprime-Cam Subaru Strategic Survey ([Miyazaki, 2012](#)).

- **Lyman- $\alpha$  forest** : as the name suggests this probe is based on the Lyman- $\alpha$  electron transition of neutral hydrogen atoms. The light emitted by distant quasars and galaxies ([Lynds, 1971](#)), during its travel towards an observer, crosses gas clouds at different redshifts which creates in the light spectra different absorption lines. Therefore the Lyman- $\alpha$  forest can be used to investigate for example the intergalactic medium and the presence of neutral hydrogen at different epochs. Cosmological parameter values can also be constrained using the Lyman- $\alpha$  forest ([Weinberg et al., 1998](#)), for example to measure BAO at high redshift ([Slosar et al., 2009](#)), as well as to test extensions to the standard model (neutrinos [Rossi, 2014](#), warm dark matter [Baur et al., 2016](#), etc.).

In Table [1.1](#) are reported the main present and future cosmological surveys.

### 1.1.7 Best-fit cosmology

In Table [1.2](#) are reported the best-fit values for the principal cosmological parameters for a  $\Lambda$ CDM cosmology obtained by combining *Planck* data with other cosmological surveys. The values obtained by *Planck* alone are also shown in order to highlight the huge importance and prominence of CMB data for cosmology. The parameter shown are:

- $\Omega_b h^2$  : baryonic matter density parameter, multiplied by  $h^2$  where  $H = 100h$  (km/s)Mpc $^{-1}$ ;
- $\Omega_c h^2$  : cold dark matter density parameter, multiplied by  $h^2$  where  $H = 100h$  (km/s)Mpc $^{-1}$ ;
- $\tau$  : Thomson scattering optical depth due to reionization;
- $\ln(10^{10} A_s)$  : scalar amplitude of the primordial perturbation power spectrum;
- $n_s$  : scalar spectral index;
- $H_0$  [km Mpc $^{-1}$  s $^{-1}$ ] : Hubble constant;
- $\Omega_\Lambda$  : dark energy density parameter;
- $\Omega_m$  : total matter density parameter;

- $\sigma_8$  : matter density fluctuations amplitude averaged over a sphere of 8 Mpc/ $h$ ;
- $z_{\text{re}}$  : reionization redshift;
- $t_0$  [Gyears] : age of the Universe;
- $z_*$  : redshift at photon decoupling;
- $r_*$  [Mpc] : comoving size of the sound horizon at the decoupling;
- $100\theta_*$  [rad]: angular scale of sound horizon at last-scattering  $\times 100$ .

Table 1.2: best-fit  $\Lambda$ CDM parameters from [Planck Collaboration et al. \(2014\)](#) combined with other surveys. For *Planck* alone, TT represents the temperature power spectrum, TE is the temperature-polarization cross spectrum, and EE is the polarisation power spectrum, lowP are the polarisation data for the large-scales (small  $\ell$ ) likelihood and "lensing" is the CMB lensing reconstruction. The additional data sets included in the joint analysis are the baryonic acoustic oscillations (BAO), the Joint Light-curve Analysis of supernovae (JLA) and the Hubble constant ([Efstathiou, 2014](#)).

Parameter	TT,TE,EE + lowP + lensing	TT,TE,EE + lowP + lensing + BAO+JLA+ $H_0$
$\Omega_b h^2$	$0.02226 \pm 0.00016$	$0.02230 \pm 0.00014$
$\Omega_c h^2$	$0.1193 \pm 0.0014$	$0.1188 \pm 0.0010$
$\tau$	$0.063 \pm 0.014$	$0.066 \pm 0.012$
$\ln(10^{10} A_s)$	$3.059 \pm 0.025$	$3.064 \pm 0.023$
$n_s$	$0.9653 \pm 0.0048$	$0.9667 \pm 0.0040$
$H_0$	$67.51 \pm 0.64$	$67.74 \pm 0.46$
$\Omega_\Lambda$	$0.6879 \pm 0.0087$	$0.6911 \pm 0.0062$
$\Omega_m$	$0.3121 \pm 0.0087$	$0.3089 \pm 0.0062$
$\sigma_8$	$0.8150 \pm 0.0087$	$0.8159 \pm 0.0086$
$z_{\text{re}}$	$8.5^{+1.4}_{-1.2}$	$8.8^{+1.2}_{-1.1}$
$t_0$	$13.807 \pm 0.026$	$13.799 \pm 0.021$
$z_*$	$1090.00 \pm 0.29$	$1089.90 \pm 0.23$
$r_*$	$144.71 \pm 0.31$	$144.81 \pm 0.24$
$100\theta_*$	$1.04106 \pm 0.00031$	$1.04112 \pm 0.00029$

## 1.2 Perturbation theory

In this section are summarised the main concepts that will be the pillars of the work presented in this thesis. For a more complete description see the clear analytic presentation in the "Physical Foundations of Cosmology" book by [Mukhanov \(2005\)](#). For the first systematic approach and application see the seminal "Large-Scale Structure of the Universe" book by [Peebles \(1980\)](#). For a recent review of statistical tools derived from perturbation theory applied to real-world surveys see the review by [Bernardeau et al. \(2002\)](#).

At recombination the CMB tells us that the inhomogeneities present in the photon-baryon fluid and dark matter distribution could be very well described by a homogeneous and isotropic Gaussian field. Nevertheless, at late-times spectroscopic and photometric surveys tell us that the Universe is rich of extremely non-linear structures, such as galaxies, galaxy clusters, filaments, and voids. How did the Gaussian fluctuations at the time of decoupling develop into the high-density objects we observe today? Gravitational collapse is the most simple and straightforward process. Gravity tends to grow the above-average density regions by attracting additional matter from their neighbourhood. At the same time it empties those patches of space which had an original unde-average density. In order to describe this process it is useful to study the evolution of the density perturbation field defined

$$\delta(\mathbf{x}) = \frac{\rho(\mathbf{x}) - \bar{\rho}}{\bar{\rho}}, \quad (1.24)$$

where  $\rho(\mathbf{x})$  is the local energy density and  $\bar{\rho}$  is the average energy density. For the purpose of this thesis work, it is enough to consider the evolution of the density perturbations of non-relativistic matter at sub-horizon scales. Therefore the Newtonian description of gravitational collapse is a very good approximation for what is observed by current cosmological surveys. The matter contained in the Universe can then be thought to follow the behaviour of a perfect fluid and hence be described by the variables:

- $\rho(\mathbf{x}, t)$  : energy density;
- $S(\mathbf{x}, t)$  : entropy per unit of mass;
- $\mathbf{V}(\mathbf{x}, t)$  : spatial 3-velocity.

These quantities are connected between each other by the hydrodynamical equations:

$$\begin{aligned}
 \frac{\partial \rho}{\partial t} + \nabla(\rho \mathbf{V}) &= 0, \\
 \frac{\partial \mathbf{V}}{\partial t} + (\mathbf{V} \cdot \nabla) \mathbf{V} + \frac{\nabla p}{\rho} + \nabla \phi &= 0, \\
 \frac{dS(\mathbf{x}(t), t)}{dt} = \frac{\partial S}{\partial t} + (\mathbf{V} \cdot \nabla) S &= 0, \\
 \nabla^2 \phi - 4\pi G \rho &= 0,
 \end{aligned} \tag{1.25}$$

where  $p$  is the fluid pressure and  $\phi$  is the local gravitational potential. The first equation is the continuity equation which imposes the conservation of energy/mass inside an infinitesimal element of volume. In the second line, it is reported the Euler equation for a fluid which describes the dynamics of a fluid under the force of gravity and the presence of an external pressure. The third equation states the conservation of entropy for a matter element. The last equation, named the Poisson equation, describes the connection between gravitational potential and the matter generating it. If the pressure is considered to be a function of the energy density and entropy through the equation of state for a fluid  $p = f(\rho, S)$ , then the set of Equations 1.25 form a closed system for the variables  $\rho, \mathbf{V}, \phi, S$ .

### 1.2.1 Static Universe

It is useful to consider the case, as was first introduced by Einstein, of a Universe in which the gravitational attraction created by a matter distribution with constant homogeneous and isotropic energy density is perfectly counterbalanced by a cosmological constant. This in order for the Universe to be static. Perturbing the variables as for the energy density with

$$\rho(\mathbf{r}, t) = \rho_0 + \delta\rho(\mathbf{r}, t), \tag{1.26}$$

the pressure perturbation can be expressed as a function of the energy density and entropy perturbations

$$\delta p = c_s^2 \delta\rho + \sigma \delta S, \tag{1.27}$$

where  $c_s^2 = \partial p / \partial \rho$  defines the speed of sound (much smaller than  $c$ , the speed of light, for  $p \ll \rho$ ) and  $\sigma = \partial p / \partial S$ . Substituting all the perturbations into the set of Equations 1.25 and keeping only the terms linear in one obtains the closed equation

$$\frac{\partial^2 \rho}{\partial t^2} - c_s^2 \Delta \delta\rho - 4\pi G \rho_0 \delta\rho = \sigma \Delta \delta S, \tag{1.28}$$

where  $\Delta$  is the Laplacian operator ( $\nabla \cdot \nabla = \Delta$ ). In a static Universe, the entropy perturbations act as sources of the energy density perturbations. If one considers adiabatic perturbations ( $\delta S = 0$ ), taking the Fourier transform of Equation 1.28 gives

$$\delta \ddot{\rho}_{\mathbf{k}} + \left(k^2 c_s^2 - 4\pi G \rho_0\right) \delta \rho_{\mathbf{k}} = 0, \quad (1.29)$$

where dots indicate coordinate time derivatives. The precise form of the solution

$$\delta \rho_{\mathbf{k}} \propto e^{\pm i\omega(k)t}, \quad (1.30)$$

depends on the sign of the quantity under the square root of the expression defining  $\omega(t)$

$$\omega(k) = \sqrt{k^2 c_s^2 - 4\pi G \rho_0}, \quad (1.31)$$

which can be thought to depend on the associated wavelength (scale) to each perturbation mode. The delimiting case,  $\lambda_J = c_s \sqrt{\pi/G\rho_0}$ , is called the Jeans length. Perturbations at scales smaller than the Jeans length evolve as sound waves with

$$\delta \rho_{\mathbf{k}} \propto \sin(\omega t + \mathbf{k}\mathbf{r}). \quad (1.32)$$

Perturbation modes with characteristic scales larger than the Jeans length instead have an exponential growth/decrease with time as

$$\delta \rho_{\mathbf{k}} \propto e^{\pm|\omega|t}. \quad (1.33)$$

From these large-scale modes, it is possible to see that gravity is very fast in increasing even the smallest perturbations. In particular, considering the very large-scale perturbation limit  $\lambda \gg \lambda_J$ , we have that  $|\omega| \rightarrow \sqrt{4\pi G \rho_0}$ . We define the characteristic collapse time

$$\tau_J = \frac{1}{\sqrt{4\pi G \rho_0}}, \quad (1.34)$$

which tells us how long it takes for a perturbation to substantially collapse. Since the only opposition to gravitational collapse is given by the fluid internal pressure, the fate of an over-density is determined by the relation between the collapse time and the time that a pressure wave takes to propagate through the fluid  $\tau_p \sim \lambda/c_s$ . If  $\tau_J < \tau_p$  (in terms of the size of the perturbation  $\lambda > c_s/\sqrt{G\rho}$ ), then the perturbation can collapse before a pressure wave can counterbalance the force of gravity.

## 1.2.2 Expanding Universe

In an expanding Universe, the energy density is no longer constant, but evolves with time  $\rho = \rho(t)$ . Combining from Equation 1.25 the divergence of Euler equation with the Poisson equation, together with considering that the velocity field obeys Hubble's law  $\Delta \mathbf{V} = H_0 \Delta \mathbf{r}$ , the first Friedmann equation 1.10 can be derived.

In order to obtain a set of decoupled differential equations, it is useful to extract the expansion of the Universe by adopting both a physical and a comoving sets of coordinates related by

$$\mathbf{r} = a(t) \mathbf{x}, \quad (1.35)$$

where  $\mathbf{r}$  is the physical distance which is equal to the product of the scale factor  $a$  with the comoving separation  $\mathbf{x}$ . Switching to the comoving set of coordinates gives the following modifications to the derivative operators:

$$\left( \frac{\partial}{\partial t} \right)_{\mathbf{r}} = \left( \frac{\partial}{\partial t} \right)_{\mathbf{x}} - (\mathbf{V}_0 \nabla_{\mathbf{r}}) \quad \text{and} \quad \nabla_{\mathbf{r}} = \frac{1}{a} \nabla_{\mathbf{x}}. \quad (1.36)$$

Applying the above transformations together with using the fractional amplitude of the matter density perturbation field  $\delta = \delta\rho/\rho_0$ , Equations 1.25 can be rewritten (substituting in the perturbed quantities and keeping only the linear order terms):

$$\begin{aligned} \frac{\partial \delta}{\partial t} + \frac{1}{a} \nabla \delta \mathbf{V} &= 0, \\ \frac{\partial \delta \mathbf{V}}{\partial t} + H \delta \mathbf{V} + \frac{c_s^2}{a} \nabla \delta + \frac{1}{a} \nabla \delta \phi &= 0, \\ \Delta \delta \phi - 4\pi G \rho_0 a^2 \delta &= 0, \end{aligned} \quad (1.37)$$

where now all the spatial derivatives are with respect to the comoving coordinate system. Taking the divergence of the perturbed hydrodynamic equation together with using the Poisson equation it is possible to derive that

$$\ddot{\delta} + 2H\dot{\delta} - \frac{c_s^2}{a^2} \Delta \delta - 4\pi G \rho_0 \delta = 0, \quad (1.38)$$

which is a closed equation for the relative matter density perturbation variable in an expanding Universe.

Let us first consider the behaviour of matter perturbations in three different scenarios for the Universe: radiation-, matter- and dark energy-dominated.

**Radiation-dominated Universe.** Considering as in the following cases only large-scales perturbations ( $k \ll k_J$ ) for which the pressure term can be safely ignored, Equation 1.38 reduces to

$$\ddot{\delta} + 2H\dot{\delta} = 0, \quad (1.39)$$

with the scale factor  $a \propto t^{\frac{1}{2}}$  as seen in Sec. 1.1.1. Depending on the size of the perturbation, two solutions are possible. The first is a static one,  $\delta(t) = \text{const.}$  while the second one describes perturbation growing logarithmically as a function of time  $\delta(t) \propto \ln t$ .

**Einstein-de Sitter Universe,  $\Omega_m = 1$ .** We already saw in Sec. 1.1.1 that when the energy content of the Universe is completely given by pressureless-like matter, then the scale factor evolves as  $a \propto t^{\frac{2}{3}}$ . Equation 1.38 then becomes in terms of the time variable  $t$

$$\ddot{\delta} + \frac{4}{3t}\dot{\delta} - \frac{2}{3t^2}\delta = 0. \quad (1.40)$$

The most general solution is given by a two term polynomial

$$\delta(\mathbf{x}, t) = At^{\frac{2}{3}} + Bt^{-1}, \quad (1.41)$$

where  $A$  and  $B$  are two constants of integration and their values can be fixed by imposing some initial conditions. The first of the two terms represents a growing mode for the perturbations while the second is a decreasing mode which can then be ignored. Comparing this growing mode with the one obtained for a static Universe in Equation 1.33, the effect of expansion is clear, perturbations grow much slower with respect to time, in particular the behaviour switches from exponential to a power law.

**De Sitter Universe,  $\Omega_\Lambda = 1$ .** If dark energy keeps behaving as a cosmological constant, the de Sitter Universe is the cosmological future limit towards which our Universe is evolving. Equation 1.38 for this case is:

$$\ddot{\delta} + 2H_0\dot{\delta} = 0, \quad (1.42)$$

where  $H_0$  is constant (as can be checked from the first Friedmann equation Equation 1.10). The above equation has two solutions, a constant one ( $\delta = \text{const}$ ) and a decaying one. For a cosmological constant dominated Universe the decaying solution is an exponential function of time  $\delta(t) \propto \exp(-H_0 t)$ . Therefore over-densities in a de Sitter Universe can remain constant but cannot further increase.

Baryonic matter perturbations can be described with good approximation by considering a fluid with equation of state  $p = p(\rho)$ . At first order then we saw that the pressure perturbation can be expressed in terms of the energy density ones ignoring entropy perturbations

$$\delta p = \left. \frac{\partial p}{\partial \rho} \right| \delta \rho = c_s^2 \delta \rho. \quad (1.43)$$

Consider the Fourier transform of the over-density field defined by the transformations:

$$\delta_{\mathbf{k}} \equiv \delta(\mathbf{k}, t) = \int d^3x e^{i\mathbf{k}\cdot\mathbf{x}} \delta(\mathbf{x}, t) \quad \text{and} \quad \delta \equiv \delta(\mathbf{x}, t) = \frac{1}{(2\pi)^3} \int d^3k e^{-i\mathbf{k}\cdot\mathbf{x}} \delta(\mathbf{k}, t). \quad (1.44)$$

Taking the Fourier transform of Equation 1.38

$$\ddot{\delta}_{\mathbf{k}} + 2H\dot{\delta}_{\mathbf{k}} + \frac{c_s^2 k^2}{a^2} \delta_{\mathbf{k}} - 4\pi G \rho_0 \delta_{\mathbf{k}} = 0, \quad (1.45)$$

which can be rewritten as

$$\ddot{\delta}_{\mathbf{k}} + 2H\dot{\delta}_{\mathbf{k}} + \omega_{\mathbf{k}}^2 \delta_{\mathbf{k}} = 0 \quad \text{where} \quad \omega_{\mathbf{k}}^2 = \frac{c_s^2 k^2}{a^2} - 4\pi G \rho_0. \quad (1.46)$$

Whether  $\omega_{\mathbf{k}}$  is an imaginary number or not determines the behaviour of the baryonic perturbations. Physically, it is easier to think about the size of these perturbation modes defined by their physical scale  $\lambda_{\text{phys}} = 2\pi a/k$ . Then  $\omega_{\mathbf{k}}^2$  can be rewritten as:

$$\omega_{\mathbf{k}}^2 = (2\pi)^2 c_s^2 \left[ \lambda_{\text{phys}}^{-2} - \lambda_{\text{J}}^{-2} \right] \quad \text{where} \quad \lambda_{\text{J}} = c_s \sqrt{\frac{\pi}{G\rho}}, \quad (1.47)$$

where the Jeans length  $\lambda_{\text{J}}$  has been used again. The fate of the perturbations is related to their scale compared to the Jeans length. If they are much greater ( $\lambda_{\text{phys}} \gg \lambda_{\text{J}}$ ) then it is clear that the perturbation equation tends to the one for the collisionless case (dark matter) and the solution will be the same.

On the other hand for much smaller perturbations ( $\lambda_{\text{phys}} \ll \lambda_{\text{J}}$ ) and the solution of Equation 1.46 is given by sound waves functions. For a speed of sounds that changes adiabatically (Mukhanov, 2005) the solution for  $\delta_{\mathbf{k}}$  is:

$$\delta_{\mathbf{k}} \propto \frac{1}{\sqrt{c_s a}} \exp\left(\pm k \int \frac{c_s dt}{a}\right). \quad (1.48)$$

An important aspect of Equation 1.47 is the fact that the relation between the physical wavelength and the Jeans wavelength evolves with time and is related to the Universe expansion.  $\lambda$  grows proportionally to the scale factor  $a$ , while  $\lambda_{\text{J}}$

is inversely proportional to  $\sqrt{\rho}$ . For example in a pressureless matter-dominated Universe with  $\rho \propto a^{-3} \Rightarrow \lambda_J \propto a^{3/2}$ , the Jeans wavelength grows faster than the physical wavelength and eventually modes which were smaller than  $\lambda_J$  become larger passing from an acoustic wave behaviour to a collision-less matter one.

**Modes larger than the Hubble radius.** From the Hubble parameter it is possible to define an important characteristic length scale, the Hubble radius. This is defined for the present time as  $r_H(t_0) = c/H(t_0) = c/H_0$ . The Hubble radius defines a three-dimensional sphere beyond which objects are receding from us with a velocity greater than the speed of light. It is used in cosmology as a proxy to determine the visible Universe. For the growth of structure, it is usually used as an *apparent* horizon to separate between modes smaller (that we can observe) and larger (which cannot be observed and therefore referred to as being *outside* the horizon) than it.

Consider then two neighbouring regions of the Universe characterised by the same Hubble parameter  $H$ , but one flat with energy density  $\rho_a$ , while the other with energy density  $\rho_b$  and curvature  $K_b$ . Equating the Friedmann equations for the two regions

$$\frac{8\pi G}{3} \rho_a = \frac{8\pi G}{3} \rho_b - \frac{K_b c^2}{a^2}. \quad (1.49)$$

Rewriting the equation in terms of  $\delta$  it is possible to see that  $\delta_a \propto (a^2 \rho)^{-1}$ . This result turns out to be correct even if the Newtonian classical approach has been used for modes larger than the Hubble radius.

From Sec. 1.1.1 we know that during radiation and matter-dominated epochs the energy density as a function of the scale factor goes as  $\rho \propto a^{-4}$  and  $\rho \propto a^{-3}$  respectively. Therefore perturbation modes outside the Hubble radius evolve as  $a^2$  and  $a$  during radiation and matter-dominated epochs, respectively. From this it is possible to see that dark matter perturbation behave in the same way both outside and within the Hubble radius during the matter-dominated epoch.

### 1.3 Statistics of the density field

Since the Universe has been observed to be isotropic and homogeneous at large enough scales, it is reasonable to assume that the perturbations of the density field can be very well approximated (still at large enough scales) by a statistically homogeneous and isotropic random field. This means that instead of showing particular features at a certain position in the Universe at a certain time, the over-density field  $\delta$  has certain statistical properties (moments) which are invariant under translation (homogeneity) and rotation (isotropy) transformations. In

order to connect theoretical predictions with observations it is necessary to develop a statistical framework for  $\delta$ . First consider the spatial average of the over-density field, by definition

$$\langle \delta(\mathbf{x}, t) \rangle = 0. \quad (1.50)$$

Usually statistical quantities like the mean are evaluated by considering several realisations of the same ensemble. However it is evident that this cannot be done in cosmology since we can observe and collect data from our one Universe. Nevertheless if far enough different patches of the Universe can be assumed to be statistically uncorrelated then they can be thought as independent realisations of the same ensemble. This property is called ergodicity, averaging over independent sub-volumes of a realisation of a random variable is equivalent to averaging over different realisations of one sub-volume.

Since the average of the over-density field is always null by definition, it is necessary to consider the second moment of the perturbations of the density field, namely the correlation function. The two-point (2pt) correlation function is defined as the joint ensemble average of the over-density field evaluated at two different points in space at the same time,

$$\langle \delta(\mathbf{x}, t) \delta(\mathbf{x} + \mathbf{r}, t) \rangle = \xi(\mathbf{x}, \mathbf{x} + \mathbf{r}, t). \quad (1.51)$$

Knowing that the over-density field  $\delta(\mathbf{x}, t)$  is statistically homogeneous and isotropic, it follows that the actual positions  $\mathbf{x}$  and  $\mathbf{x} + \mathbf{r}$  do not influence  $\xi$ , since there is no preferred location or direction. The only quantity affecting the 2pt correlation function is the separation between the two-point in space,  $|\mathbf{r}|$  and therefore  $\xi(\mathbf{x}, \mathbf{x} + \mathbf{r}, t) = \xi(|\mathbf{r}|, t)$ . The use of the 2pt correlation function in cosmology was first introduced by Peebles (1980).

As we saw in the previous Section when deriving the equations for the time-evolution of  $\delta$ , switching the analysis to Fourier space can provide useful insights. We use Equation 1.44 to convert  $\delta(\mathbf{x}, t)$  and  $\delta(\mathbf{x} + \mathbf{r})$  into  $\delta(\mathbf{k})$  and  $\delta(\mathbf{k}')$ , respectively. At least for the moment we can ignore the time dependence since all the quantities are evaluated at the same instant  $t$ . We can then compute the second moment of the density distribution in Fourier space,

$$\langle \delta(\mathbf{k}) \delta(\mathbf{k}') \rangle = (2\pi)^3 \delta_D(\mathbf{k} + \mathbf{k}') \int d^3r e^{i\mathbf{k}\cdot\mathbf{r}} \xi(r) = (2\pi)^3 \delta_D(\mathbf{k} + \mathbf{k}') P(k). \quad (1.52)$$

where the power spectrum of the density perturbations  $P(k)$  has been introduced as the Fourier transform of the 2pt correlation function.  $\delta_D$  is the Dirac's delta which imposes equality between the wave-vectors  $\mathbf{k}$  and  $\mathbf{k}'$ . Notice also that

the power spectrum is only a function of the modulus of the wave-vector  $|\mathbf{k}|$  because of the homogeneity and isotropy assumptions. Both the 2pt correlation function and the power spectrum describe how much structure is present at scales  $r \sim 2\pi/k$ .

If one considers the inflationary paradigm described earlier in Sec. 1.1.2 in order to produce the primordial density perturbations, with details depending on the precise shape of the inflationary potential  $V(\phi)$ , the theory predicts a power spectrum scaling as a power law of  $k$ . In particular, introducing the scalar spectral index  $n_s$  mentioned in Sec. 1.1.7, the power law can be expressed as  $P(k) \propto k^{n_s}$ . The case in which the power spectrum is a linear function of  $k$  ( $n_s = 1$ ) is called Harrison and Zel'dovich after the physicists who first derived this power law behaviour.

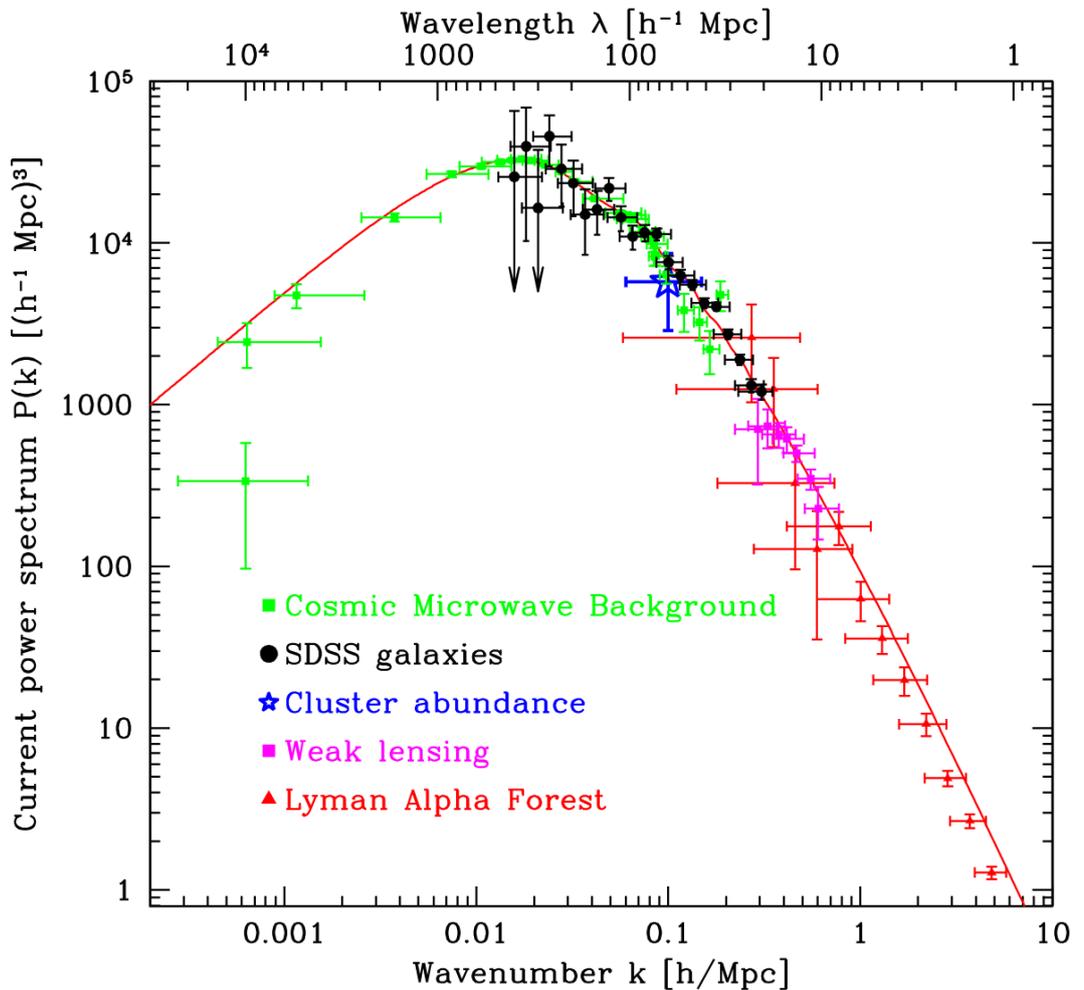


Figure 1.3: Measurements from different cosmological probes of the matter power spectrum presented in Tegmark et al. (2004). The red line shows the best fit given by a  $\Lambda$ CDM cosmology with  $\Omega_m = 0.28$ ,  $h = 0.72$ ,  $\Omega_b/\Omega_m = 0.16$  and  $\tau = 0.17$ .

### 1.3.1 From primordial to late-time matter power spectrum

In order to be able to constrain cosmological parameters with data from current and future surveys the matter power spectrum, it is necessary to derive the primordial one (or the power spectrum corresponding to an early time  $t_0$  very close to inflation's end). When solving the perturbation equations for the over-density parameter  $\delta$ , since it is assumed to be homogeneous on large enough scales, the time and spacial dependencies are separated in two different functions as in

$$\delta(\mathbf{x}, t) = \delta(\mathbf{x}) D(t), \quad (1.53)$$

where  $D(t)$  is called the *growth factor* and is the variable whose solution is found when solving for example Equation 1.38. If the same assumption as was used for the over-density field can be assumed to be valid also for the power spectrum, than its spatial and time dependencies can also be disentangled and assumed to be independent between each others as in

$$P(k, t) = D^2(t) P_0(k), \quad (1.54)$$

where  $P_0(k)$  is the power spectrum for the wave number  $k$  evaluated at an arbitrary initial time  $t_0$ .

Nevertheless this is not sufficient to connect the late-time power spectrum with the primordial one since, as we saw in Sec. 1.2.2, perturbations grow differently depending on which kind of energy density is dominating and also on when the size of the perturbations becomes smaller than the Hubble radius. In order to account for these factors, the initial power spectrum taken at an initial time  $t_0$  is corrected via the square of the *transfer function*  $T(k)$ :

$$P_0(k) \propto k^{n_s} T^2(k). \quad (1.55)$$

A first approximation for the transfer function can be computed by comparing the wave number  $k$  of a determinate perturbation mode with the size of the apparent horizon at the time of radiation-matter equality (redshift  $z_{\text{eq}}$ ). The comoving Hubble scale at  $z_{\text{eq}}$  is given by  $c/(aH)_{\text{eq}}$  and therefore it is possible to define the associated wave number  $k_{\text{eq}} = 2\pi(aH)_{\text{eq}}/c$ . If  $k < k_{\text{eq}}$  than the perturbation mode will enter the apparent horizon after the radiation-matter equality while on the contrary if  $k > k_{\text{eq}}$  it will enter the apparent horizon during the radiation-dominated epoch. Using the evolution of the growth factor with respect to time while comparing an initial perturbation  $\delta_0$  with its correspondent

one after  $z_{\text{eq}}$  it is possible to see that  $T(k) = \text{const.}$  for  $k < k_{\text{eq}}$  while  $T(k) \propto k^{-2}$  for  $k > k_{\text{eq}}$ . In Figure 1.3 it is indeed possible to see that at large-scales (small  $k$ 's) the power spectrum evolves almost linearly with respect to the wave number  $k$ , while at small scales (large  $k$ 's) it evolves as a negative power law. From the measured data points the transition around  $k_{\text{eq}}$  appears to happen gradually and not suddenly.

As we saw from Equation 1.48, on small scales the baryonic perturbation modes have an oscillatory behaviour (sound waves). As we saw in the CMB section 1.1.3 when describing the acoustic peaks, until recombination the baryons were tightly coupled to the photons. On the contrary the dark matter perturbations were free to collapse already after  $z_{\text{eq}}$ . When finally after recombination the baryons decoupled from the radiation component, their oscillations froze. In particular a prominent feature of these oscillations is given by those spherical sound waves propagating outwards from dark matter over-densities (gravitational potential wells) because of the fluid pressure. The distance of a wave-front of these baryon-photon fluid oscillations from the center of the corresponding dark matter well is equal to the product of the fluid sound speed ( $c_s$ ) and the interval of time between the radiation-matter equality (when dark matter began collapsing) and recombination. This quantity is normally called *sound horizon*.

Once the baryons decouple from the photons, at the sound horizon distance, the over-density of baryons, previously corresponding to the wave front of the baryon-photon fluid, is free to collapse because of the dark matter gravitational potential. This creates a characteristic scale in the matter density field, which physically translates into a higher probability of finding baryonic matter over-densities separated by this length, which is usually labelled *BAO-scale/peak*. The first detection of the BAO peak in the 2pt correlation function, using data from the SDSS-BOSS survey, was presented by Eisenstein et al. (2005) who obtained a value for the characteristic separation of  $\sim 100$  Mpc/h averaging over the redshift interval  $0.16 < z < 0.47$ .

In the power spectrum, the BAO peak present in the 2pt correlation function transforms into a series of wiggles as it can be seen in Figure 1.3. Therefore the BAO feature can be used to constrain cosmological parameters, such as the baryon density parameter  $\Omega_b$  or other parameters strictly related to the recombination time like  $z_*$  and  $r_*$ .

Finally the BAO characteristic length can be used to build a standard *cosmological ruler*: by measuring the BAO scale at two different redshifts, it becomes possible to study the evolution of the Universe and in particular its rate of expansion since the BAO comoving length remains constant.

### 1.3.2 Three-point statistics

Before introducing higher-order correlation functions of the over-density field, it is fundamental to remind an important property of Gaussian fields, the Wick's theorem.

**Wick's theorem:** All the statistical moments of the distribution of a zero-mean Gaussian field:

- are equal to zero if the number of variables used to compute the moment is odd (odd moments);
- are equal to the sum of all the terms given by the product of second moments which is possible to write starting from the original Gaussian variables.

In Fourier space, which as we saw is extremely useful in Cosmology, the Wick's theorem can be written as

$$\begin{aligned} \langle \delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\dots\delta(\mathbf{k}_{2n+1}) \rangle &= 0 \\ \langle \delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\dots\delta(\mathbf{k}_{2n}) \rangle &= \sum_{\text{terms}} \prod_{\substack{l,m=1\dots 2n \\ l \neq m}} \langle \delta(\mathbf{k}_l)\delta(\mathbf{k}_m) \rangle, \end{aligned} \quad (1.56)$$

where the product is over all the  $\delta$ -pairs that are possible to form with  $\mathbf{k}_l \neq \mathbf{k}_m$ . The importance of this result is given by the fact that the primordial energy density field fluctuations are predicted to be Gaussian (Guth and Pi, 1982; Starobinsky, 1982; Hawking, 1982; Bardeen et al., 1983). If that is the case then all the statistical properties of the primordial  $\delta$  are encoded in its 2pt correlation function or equivalently in its power spectrum.

However we know that gravitational instability along the history of the Universe transforms the distribution of the matter density field, in particular causing the collapse of its over-densities. Therefore a *late*-time non-Gaussian component is present in the matter over-density field. The easiest way to study it is to use higher-order moments of the  $\delta$ -distribution and the lowest is given by the 3pt correlation function in configuration space

$$\zeta(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) = \langle \delta(\mathbf{x}_1)\delta(\mathbf{x}_2)\delta(\mathbf{x}_3) \rangle_c \quad (1.57)$$

where the *connected* component of a correlation function  $\langle \rangle_c$  is defined as the ensemble average over the three  $\delta$ 's minus all the possible combination of subsets (*disconnected*). This concept will become explicit in Chapter 2 when we will consider second order perturbation theory. It is possible then to define the analogue of  $\zeta$  in Fourier space, the bispectrum

$$\langle \delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_3) \rangle_c = (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) B(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) \quad (1.58)$$

where the Dirac's delta enforces the fact that the three wave-vectors must form a connected configuration, a triangle.

## 1.4 Galaxy bias

All the statistical quantities introduced in the previous section are relative to the matter over-density field, which now for clarity we relabel  $\delta_m$ . However real 3D surveys measure the position (redshift, right ascension and declination) of galaxies which represent only the luminous component of the matter present in the Universe. Therefore it becomes necessary to understand and model the relation between luminous and dark matter in order to compare observations with theoretical predictions. How biased is the distribution of galaxies with respect to the underlying dark matter one?

The most immediate hypothesis that can be made is that they are proportional to each other

$$\delta_g(\mathbf{x}) = b \delta_m(\mathbf{x}), \quad (1.59)$$

where  $b$  is called the linear *bias* parameter and  $\delta_g$  is the local galaxy density field. However this simple form does not hold at all scales and in particular not at the small ones, where for example the creation of galaxies is influenced by other forces beside the gravitational one and by still not completely understood physical processes, like for example supernovae feedback.

Several attempts have been done by the research community to study the bias relation using hydro-dynamical simulations (Springel, 2018; Katz et al., 1999; Stinson et al., 2010; Genel et al., 2014; Pearce, 1999).

An alternative approach introduces a set of parameters describing the relation between galaxies and dark matter distribution. These parameters are then tested and fitted using real data together with the cosmological parameter of interest. Finally in order to derive constraints on the set of cosmological parameters of interest it is possible to marginalise over the posterior distributions of the bias ones.

The first attempt of modelling the bias relation was done by [Kaiser \(1984a\)](#), who described galaxy clusters as peaks of an underlying random Gaussian field distribution (dark matter). His calculations were afterwards extended by [Peacock and Heavens \(1985\)](#); [Bardeen et al. \(1986a\)](#).

If one considers large enough scales, then it is reasonable to assume that the only force determining the relation between luminous and dark matter is gravity. Analysing the bias relation at large-scales implies considering smoothed versions of the galaxy and matter density fields:

$$\delta(\mathbf{x}, R) = \int d^3y \delta(\mathbf{y}) W(\mathbf{y} - \mathbf{x}, R), \quad (1.60)$$

where  $W(\mathbf{y} - \mathbf{x}, R)$  is a smoothing window function with characteristic length  $R$  (if spherical, then  $R$  is the radius) centered at the position defined by the vector  $\mathbf{x}$ . Then the Eulerian bias model prescribes that if the galaxy over-density variable  $\tilde{\delta}_g$  is a function of the matter density one  $\tilde{\delta}_m$  at a certain position  $\mathbf{x}$ , then  $\tilde{\delta}_g$  can be Taylor expanded as

$$\tilde{\delta}_g = \sum_{i=0}^{\infty} \frac{b_i}{i!} \tilde{\delta}_m^i, \quad (1.61)$$

where it is assumed  $\delta_m \ll 1$  for large enough  $R$ . In particular at first approximation the above equation implies that the galaxy power spectrum is related to the matter one simply by:

$$P_g(k) = b_1^2 P_m(k). \quad (1.62)$$

It will be necessary to implement more bias parameters when higher-order correlation functions are considered and when higher-order perturbative corrections to the power spectrum are included.

The Euclidean model was recently extended in [McDonald and Roy \(2009a\)](#) with the addition of the dependence of the galaxy density on the local value of the gravitational potential and on the peculiar velocity field. This extension proves that by using two more bias parameters, which encode the dependencies above described, it is possible to model the power spectrum and bispectrum accurately up to fourth order perturbation theory.

Another contribution to the bias relation could be caused by the fact that galaxy formation does not depend only on the underlying dark matter distribution. Therefore the relation between  $\delta_m$  and  $\delta_g$  could have a stochastic component ([Dekel and Lahav, 1999a](#); [Taruya et al., 1999](#)):

$$\delta_g(\mathbf{x}) = \mathcal{F}[\delta_m(\mathbf{x})] + \epsilon(\mathbf{x}), \quad (1.63)$$

where  $\epsilon(\boldsymbol{x})$  is a random field parametrising the discrepancy between luminous and dark matter distributions due to non gravitational physics. However this scatter has been proved to become negligible at large enough scales for the two-point statistics (Scherrer and Weinberg, 1998).

Finally another important approach to the bias problem is the one considering an intermediate step between the overall matter distribution and the galaxy one. This intermediate step consists in studying the formation of dark matter halos (White and Rees, 1978). The formation of halos is dominated only by gravitational collapse, avoiding all the complications due to the other physical processes. In order to connect matter distribution to the galaxy one using the statistical properties of the halo distribution, the remaining challenge to overcome is to derive a model for populating the halos with galaxies (Scherrer and Bertshinger, 1991).

## 1.5 Redshift-space distortions

When real world survey data are used, the quantities measured to describe the distribution of matter in the Universe are the right ascension and declination angles  $\phi$ ,  $\theta$  and instead of the radial distance  $r$ , the data contains the redshift  $z$  (spectroscopic rather than photometric for 3D statistics studies). The conversion from  $z$  to  $r$  depends on the assumed cosmology but it is also affected by distortions caused by gravity. Indeed in the ideal case in which a galaxy would not feel the influence of any gravitational field, it would simply move further away from an observer with the Hubble flow velocity  $\boldsymbol{v} = H_0 \boldsymbol{r}$ , where  $H_0$  is the Hubble constant today.

In the real world, due to the presence of gravitational fields, galaxies have different peculiar velocities in addition to the receding movement from us due to cosmic expansion. On large-scales this effect manifests when galaxies have peculiar velocities with components due to the infall motion towards galaxy clusters center. This causes large-scale structures to appear *squashed* along the line of sight because of the combined effect of galaxies in-falling peculiar velocities and Hubble flow (Kaiser effect Kaiser, 1987). On small scales the galaxies velocity dispersion produces a "Finger of God effect" which causes structures to appear elongated in redshift-space with respect to real space. Therefore the conversion from redshift-space to real space is influenced by the line of sight component of these peculiar velocities and the conversion formula is (from now on  $s$  labels the redshift-space quantities)

$$\mathbf{s}(\mathbf{r}) = \mathbf{r} \left[ 1 + \frac{U(\mathbf{r}) - U(\mathbf{r}_0)}{r} \right], \quad \text{where} \quad U(\mathbf{r}) = \frac{\mathbf{v} \cdot \mathbf{r}}{H_0 r}. \quad (1.64)$$

In the paper by [Heavens et al. \(1998\)](#), it has been shown how to express the redshift-space galaxy overdensity perturbation variable measured in the real surveys  $\delta_{\mathbf{s}, \mathbf{k}}^g$  (the Fourier transform) to the real space matter one  $\delta_{\mathbf{k}}$ ,

$$\begin{aligned} \delta_{\mathbf{s}, \mathbf{k}}^g &= F_s^{(1)}(\mathbf{k}) \delta_{\mathbf{k}} + \frac{1}{(2\pi)^3} \int d^3 \mathbf{k}_1 d^3 \mathbf{k}_2 \delta_D(\mathbf{k} - \mathbf{k}_2 - \mathbf{k}_1) F_s^{(2)}(\mathbf{k}_1, \mathbf{k}_2) \delta_{\mathbf{k}_1} \delta_{\mathbf{k}_2} \\ &+ \frac{1}{(2\pi)^6} \int d^3 \mathbf{k}_1 d^3 \mathbf{k}_2 d^3 \mathbf{k}_3 \delta_D(\mathbf{k} - \mathbf{k}_3 - \mathbf{k}_2 - \mathbf{k}_1) F_s^{(3)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) \delta_{\mathbf{k}_1} \delta_{\mathbf{k}_2} \delta_{\mathbf{k}_3}, \end{aligned} \quad (1.65)$$

where the kernels used are:

$$\begin{aligned} F_s^{(1)}(\mathbf{k}) &= b_1 + f\mu^2 \\ F_s^{(2)}(\mathbf{k}_1, \mathbf{k}_2) &= b_1 J_s^{(2)}(\mathbf{k}_1, \mathbf{k}_2) + f\mu^2 K_s^{(2)}(\mathbf{k}_1, \mathbf{k}_2) + \frac{1}{2} b_2 \\ &+ \frac{b_1 f}{2} \left[ \mu_1^2 + \mu_2^2 + \mu_1 \mu_2 \left( \frac{k_1}{k_2} + \frac{k_2}{k_1} \right) \right] + f^2 \left[ \mu_1^2 \mu_2^2 + \frac{\mu_1 \mu_2}{2} \left( \mu_1^2 \frac{k_1}{k_2} + \mu_2^2 \frac{k_2}{k_1} \right) \right] \\ F_s^{(3)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) &= \\ &= b_1 J_s^{(3)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) + f\mu^2 K_s^{(3)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) + \frac{b_2}{2} f\mu_3^2 + \frac{b_3}{6} + \frac{b_2}{2} f\mu_1 \mu_2 \frac{k_2}{k_1} \\ &+ \frac{b_2}{2} f\mu_1 \mu_3 \frac{k_3}{k_1} + b_1 f^2 \mu_2^2 \mu_3^2 + 2b_1 f^2 \mu_1 \mu_2 \mu_3^2 \frac{k_1}{k_2} + b_1 f^2 \mu_2 \mu_3^3 \frac{k_3}{k_2} + \frac{b_1}{2} \mu_1^2 \mu_2 \mu_3 \frac{k_1^2}{k_2 k_3} \\ &+ f^3 \mu_1^2 \mu_2^2 \mu_3^2 + 3f^3 \mu_1 \mu_2^2 \mu_3^3 \frac{k_3}{k_1} + \frac{1}{2} f^3 \mu_1 \mu_2 \mu_3^4 \frac{k_3^2}{k_1 k_2} \\ &+ J_s^{(2)}(\mathbf{k}_2, \mathbf{k}_3) \left( b_2 + b_1 f \mu_1^2 + b_1 f \mu_1 \mu_{2+3} \frac{k_{2+3}}{k_1} \right) \\ &+ K_s^{(2)}(\mathbf{k}_2, \mathbf{k}_3) \left( b_1 f \mu_{2+3}^2 + b_1 f \mu_1 \mu_{2+3} \frac{k_1}{k_{2+3}} + 2f^2 \mu_1^2 \mu_{2+3}^2 \right. \\ &\left. + f^2 \mu_1 \mu_{2+3}^3 \frac{k_{2+3}}{k_1} + f^2 \mu_1^3 \mu_{2+3} \frac{k_1}{k_{2+3}} \right), \end{aligned} \quad (1.66)$$

where the  $K$ 's and  $J$ 's are given by:

$$\begin{aligned}
 J_s^{(2)}(\mathbf{k}_1, \mathbf{k}_2) &= \frac{5}{7} + \frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{2k_1 k_2} \left( \frac{k_1}{k_2} + \frac{k_2}{k_1} \right) + \frac{2}{7} \left( \frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{k_1 k_2} \right)^2 \\
 K_s^{(2)}(\mathbf{k}_1, \mathbf{k}_2) &= \frac{3}{7} + \frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{2k_1 k_2} \left( \frac{k_1}{k_2} + \frac{k_2}{k_1} \right) + \frac{4}{7} \left( \frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{k_1 k_2} \right)^2 \\
 J_s^{(3)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) &= J_s^{(2)}(\mathbf{k}_2, \mathbf{k}_3) \left[ \frac{1}{3} + \frac{1}{3} \frac{\mathbf{k}_1 \cdot (\mathbf{k}_2 + \mathbf{k}_3)}{(\mathbf{k}_2 + \mathbf{k}_3)^2} + \frac{4}{9} \frac{\mathbf{k} \cdot \mathbf{k}_1}{k_1^2} \frac{\mathbf{k} \cdot (\mathbf{k}_2 + \mathbf{k}_3)}{(\mathbf{k}_2 + \mathbf{k}_3)^2} \right] \\
 &\quad - \frac{2}{9} \frac{\mathbf{k} \cdot \mathbf{k}_1}{k_1^2} \frac{\mathbf{k} \cdot (\mathbf{k}_2 + \mathbf{k}_3)}{(\mathbf{k}_2 + \mathbf{k}_3)^2} \frac{\mathbf{k}_3 \cdot (\mathbf{k}_2 + \mathbf{k}_3)}{k_3^2} + \frac{1}{9} \frac{\mathbf{k} \cdot \mathbf{k}_2}{k_2^2} \frac{\mathbf{k} \cdot \mathbf{k}_3}{k_3^2} \\
 K_s^{(3)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) &= 3J_s^{(3)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) - \frac{\mathbf{k} \cdot \mathbf{k}_1}{k_1^2} J_s^{(2)}(\mathbf{k}_2, \mathbf{k}_3) \\
 &\quad - \frac{\mathbf{k} \cdot (\mathbf{k}_1 + \mathbf{k}_2)}{(\mathbf{k}_1 + \mathbf{k}_2)^2} K_s^{(2)}(\mathbf{k}_1, \mathbf{k}_2)
 \end{aligned} \tag{1.67}$$

where the growth factor is  $f \equiv d \ln D / d \ln \simeq \Omega_m^{0.55}$  in standard  $\Lambda$ CDM models, ( $D(a)$  growing mode of the amplitude and  $a$  is the scale factor) and  $\mu = \mathbf{k} \cdot \hat{r} / k$ ,  $\mathbf{k} = \mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3$  and  $\mu_{2+3} \equiv (\mathbf{k}_2 + \mathbf{k}_3) \cdot \hat{r} / |\mathbf{k}_2 + \mathbf{k}_3|$  with  $\hat{r} = \mathbf{r} / r$ . Each of the  $J$  and  $K$  kernels represents the Fourier transform of the solution of the perturbation equation of the same order for the over-density perturbation and velocity field divergence variables, respectively. For example  $J^{(2)}$  describes how the second order matter density perturbation variable  $\delta_m^{(2)}$  is related to the first order one  $\delta_m^{(1)}$ . The  $F$  kernels are the generalisation of the  $J$  and  $K$  ones to the galaxy field, which implies taking into account the effect of redshift space distortions. In all kernels it is possible to see that they have on average stronger amplitude when the wave-vectors are as parallel as possible (scalar products) and when they are very different in magnitude (ratio of the modules).

For the galaxy redshift-space power spectrum defined as

$$\langle \delta_{s, \mathbf{k}_1}^g \delta_{s, \mathbf{k}_2}^g \rangle = (2\pi)^3 P_s^g(\mathbf{k}_1) \delta_D(\mathbf{k}_1 + \mathbf{k}_2), \tag{1.68}$$

substituting (1.65) into the above expression and applying Wick's theorem assuming that the initial perturbation were Gaussian gets

$$\begin{aligned}
 P_s^g(\mathbf{k}) &\equiv P_{s,11}^g + P_{s,22}^g + P_{s,13}^g \\
 &= (1 + \beta\mu^2)^2 b_1^2 P_{11}(k) + 2 \int \frac{d^3 \mathbf{q}}{(2\pi)^3} P_{11}(q) P_{11}(|\mathbf{k} - \mathbf{q}|) [F_s^{(2)}(\mathbf{q}, \mathbf{k} - \mathbf{q})]^2 \\
 &\quad + 6 (1 + \beta\mu^2) b_1 P_{11}(k) \int \frac{d^3 \mathbf{q}}{(2\pi)^3} P_{11}(q) F_s^{(3)}(\mathbf{q}, -\mathbf{q}, \mathbf{k}),
 \end{aligned} \tag{1.69}$$

where  $F_s^{(3)}(\mathbf{q}, -\mathbf{q}, \mathbf{k})$  is kernel corresponding to the third order level in perturbation theory and which was originally reported in reported in Equation 1.66

from [Heavens et al. \(1998\)](#).  $P_{11}(k)$  is the linear matter power spectrum defined analogously as (1.68).

Analogously for the redshift-space galaxy Bispectrum defined as:

$$\langle \delta_{s,\mathbf{k}_1}^g \delta_{s,\mathbf{k}_2}^g \delta_{s,\mathbf{k}_3}^g \rangle = (2\pi)^3 B_s^g(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3). \quad (1.70)$$

The relation to the linear matter density power spectrum is at first order:

$$B_s^g(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = 2P_{11}(k_1)P_{11}(k_2)F_s^{(1)}(\mathbf{k}_1)F_s^{(1)}(\mathbf{k}_2)F_s^{(2)}(\mathbf{k}_1, \mathbf{k}_2) \\ + \text{cyclic terms } (1, 2) \rightarrow (2, 3) \text{ and } (3, 1). \quad (1.71)$$

## 1.6 Data and simulations

In this section are described briefly the datasets and the simulations used in this work. Additional details and potential overlaps are present in the following chapters of this thesis which are self-consistent.

### 1.6.1 SDSS-III BOSS

The Baryon Oscillation Spectroscopic Survey (BOSS [Dawson et al., 2013](#)) which is part of the Sloan Digital Sky Survey III (SDSS-III [Eisenstein et al., 2011](#)) is the source of the data used in the following chapters. SDSS created one of the largest and most accurate 3D maps of the Universe. The main purpose of BOSS was to constrain cosmological parameters through the measurement of the BAO using the clustering analysis of large-scales structures. In particular BOSS created a catalogue of 1.5 millions luminous galaxies up to redshift  $z < 0.7$  and 150,000 quasars mapping the Ly $\alpha$  forest in the redshift range  $2.15 \leq z \leq 3.5$  using the 2.5 m-aperture Sloan Foundation Telescope at Apache Point Observatory in New Mexico ([Gunn et al., 2006](#)). It covered in total an area of 14,555 square degrees in the *ugriz* band ([Fukugita et al., 1996](#)). BOSS covers a large volume ( $V_{\text{eff}} \simeq 7.4 \text{ Gpc}^3$ ) with a number density of galaxies  $n_g \sim 3 \times 10^{-4} [h\text{Mpc}^{-1}]^3$ . These specifications were chosen in order to be sure that shot noise contributions would have been subdominant at BAO scale ([White, 2011](#)).

In particular in the analysis it has been used the data-set corresponding to the CMASS sample ("constant mass"), both south and north galactic caps (SGC and NGC). The magnitude range of the CMASS galaxies is between  $17.5 < i < 19.9$  where  $i$  is the composite model magnitude. The redshift range of this sample is  $0.43 < z \leq 0.7$ . The average bias of the CMASS galaxies is  $b \sim 2$  ([Gil-Marín et al., 2015](#); [Slepian et al., 2017a](#)) with a prominent break at 4000 Å.

Table 1.3: Specifications for the DR11 and DR12 BOSS-SDSS III data releases CMASS samples (Reid, 2016; Gil-Marín et al., 2015).

	DR11	DR12	
$A_{\text{tot}}$	8,498	10,252	[deg <sup>2</sup> ]
$A_{\text{NGC}}$	6,391	7,429	[deg <sup>2</sup> ]
$A_{\text{SGC}}$	2,107	2,823	[deg <sup>2</sup> ]
$N_{\text{gal}}^{\text{tot}}$	690,827	836,347	
$N_{\text{gal}}^{\text{NGC}}$	520,806	607,357	
$N_{\text{gal}}^{\text{SGC}}$	170,021	228,990	
$V_{\text{eff}}$	6.0	7.4	[Gpc <sup>3</sup> ]

In Chapter 2 we used the Data Release 11 (DR11) data release of BOSS SDSS-III while in Chapter 4 we used the last data release DR12 (Reid, 2016). The effective redshift for both DR11 and DR12 for the CMASS sample was set to  $z_{\text{eff}} = 0.57$  (Anderson, 2012). In Table 1.3 are reported the specifications for the DR11 and DR12 CMASS samples data sets. In Figure 1.4 is shown the footprint completeness map of the DR12 CMASS sample.

## 1.6.2 Galaxy mocks

For the analysis on DR11 data, in order to estimate the covariance matrix a set of 600 galaxy mocks (Manera et al., 2013) were used for each of the CMASS galaxy caps (1200 in total). The synthetic data were based on the BOSS survey geometry and used the same weighting system for the galaxies in order to take into account systematic errors. Galaxies were assigned to dark matter halos using a Halo Occupation Distribution (HOD) prescription. The mocks are based on the *PT-halos* algorithm, using second order Lagrangian perturbation theory to implement the large-scale physical processes (Scoccimarro and Sheth, 2002; Manera et al., 2013).

In Chapter 4 the covariance matrix for the DR12 CMASS sample was estimated using a set of 2048 realizations (Kitaura et al., 2016) of the MultiDark-Patchy BOSS DR12 mocks based on augmented Lagrangian perturbation theory (ALPT Kitaura and Heß, 2013).

In total there are 12,288 light-cones available corresponding to an effective volume of  $\sim 192,000 [h^{-1}\text{Gpc}]^3$ , including a cosmic evolution in the redshift range  $0.15 < z < 0.75$ . A reference galaxy catalogue based on the Halo Abundance Matching modelling of the BOSS DR12 galaxy clustering data and on the data themselves has been used to calibrate the mocks.

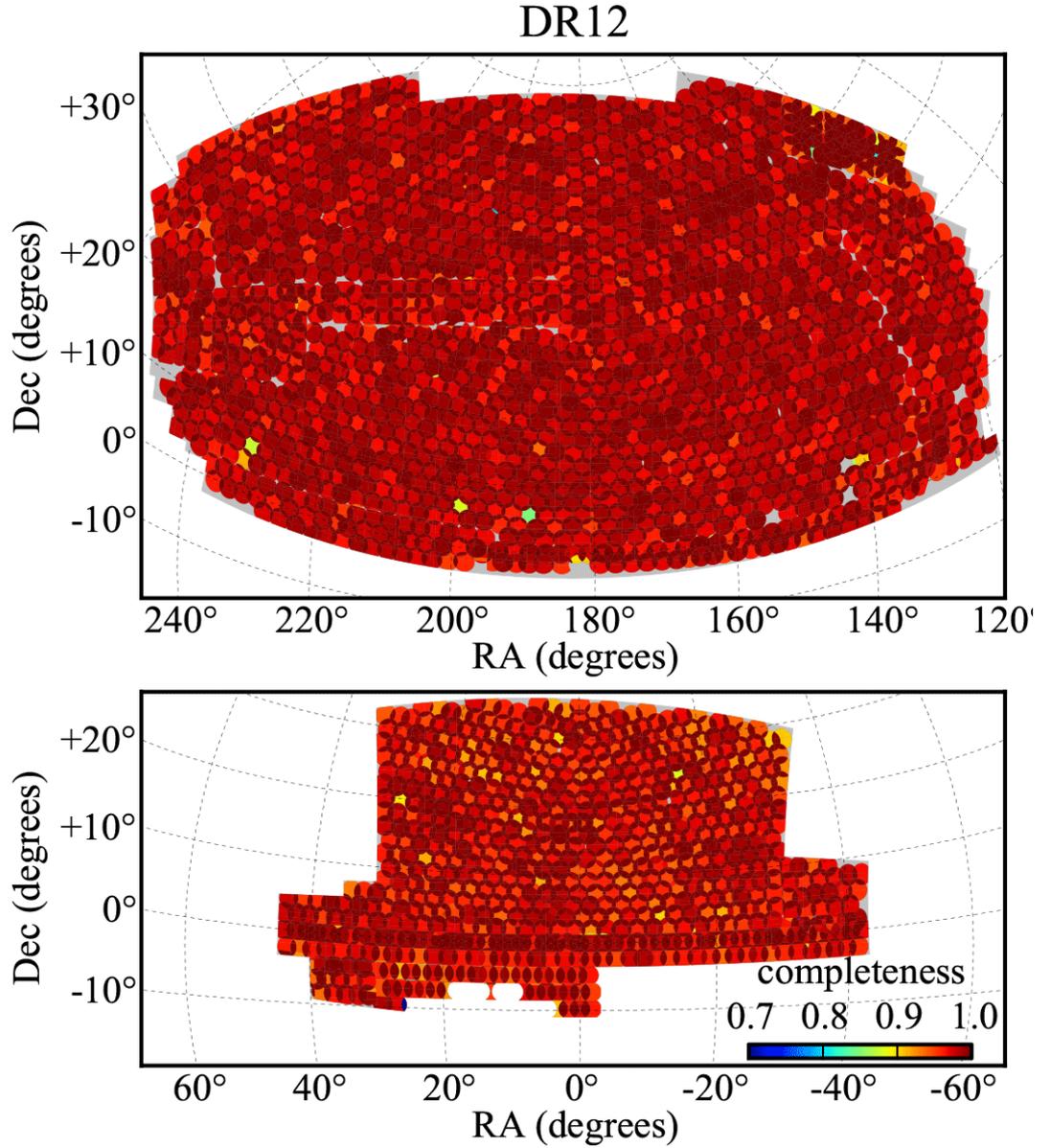


Figure 1.4: Completeness map of the BOSS SDSS-III CMASS DR12 sample from Reid (2016).

The mocks have been tested to reproduced within the 2pt correlation function  $1\sigma$  confidence interval and the power spectrum up to  $k = 0.3 h\text{Mpc}^{-1}$ .

More details are given in Chapters 2 and 4 where the mocks have been used to estimate the covariance matrices necessary to evaluate the data-vectors likelihoods.

*“There is nothing like looking, if you want to find something. You certainly usually find something, if you look, but it is not always quite the something you were after.”*

- J.R.R. Tolkien, *The Hobbit*

## 2 | Skewness: 3pt auto-correlation function

This Chapter focuses on the integrated quantity called "skewness" (Peebles, 1980) which is a function of the ratio between third and second cumulative moments of the matter overdensity distribution,  $\langle \delta^3 \rangle$  and  $\langle \delta^2 \rangle$ . Where no references are reported, the computations shown are original of this work. In this section we drop the subscript "m" since all computations concern the matter overdensity field. In turns these are the volume averages of the 3pt and 2pt correlation functions  $\zeta$  and  $\xi$ . The skewness is defined as

$$S_3(R) \equiv \frac{\langle \delta^3(R) \rangle}{\langle \delta^2(R) \rangle^2}. \quad (2.1)$$

where the ensemble average is over different regions of space (e.g. spheres) of a particular size  $R$ .

There are two main reasons for this choice of statistic. First, it represents a way to compress the information available in the 3pt correlation function. Secondly it will be shown later that another quantity can be derived from the skewness which turns out to be fairly independent from the linear galaxy bias and therefore useful in constraining cosmological parameters (by lifting the bias degeneracies).

Skewness in the matter/galaxy field naturally arises as a consequence of gravitational instability. Therefore, it is a good estimator of the late-time departure of the matter/galaxy density field from the initial primordial Gaussian conditions, for example observable in the CMB. In the case of primordial non Gaussianities (PNG), also at late-times the skewness is expected to keep trace of their presence. A target of this Chapter is to separate late-time non-Gaussianity from the primordial one in the auto-correlation functions analytic expressions.

We derive analytic expressions for  $S_3$ ,  $\langle \delta^2 \rangle$  and  $\langle \delta^3 \rangle$  including PNG contributions. In the second part of this Chapter we will use the derived model to constrain cosmological parameters of interest using BOSS DR11 data.

## 2.1 Second-order perturbation theory

Primordial perturbations are assumed to have been almost perfectly Gaussian. However gravitational collapse during the cosmic history introduced a non-linear component to the overdensity field. A good way to describe and to study this departure from linearity is to expand the overdensity variable  $\delta$  in terms of perturbative orders:

$$\delta = \delta_1 + \delta_2 + \delta_3 + \dots, \quad (2.2)$$

where the subscript indicates the order of perturbation theory (and naturally we assume that  $\delta_1 \gg \delta_2 \gg \delta_3$ ). When we write down  $\langle \delta^3 \rangle$  in terms of this expansion

$$\langle \delta^3 \rangle = \langle \delta_1 \delta_1 \delta_1 \rangle + 3 \langle \delta_1 \delta_1 \delta_2 \rangle + \dots, \quad (2.3)$$

the first term on the right hand side vanishes since all the odd moments of a Gaussian field are equal to zero (see Wick's theorem Sec. 1.3.2). In order to proceed we need to derive the relation between  $\delta_2$  and  $\delta_1$ . First we separate the space and time dependencies of the  $i$ -th order perturbation variable  $\delta_i(\mathbf{x}, t) = \delta_i(\mathbf{x}) D_i(t)$  as also done in Sec. 1.3.1. Indeed we are just interested in the time dependence since the matter density field is assumed to be homogeneous on large enough scales. For the first-order (equivalent to Equation 1.38)

$$\ddot{D}_1 + 2\frac{\dot{a}}{a}\dot{D}_1 - \frac{3}{2}\Omega_0 H_0^2 \left(\frac{a_0}{a}\right)^3 D_1 = 0. \quad (2.4)$$

Regarding the second-order, we want to reproduce the result obtained first by Peebles, 1980 following the procedure presented in Kamionkowski and Buchalter, 1999, starting from the second-order equation

$$\begin{aligned} \ddot{\delta}_2 + 2\frac{\dot{a}}{a}\dot{\delta}_2 - \frac{3}{2}\Omega_0 H_0^2 \left(\frac{a_0}{a}\right)^3 \delta_2 &= \left[ \frac{3}{2}\Omega_0 H_0^2 \left(\frac{a_0}{a}\right)^3 + \left(\frac{\dot{D}_1}{D_1}\right)^2 \right] D_1^2 \delta_1^2 \\ &+ \left[ \frac{3}{2}\Omega_0 H_0^2 \left(\frac{a_0}{a}\right)^3 + 2\left(\frac{\dot{D}_1}{D_1}\right)^2 \right] D_1^2 \delta_{1,i} \Delta_{1,i} \\ &+ \left(\frac{\dot{D}_1}{D_1}\right)^2 D_1^2 \Delta_{1,ij} \Delta_{1,ij}, \end{aligned} \quad (2.5)$$

where

$$\Delta_1(\mathbf{x}) = -\frac{1}{4\pi} \int d^3x' \frac{\delta(\mathbf{x})}{|\mathbf{x} - \mathbf{x}'|}. \quad (2.6)$$

Unfortunately Equation 2.5 cannot be solved directly because of the two different time dependencies on the right hand side, the time dependent part of the first-order perturbation solution and its derivative. The solution of 2.5 can be found by considering the two equations each given by the left hand side of Equation 2.5 together with one of the two different time dependencies Kamionkowski and Buchalter (1999). Writing also the second-order perturbation variable as  $\delta_2(\mathbf{x}, t) = D_2(t)\delta_2(\mathbf{x})$ , the two equations are

$$\begin{aligned} \delta_2(\mathbf{x}) \times \left[ \ddot{D}_{2,a} + 2H\dot{D}_{2,a} - \frac{3}{2}\Omega_0 H_0^2 \left(\frac{a_0}{a}\right)^3 D_{2,a} \right] &= \frac{3}{2}\Omega_0 H_0^2 \left(\frac{a_0}{a}\right)^3 D_1^2 [\delta_1^2 + \delta_{1,i}\Delta_{1,i}] \\ \delta_2(\mathbf{x}) \times \left[ \ddot{D}_{2,b} + 2H\dot{D}_{2,b} - \frac{3}{2}\Omega_0 H_0^2 \left(\frac{a_0}{a}\right)^3 D_{2,b} \right] &= \dot{D}_1^2 [\delta_1^2 + \delta_{1,i}\Delta_{1,i} + \Delta_{1,ij}\Delta_{1,ij}]. \end{aligned} \quad (2.7)$$

In order to decompose Equation 2.5 in the last two equations we impose the initial conditions that the time dependent part of the first-order perturbation solution and its derivative are zero for  $t = 0$ . As a consequence, the solution can be written as

$$\delta_2 = (D_{2,a} + D_{2,b})\delta_1^2 + (D_{2,a} + 2D_{2,b})\delta_{1,i}\Delta_{1,i} + D_{2,b}\Delta_{1,ij}\Delta_{1,ij}. \quad (2.8)$$

It is important to notice that Equations 2.7 and 2.8 are not really independent. For example it can be verified that the following relation holds

$$D_{2,b} = \frac{1}{2}(D_1^2 - D_{2,a}). \quad (2.9)$$

The first line in Equations 2.7 can be analytically solved in the case of an Einstein-de Sitter universe ( $\Omega_m = 1$ ). For this kind of model we have the standard relations

$$\begin{aligned} a \propto t^{\frac{2}{3}}, \quad D_1(t) = t^{\frac{2}{3}}, \quad 6\pi G\rho = t^{-2} \\ \text{since } \Omega_0 = \frac{8\pi G}{3H_0^2}\rho_0 \implies t^{-2} = \frac{9}{4}\Omega_0 a^{-3}H_0^2. \end{aligned} \quad (2.10)$$

Substituting Equation 2.10 into the first line of Equation 2.7 returns

$$\ddot{x} + \frac{4}{3}t^{-1}\dot{x} - \frac{2}{3}t^{-2}x = \frac{2}{3}t^{-\frac{2}{3}}, \quad (2.11)$$

which has the solution

$$x = \frac{3}{7}t^{\frac{4}{3}} = \frac{3}{7}D_1^2 = D_{2,a}. \quad (2.12)$$

From the relation in Equation 2.9, gives  $D_{2,b} = \frac{2}{7}t^{\frac{4}{3}}$ . Finally substituting these two results into Equation 2.8 we have the expression for the second-order matter density perturbation variable for an Einstein-de Sitter universe in terms of the first-order one

$$\delta_2 = \frac{5}{7}\delta_1^2 + \delta_{1,i}\Delta_{1,i} + \frac{2}{7}\Delta_{1,ij}\Delta_{1,ij}. \quad (2.13)$$

## 2.2 Skewness derivation

Before deriving an expression for  $S_3$ , it is useful to work out the relation between the variance of the matter density field  $\langle \delta^2 \rangle$  and the matter power spectrum  $P(k)$ , starting from the definition of  $\langle \delta^2 \rangle$  as the averaged two-point correlation function. Here and in the following steps we introduce the concept of "smoothing window function", which defines a maximum scale for computing the cumulative function. By definition we have that:

$$\begin{aligned} \langle \delta^2(\mathbf{x}_0) \rangle &= \left\langle \int d^3x d^3y W(\mathbf{y} - \mathbf{x}_0) W(\mathbf{x} - \mathbf{x}_0) \delta(\mathbf{x}) \delta(\mathbf{y}) \right\rangle \\ &= \int d^3x d^3y W(\mathbf{y} - \mathbf{x}_0) W(\mathbf{x} - \mathbf{x}_0) \xi(\mathbf{x} - \mathbf{y}). \end{aligned} \quad (2.14)$$

Recalling that the two-point correlation function  $\xi$  is also the Fourier transform of the power spectrum

$$\xi(\mathbf{x} - \mathbf{y}) = \frac{1}{(2\pi)^3} \int d^3k P(k) e^{i\mathbf{k}(\mathbf{x}-\mathbf{y})}, \quad (2.15)$$

and substituting it into the Equation 2.14 it gives the final result:

$$\begin{aligned} \langle \delta^2(\mathbf{x}_0) \rangle &= \frac{1}{(2\pi)^3} \int d^3k P(k) \left[ \int d^3x W(\mathbf{x} - \mathbf{x}_0) e^{i\mathbf{k}\mathbf{x}} \right] \left[ \int d^3y W(\mathbf{y} - \mathbf{x}_0) e^{-i\mathbf{k}\mathbf{y}} \right] \\ &= \frac{1}{(2\pi)^3} \int d^3k P(k) W_k^2 = \frac{1}{2\pi^2} \int dk P(k) W_k^2 k^2 = \sigma^2, \end{aligned} \quad (2.16)$$

where the Fourier transform of a spherical top-hat window function of radius  $R$  whose Fourier transform is given by

$$W(k) = \frac{3}{(kR)^3} [\sin(kR) - kR \cos(kR)]. \quad (2.17)$$

When  $R = 8 \text{ Mpc}/h$  the square root of Equation 2.16 gives the cosmological parameter  $\sigma_8$ , which is usually considered as a proxy of the dark matter field oscillation amplitude.

Consider the numerator of  $S_3$ , the three-point auto-correlation function  $\langle \delta^3 \rangle$ . When this is expanded up to the second-order in perturbation theory as

$$\langle \delta(\mathbf{x}_0)\delta(\mathbf{x}_0)\delta(\mathbf{x}_0) \rangle = \langle \delta_1(\mathbf{x}_0)\delta_1(\mathbf{x}_0)\delta_1(\mathbf{x}_0) \rangle + 3\langle \delta_2(\mathbf{x}_0)\delta_1(\mathbf{x}_0)\delta_1(\mathbf{x}_0) \rangle + \dots, \quad (2.18)$$

the first term on the right hand side vanishes because the first-order perturbation variable  $\delta_1$  is assumed to be a Gaussian field. We consider only the lowest order remaining term:

$$\langle \delta(\mathbf{x}_0)^3 \rangle = \int d^3x_1 d^3x_2 d^3x_3 \prod_{i=1}^3 W(\mathbf{x}_i - \mathbf{x}_0) 3\langle \delta_1(\mathbf{x}_1)\delta_1(\mathbf{x}_2)\delta_2(\mathbf{x}_3) \rangle. \quad (2.19)$$

Now recalling the expression in Equation 2.13 for the second-order perturbation variable as a function of the first, it is useful to switch to Fourier space using the following transformations previously introduced in Equation 1.44.

In order to transform the non-linear term in the expression in Equation 2.13 we use the convolution formula found in Fry, 1984 :

$$\begin{aligned} \text{F.T.}\{F_1(\mathbf{x})\dots F_N(\mathbf{x})\}(\mathbf{k}) &= \int \frac{d^3k_1}{(2\pi)^3} \dots \frac{d^3k_N}{(2\pi)^3} \left[ (2\pi)^3 \delta_D \left( \sum \mathbf{k}_i - \mathbf{k} \right) \right] \tilde{F}_1(\mathbf{k}_1)\dots\tilde{F}_N(\mathbf{k}_N) \\ &= \tilde{F}_1 * \dots * \tilde{F}_N, \end{aligned} \quad (2.20)$$

together with the useful relation (Fry, 1984)  $\tilde{\Delta}(\mathbf{k}) = -\frac{1}{k^2}\tilde{\delta}(\mathbf{k})$ . In this way one obtains

$$\begin{aligned} \text{F.T.}\{\delta_2(\mathbf{x})\} &= \\ &= \int \frac{d^3k_1}{(2\pi)^3} \frac{d^3k_2}{(2\pi)^3} (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{k}) \left[ \frac{5}{7} + \frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{k_2^2} + \frac{2(\mathbf{k}_1 \cdot \mathbf{k}_2)^2}{7k_1^2 k_2^2} \right] \tilde{\delta}(\mathbf{k}_1)\tilde{\delta}(\mathbf{k}_2). \end{aligned} \quad (2.21)$$

Therefore the three-point cumulative Fourier transform becomes:

$$\begin{aligned} \langle \delta(\mathbf{x}_0)^3 \rangle &= 3 \int \prod_{i=1}^3 d^3x_i W(\mathbf{x}_i - \mathbf{x}_0) \langle \int \prod_{i=1}^3 \left( \frac{d^3k_i}{(2\pi)^3} \right) e^{-i\mathbf{k}_i \cdot \mathbf{x}_i} \tilde{\delta}_1(\mathbf{k}_1)\tilde{\delta}_1(\mathbf{k}_2) \times \\ &\quad \times \int \frac{d^3k_4}{(2\pi)^3} \frac{d^3k_5}{(2\pi)^3} (2\pi)^3 \delta_D(\mathbf{k}_4 + \mathbf{k}_5 - \mathbf{k}_3) A(\mathbf{k}_4, \mathbf{k}_5) \tilde{\delta}(\mathbf{k}_4)\tilde{\delta}(\mathbf{k}_5) \rangle, \end{aligned} \quad (2.22)$$

where

$$A(\mathbf{k}_4, \mathbf{k}_5) = \left[ \frac{5}{7} + \frac{\mathbf{k}_4 \cdot \mathbf{k}_5}{k_5^2} + \frac{2(\mathbf{k}_4 \cdot \mathbf{k}_5)^2}{7k_4^2 k_5^2} \right]. \quad (2.23)$$

Using the delta function we arrive to the expression

$$\begin{aligned} \langle \delta(\mathbf{x}_0)^3 \rangle &= 3 \int \prod_{i=1}^4 d^3 x_i W(\mathbf{x}_i - \mathbf{x}_0) \int \prod_{i=1}^3 \frac{d^3 k_i}{(2\pi)^3} e^{-i(\mathbf{k}_1 \mathbf{x}_1 + \mathbf{k}_2 \mathbf{x}_2 + (\mathbf{k}_3 + \mathbf{k}_4) \mathbf{x}_3)} \\ &\quad \times A(\mathbf{k}_4, \mathbf{k}_5) \langle \tilde{\delta}(\mathbf{k}_1) \tilde{\delta}(\mathbf{k}_2) \tilde{\delta}(\mathbf{k}_3) \tilde{\delta}(\mathbf{k}_4) \rangle. \end{aligned} \quad (2.24)$$

As explained in [Amendola and Tsujikawa, 2010](#) chapter 12, the correlation term  $\langle \dots \rangle$  for Gaussian variables vanishes except for identical  $k$ -pairs since different modes are independent. This implies that:

$$\langle \tilde{\delta}(\mathbf{k}_1) \tilde{\delta}(\mathbf{k}_2) \tilde{\delta}(\mathbf{k}_4) \tilde{\delta}(\mathbf{k}_5) \rangle = 2 \times (2\pi)^6 \delta_D(\mathbf{k}_1 - \mathbf{k}_4) \delta_D(\mathbf{k}_2 - \mathbf{k}_5) P(k_1) P(k_2), \quad (2.25)$$

where  $P(k)$  is the power spectrum. Therefore

$$\begin{aligned} \langle \delta(\mathbf{x}_0)^3 \rangle &= \frac{6}{(2\pi)^6} \int \prod_{i=1}^3 d^3 x_i W(\mathbf{x}_i - \mathbf{x}_0) \int d^3 k_4 d^3 k_5 e^{-i(\mathbf{k}_4 \mathbf{x}_1 + \mathbf{k}_5 \mathbf{x}_2 + (\mathbf{k}_4 + \mathbf{k}_5) \mathbf{x}_3)} \\ &\quad \times A(\mathbf{k}_4, \mathbf{k}_5) P(\mathbf{k}_4) P(\mathbf{k}_5). \end{aligned} \quad (2.26)$$

Defining the Fourier transform of a window function as

$$W_{\mathbf{k}} = \int d^3 x W(\mathbf{x} - \mathbf{x}_0) e^{-i\mathbf{k}\mathbf{x}}, \quad (2.27)$$

the three-point auto-correlation function becomes

$$\langle \delta(\mathbf{x}_0)^3 \rangle = \frac{6}{(2\pi)^6} \int d^3 k_4 d^3 k_5 W_{\mathbf{k}_4} W_{\mathbf{k}_5} W_{|\mathbf{k}_4 + \mathbf{k}_5|} A(\mathbf{k}_4, \mathbf{k}_5) P(k_4) P(k_5). \quad (2.28)$$

For spherical top-hat window functions (Equation 2.17), we then have two useful relations for this type of window functions [Bernardeau et al., 2002](#):

$$\begin{aligned} \int \frac{d\Omega_{45}}{(4\pi)^2} W_{|\mathbf{k}_4 + \mathbf{k}_5|} \left[ 1 - \frac{(\mathbf{k}_4 \cdot \mathbf{k}_5)^2}{\mathbf{k}_4^2 \mathbf{k}_5^2} \right] &= \frac{2}{3} W(k_4) W(k_5) \\ \int \frac{d\Omega_{45}}{(4\pi)^2} W_{|\mathbf{k}_4 + \mathbf{k}_5|} \left[ 1 + \frac{\mathbf{k}_4 \cdot \mathbf{k}_5}{\mathbf{k}_5^2} \right] &= W(k_4) \left[ W(k_5) + \frac{1}{3} k_5 R W'(k_5) \right], \end{aligned} \quad (2.29)$$

where  $W(k) \equiv W(kR)$  and  $W'(k) = \frac{dW(kR)}{d(kR)}$ . Rearranging the kernel

$$A(\mathbf{k}_4, \mathbf{k}_5) = \left[ \left( 1 + \frac{\mathbf{k}_4 \mathbf{k}_5}{\mathbf{k}_5^2} \right) - \frac{2}{7} \left( 1 - \frac{(\mathbf{k}_4 \mathbf{k}_5)^2}{\mathbf{k}_4^2 \mathbf{k}_5^2} \right) \right], \quad (2.30)$$

it follows that

$$\langle \delta(\mathbf{x}_0)^3 \rangle = \frac{4}{(2\pi)^4} \int dk_4 dk_5 P(k_4) P(k_5) \left[ \frac{34}{7} W_{k_4}^2 W_{k_5}^2 + 2k_5 R W_{k_4}^2 W_{k_5} W'_{k_5} \right] k_4^2 k_5^2. \quad (2.31)$$

Finally it is possible to compute  $S_3$ :

$$S_3 = \frac{\langle \delta^3 \rangle}{\langle \delta^2 \rangle^2} = \frac{34}{7} + \frac{2}{2\pi^2} \frac{1}{\sigma^2} \int dk R k^3 W_k W'_k P(k). \quad (2.32)$$

which was first derived by Peebles (1980) for infinitesimal cells. The last step involves noticing that the window function is a function of  $kR$  and deriving the variance with respect to  $R$ :

$$\frac{d}{dR} \langle \delta^2 \rangle = \frac{1}{\pi^2} \int dk k^3 P(k) W(kR) W'(kR), \quad (2.33)$$

giving the skewness in the compact form:

$$S_3 = \frac{34}{7} + \frac{d \log \sigma^2}{d \log R}. \quad (2.34)$$

In order to compute the skewness with numerical codes via the linear matter power spectrum like for example CAMB (Lewis et al., 2000), it is necessary to compute the explicit expression of  $\sigma^2$  and of  $S_3$ . Following the procedure presented in the appendix of Juszkiewicz (2013) we have that

$$W'(kR) = \frac{dW(kR)}{d(kR)} = \frac{9}{(kR)^4} (kR \cos(kR) - \sin(kR)) + \frac{3}{(kR)^2} \sin(kR), \quad (2.35)$$

we have that:

$$S_3 = \frac{34}{7} + \frac{27}{\sigma^2} \frac{1}{R^3 \pi^2} \int dk \frac{1}{(kR)^4} \left[ \frac{1}{2} \sin(2kR) \left( 2kR - \frac{1}{3}(kR)^3 \right) + \sin^2(kR) \left( \frac{4}{3}(kR)^2 - 1 \right) - (kR)^2 \right] P(k) \quad (2.36)$$

$$\sigma^2 = \frac{1}{2\pi^2} \frac{9}{R^3} \int dk \frac{1}{(kR)^4} \left[ \sin^2(kR) + (kR)^2 \cos^2(kR) - 2kR \sin(kR) \cos(kR) \right] P(k). \quad (2.37)$$

Since the kernel of this expression is a function with infinite zero points, in order to integrate it in an accurate way, it has been necessary to write an apposite code which separately integrates each interval between the zeros of the kernel and adds all those contributions together.

## 2.3 Skewness for a $\Omega_m = 0$ and $\Omega_\Lambda = 1$ Universe

We take now the extreme limit of what it is considered today the concordance model:  $\Lambda$ CDM cosmology. In this cosmology we have that for  $t \rightarrow \infty$ ,  $\Omega_m \rightarrow 0$

while  $\Omega_\Lambda \rightarrow 1$ . In order to compute the second-order expression of the matter density perturbation field in this case Equation 2.4 becomes:

$$\begin{aligned} \ddot{D}_1(t) + 2H\dot{D}_1(t) &= 0 \\ \implies \dot{D}_1(t) &= B e^{-2Ht} \quad \text{and} \quad D_1(t) = \frac{B}{2H} [c_0 - e^{-2Ht}], \end{aligned} \quad (2.38)$$

where the integration has been performed between a constant initial time  $t_i$  and a generic final time  $t$ . In order to fix the initial conditions for the Cauchy problem to solve the differential equation (the constants  $B, c_0$ ) there are two possibilities for the choice of the initial time  $t_i$ :

- setting  $t_i = 0$  is equivalent to computing the initial conditions for a generic time deep into the dark energy-dominated epoch, in this case  $c_0 = 1$ ;
- setting  $t_i = t_{eq}$  where  $t_{eq} \Leftrightarrow \rho_m = \rho_\Lambda$  and in this case the initial conditions require a more careful analysis in order for them to be fixed.

In this type of universe, we have to consider only the second of Equations 2.7 which simplifies to:

$$\begin{aligned} \ddot{D}_2 + 2H\dot{D}_2 &= \dot{D}_1^2 \\ \implies D_2(t) &= \frac{B^2}{8H^2} e^{-4Ht} + \frac{c_1}{H} e^{-2Ht} + c_2. \end{aligned} \quad (2.39)$$

The goal is to obtain, as in the case of an Einstein - de Sitter Universe (EdS), an expression of  $D_2(t)$  as a function of  $D_1(t)$  in order to proceed in the same way for the computation of the skewness. In order to do this it is necessary to fix the constants  $B, c_0, c_1, c_2$ , imposing specific conditions on  $D_1(t_i), \dot{D}_1(t_i), D_2(t_i), \dot{D}_2(t_i)$ .

Let's consider the first of the two initial conditions fixing options. If we follow the same assumptions made in the EdS case we have that:

$$\begin{cases} D_1(0) = 0 & \implies c_0 = 1 \\ \dot{D}_1(0) = \text{const} & \implies B = \text{const} \\ D_2(0) = 0 & \implies c_2 = \frac{B^2}{8H^2} \\ \dot{D}_2(0) = 0 & \implies c_2 = -\frac{B^2}{4H}. \end{cases} \quad (2.40)$$

It is immediate then to see that in this case it is possible to express the second-order solution in term of the first-order one:

$$D_2(t) = \frac{B^2}{8H^2} (e^{-4Ht} - 2e^{-2Ht} + 1) \implies D_2(t) = \frac{1}{2} D_1(t)^2. \quad (2.41)$$

Proceeding as in the EdS case and substituting Equation 2.41 into Equation 2.8, we obtain the analogous expression of Equation 2.13:

$$\delta_2 = \frac{1}{2}\delta_1^2 + \delta_{1,i}\Delta_{1,i} + \frac{1}{2}\Delta_{1,ij}\Delta_{1,ij}, \quad (2.42)$$

which, reproducing the skewness computations made before, implies that the new analogous of the kernel in Equation 2.23 is:

$$A_2(\mathbf{k}_1, \mathbf{k}_2) = \left[ \frac{1}{2} + \frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{k_2^2} + \frac{1}{2} \frac{(\mathbf{k}_1 \cdot \mathbf{k}_2)^2}{k_1^2 k_2^2} \right], \quad (2.43)$$

In order to use the geometrical properties of the top-hat window functions we can rewrite  $A_2(\mathbf{k}_1, \mathbf{k}_2)$  as

$$A_2(\mathbf{k}_1, \mathbf{k}_2) = \left[ \left( 1 + \frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{k_2^2} \right) - \frac{1}{2} \left( 1 - \frac{(\mathbf{k}_1 \cdot \mathbf{k}_2)^2}{k_1^2 k_2^2} \right) \right]. \quad (2.44)$$

Recalling the three-point function expression in Equation 2.28

$$\langle \delta(\mathbf{x}_0)^3 \rangle = \frac{6}{(2\pi)^6} \int d^3k_1 d^3k_2 W_{k_1} W_{k_2} W_{|\mathbf{k}_1+\mathbf{k}_2|} A_2(\mathbf{k}_1, \mathbf{k}_2) P(k_1) P(k_2), \quad (2.45)$$

we can now use the geometric relations in Equation 2.29 and obtain

$$\begin{aligned} \langle \delta(\mathbf{x}_0)^3 \rangle &= \\ &= \frac{4}{(2\pi)^4} \int dk_1 dk_2 P(k_1) P(k_2) \left[ 6 \left( 1 - \frac{1}{3} \right) W_{k_1}^2 W_{k_2}^2 + 2k_2 R W_{k_1}^2 W_{k_2} W'_{k_2} \right] k_1^2 k_2^2 \\ &= \frac{4}{(2\pi)^4} \int dk_1 dk_2 P(k_1) P(k_2) \left[ 4W_{k_1}^2 W_{k_2}^2 + 2k_2 R W_{k_1}^2 W_{k_2} W'_{k_2} \right] k_1^2 k_2^2. \end{aligned} \quad (2.46)$$

Proceeding as in the standard case it is then finally possible to write the compact expression for the skewness  $S_{\Lambda,3}$  like in the case of an EdS universe:

$$S_{\Lambda,3} = 4 + \frac{d \log \sigma^2}{d \log R}, \quad (2.47)$$

from which we can observe a small difference from the standard result  $\Delta_{S_3} = S_3 - S_{\Lambda,3} = \frac{6}{7}$ . The fact that the skewness parameter results smaller in a dark energy dominated universe than in a EdS one agrees with the physical intuition that an accelerated expansion phase tends to smooth the existing gravitational inhomogeneities.

Considering a more realistic case in which the dark energy dominated epoch is the extreme limit which follows after a cold dark matter dominated one, the

computation should be done with the following boundary conditions (where  $t = 0$  corresponds to an initial time in the dark energy dominated epoch):

$$\begin{cases} D_1(0) = a_1 & \implies a_1 = \frac{B}{2H} [c_0 - 1] \\ \dot{D}_1(0) = B & \implies B = B \\ D_2(0) = a_2 & \implies a_2 = \frac{B^2}{8H^2} + \frac{c_1}{H} + c_2 \\ \dot{D}_2(0) = a_3 & \implies a_3 = -\frac{B^2}{2H} - 2c_1, \end{cases} \quad (2.48)$$

where  $a_i$  are constants which must be properly fixed depending on the choice of the initial time  $t = 0$  and precisely The cosmology corresponding to that time. Solving the system one finds for the first and second-order solutions:

$$\begin{aligned} D_1(t) &= a_1 + \frac{B}{2H} (1 - e^{-2Ht}) \\ D_2(t) &= \frac{B^2}{8H^2} e^{-4Ht} - \left( \frac{a_3}{2H} + \frac{B^2}{4H^2} \right) e^{-2Ht} + a_2 + \frac{B^2}{8H^2} + \frac{a_3}{2H}, \end{aligned} \quad (2.49)$$

from which it is possible to see that in order to maintain the relation in Equation 2.41 and the same result for the skewness we have that only one of the  $a_i$  is truly a free parameter together with  $B$ . Indeed one obtains the constraints

$$a_1 B = a_3 \quad \text{and} \quad a_2 = \frac{a_1^2}{2}, \quad (2.50)$$

which physically makes sense since the second-order perturbation solution is derived from the first-order one and hence its initial conditions should be bounded to the first-order ones.

## 2.4 Primordial non-Gaussianity contribution

Following the results for the cases in which the primordial gravitational field has been assumed to be perfectly Gaussian, it is now presented an analysis of the effects on the late-time gravitationally induced non-linearities by the presence of a particular kind of primordial non Gaussianities (PNG).

### 2.4.1 Three-point contribution

The case in which the distribution of the primordial anisotropies slightly deviated from perfect Gaussianity can be mathematically expressed with the primordial curvature perturbation field  $\Phi$  represented as a function of an auxiliary gaussian field  $\phi$ ,

$$\Phi^{\text{prim.}}(\mathbf{x}) = \phi^{\text{prim.}}(\mathbf{x}) + \frac{f_{NL}}{c^2} \left( \phi^{\text{prim.}}(\mathbf{x})^2 - \langle \phi^{\text{prim.}}(\mathbf{x})^2 \rangle \right), \quad (2.51)$$

where  $f_{NL}$  expresses in particular the "local non-Gaussianity" of the field and it is assumed to be constant. The primordial potential  $\Phi^{\text{prim.}}$  is related to the late-time one  $\Phi^{\text{l.t.}}$  by the relation (Scoccimarro et al., 2004):

$$\Phi^{\text{l.t.}}(a) = \frac{9}{10} \frac{D_+}{a} \mathbb{T}(k) \Phi^{\text{prim.}}, \quad (2.52)$$

where  $\mathbb{T}(k)$  is the transfer function,  $D_+(a)$  is the growth factor from linear perturbation theory (it has been neglected the anisotropic stresses). Also, at late-times we can express the total non-linearity of the matter density contrast field, both due to PNG and gravitational collapse, adding a second-order term in the expansion of the perturbation variable:

$$\delta(\mathbf{x}) = \delta_1(\mathbf{x}) + \delta_2(\mathbf{x}) + \text{higher orders}. \quad (2.53)$$

The target of the following computation is to obtain the contribution of PNG at first-order in  $f_{NL}$ . The steps can be schematically seen below as

$$\begin{aligned} \langle \delta^3 \rangle &\xrightarrow{\text{expand in}} f(\Phi^{\text{l.t.}}) \xrightarrow{\text{substitute with}} f(\Phi^{\text{prim.}}) \\ &\xrightarrow{\text{expand}} f(\phi^{\text{prim.}} + f_{NL}(\phi_{\text{prim.}}^2 + \langle \phi_{\text{prim.}}^2 \rangle)) \\ &\xrightarrow{\text{express as}} f(P_\phi^i(k_i)) \xrightarrow{\text{end in}} f(P_m^{\text{l.t.}}), \end{aligned} \quad (2.54)$$

where  $P_\phi^i(k_i)$  is the power spectrum of the primordial curvature perturbation  $\phi$  relative to the wave vector  $\vec{k}_i$ .  $P_m^{\text{l.t.}}(k_i)$  is the late-time matter power spectrum. At late-times we have the Poisson equation

$$\nabla^2 \Phi^{\text{l.t.}}(\mathbf{x}, a) = \frac{3}{2} \frac{\Omega_m H_0^2}{a} \delta(\mathbf{x}, a), \quad (2.55)$$

which combined with Equation 2.52 gives the relation in Fourier space

$$\delta_k(\Phi^{\text{prim.}}) = \frac{3}{5} \frac{D_+}{\Omega_m H_0^2} k^2 \mathbb{T}(k) \Phi^{\text{prim.}} = \beta k^2 \mathbb{T}(k) \Phi^{\text{prim.}}. \quad (2.56)$$

We then consider then the three-point function

$$\begin{aligned} \langle \delta(\mathbf{x}_0) \delta(\mathbf{x}_0) \delta(\mathbf{x}_0) \rangle &= \langle (\delta_1(\mathbf{x}_0) + \delta_2(\mathbf{x}_0)) (\delta_1(\mathbf{x}_0) + \delta_2(\mathbf{x}_0)) (\delta_1(\mathbf{x}_0) + \delta_2(\mathbf{x}_0)) \rangle \\ &= \langle \delta_1(\mathbf{x}_0) \delta_1(\mathbf{x}_0) \delta_1(\mathbf{x}_0) \rangle + 3 \langle \delta_1(\mathbf{x}_0) \delta_1(\mathbf{x}_0) \delta_2(\mathbf{x}_0) \rangle \\ &+ \text{higher orders}. \end{aligned} \quad (2.57)$$

The first term of Equation 2.57 would normally be equal to zero, being an odd moment of what is usually a Gaussian field. Instead in the case of PNG this is no longer true. Using the relation in Equation 2.52 together with the Poisson equation in order to pass from late-time to primordial potential (the transfer function will be added after going to Fourier space), the first term expands into

$$\begin{aligned}
 & \langle \delta_1(\mathbf{x}_0)\delta_1(\mathbf{x}_0)\delta_1(\mathbf{x}_0) \rangle = \\
 & = \int \prod_{i=1}^3 d^3 \mathbf{x}_i W(\mathbf{x}_i - \mathbf{x}_0) \langle \delta(\mathbf{x}_1)\delta(\mathbf{x}_2)\delta(\mathbf{x}_3) \rangle \\
 & = \int \prod_{i=1}^3 d^3 \mathbf{x}_i W(\mathbf{x}_i - \mathbf{x}_0) \times \int \prod_{i=1}^3 \frac{d^3 \mathbf{k}_i}{(2\pi)^3} e^{-i\mathbf{k}_i \mathbf{x}_i} \langle \delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_3) \rangle \\
 & = \int \prod_{i=1}^3 \frac{d^3 \mathbf{k}_i}{(2\pi)^3} W(\mathbf{k}_i, R) \langle \delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_3) \rangle \\
 & = \int \prod_{i=1}^3 \frac{d^3 \mathbf{k}_i}{(2\pi)^3} \mathbf{k}_i^2 \mathbb{T}(\mathbf{k}_i) \beta^3 W(\mathbf{k}_i, R) \langle \Phi_{\mathbf{k}_1}^p \Phi_{\mathbf{k}_2}^p \Phi_{\mathbf{k}_3}^p \rangle, \tag{2.58}
 \end{aligned}$$

where  $W(\mathbf{k}_i, R)$  is the Fourier transform of a top-hat window function with radius  $R$ . Expanding in Fourier space each non-linear primordial curvature potential using the convolution formula reported in (Fry, 1984)

$$\Phi_{\mathbf{k}_i}^p = \phi_{\mathbf{k}_i} + \frac{f_{NL}}{c^2} \left( \int \frac{d^3 \mathbf{k}_a d^3 \mathbf{k}_b}{(2\pi)^3} \delta_D(\mathbf{k}_a + \mathbf{k}_b - \mathbf{k}_i) \phi_{\mathbf{k}_a} \phi_{\mathbf{k}_b} - \delta_D(\mathbf{k}_i) \langle \phi^2 \rangle \right), \tag{2.59}$$

it is possible to resume the computation on the first term of the three-point auto-correlation function in Equation 2.58:

$$\begin{aligned}
 & \langle \delta_1(\mathbf{x}_0)\delta_1(\mathbf{x}_0)\delta_1(\mathbf{x}_0) \rangle = \\
 & = \beta^3 \int \prod_{i=1}^3 \frac{d^3 \mathbf{k}_i}{(2\pi)^3} \mathbf{k}_i^2 \mathbb{T}(\mathbf{k}_i) W(\mathbf{k}_i, R) \times \\
 & \times \left\{ \left\langle \left[ \phi_{\mathbf{k}_1} + \frac{f_{NL}}{c^2} \int \frac{d^3 \mathbf{k}_4 d^3 \mathbf{k}_5}{(2\pi)^3} \delta_D(\mathbf{k}_4 + \mathbf{k}_5 - \mathbf{k}_1) \phi_{\mathbf{k}_4} \phi_{\mathbf{k}_5} - \frac{f_{NL}}{c^2} \delta_D(\mathbf{k}_1) \langle \phi^2 \rangle \right] \right. \right. \\
 & \times \left[ \phi_{\mathbf{k}_2} + \frac{f_{NL}}{c^2} \int \frac{d^3 \mathbf{k}_6 d^3 \mathbf{k}_7}{(2\pi)^3} \delta_D(\mathbf{k}_6 + \mathbf{k}_7 - \mathbf{k}_2) \phi_{\mathbf{k}_6} \phi_{\mathbf{k}_7} - \frac{f_{NL}}{c^2} \delta_D(\mathbf{k}_2) \langle \phi^2 \rangle \right] \\
 & \left. \left. \times \left[ \phi_{\mathbf{k}_3} + \frac{f_{NL}}{c^2} \int \frac{d^3 \mathbf{k}_8 d^3 \mathbf{k}_9}{(2\pi)^3} \delta_D(\mathbf{k}_8 + \mathbf{k}_9 - \mathbf{k}_3) \phi_{\mathbf{k}_8} \phi_{\mathbf{k}_9} - \frac{f_{NL}}{c^2} \delta_D(\mathbf{k}_3) \langle \phi^2 \rangle \right] \right\} \right\}. \tag{2.60}
 \end{aligned}$$

From Equation 2.60, remembering that  $\phi$  is assumed to be a Gaussian random field and therefore all the odd moments vanish,  $\langle \phi^{2n+1} \rangle = 0$ , all the terms  $\propto f_{NL}^2$  vanish. The surviving term at order  $\propto f_{NL}$  is then

$$\begin{aligned}
 & \langle \delta_1(\mathbf{x}_0) \delta_1(\mathbf{x}_0) \delta_1(\mathbf{x}_0) \rangle \\
 &= \frac{\beta^3 f_{NL}}{c^2} \int \prod_{i=1}^3 \frac{d^3 \mathbf{k}_i}{(2\pi)^3} \mathbf{k}_i^2 \mathbb{T}(\mathbf{k}_i) \mathbb{W}(\mathbf{k}_i, R) \times \\
 & \times \left\{ \int \frac{d^3 \mathbf{k}_4 d^3 \mathbf{k}_5}{(2\pi)^3} \delta_D(\mathbf{k}_4 + \mathbf{k}_5 - \mathbf{k}_1) \langle \phi_{\mathbf{k}_2} \phi_{\mathbf{k}_3} \phi_{\mathbf{k}_4} \phi_{\mathbf{k}_5} \rangle - \delta_D(\mathbf{k}_1) \langle \phi_{\mathbf{k}_2} \phi_{\mathbf{k}_3} \langle \phi^2 \rangle \rangle \right. \\
 & + \int \frac{d^3 \mathbf{k}_6 d^3 \mathbf{k}_7}{(2\pi)^3} \delta_D(\mathbf{k}_6 + \mathbf{k}_7 - \mathbf{k}_2) \langle \phi_{\mathbf{k}_1} \phi_{\mathbf{k}_3} \phi_{\mathbf{k}_6} \phi_{\mathbf{k}_7} \rangle - \delta_D(\mathbf{k}_2) \langle \phi_{\mathbf{k}_1} \phi_{\mathbf{k}_3} \langle \phi^2 \rangle \rangle \\
 & \left. + \int \frac{d^3 \mathbf{k}_8 d^3 \mathbf{k}_9}{(2\pi)^3} \delta_D(\mathbf{k}_8 + \mathbf{k}_9 - \mathbf{k}_3) \langle \phi_{\mathbf{k}_1} \phi_{\mathbf{k}_2} \phi_{\mathbf{k}_8} \phi_{\mathbf{k}_9} \rangle - \delta_D(\mathbf{k}_3) \langle \phi_{\mathbf{k}_1} \phi_{\mathbf{k}_2} \langle \phi^2 \rangle \rangle \right\}. \tag{2.61}
 \end{aligned}$$

Consider for example the first term, applying the Wick's theorem gets

$$\begin{aligned}
 & \int \frac{d^3 \mathbf{k}_4 d^3 \mathbf{k}_5}{(2\pi)^3} \delta_D(\mathbf{k}_4 + \mathbf{k}_5 - \mathbf{k}_1) \left[ \langle \phi_{\mathbf{k}_2} \phi_{\mathbf{k}_4} \rangle \langle \phi_{\mathbf{k}_3} \phi_{\mathbf{k}_5} \rangle + \langle \phi_{\mathbf{k}_2} \phi_{\mathbf{k}_5} \rangle \langle \phi_{\mathbf{k}_3} \phi_{\mathbf{k}_4} \rangle + \langle \phi_{\mathbf{k}_2} \phi_{\mathbf{k}_3} \rangle \langle \phi_{\mathbf{k}_4} \phi_{\mathbf{k}_5} \rangle \right] \\
 & - \delta_D(\mathbf{k}_1) \langle \phi_{\mathbf{k}_2} \phi_{\mathbf{k}_3} \rangle \langle \phi^2 \rangle. \tag{2.62}
 \end{aligned}$$

Only the first two terms inside the square brackets survive. Using the definition of power spectrum for Gaussian variables  $\langle \phi_{\mathbf{k}_i} \phi_{\mathbf{k}_j} \rangle = (2\pi)^3 \delta_D(\mathbf{k}_i + \mathbf{k}_j) P(k_i)$  Equation 2.61 reduces to

$$\begin{aligned}
 & \langle \delta_1(\mathbf{x}_0) \delta_1(\mathbf{x}_0) \delta_1(\mathbf{x}_0) \rangle = \\
 &= \frac{\beta^3 f_{NL}}{c^2} \int \prod_{i=1}^3 \frac{d^3 \mathbf{k}_i}{(2\pi)^3} \mathbf{k}_i^2 \mathbb{T}(\mathbf{k}_i) \mathbb{W}(\mathbf{k}_i, R) \times \\
 & \times \left\{ 2 \int \frac{d^3 \mathbf{k}_4 d^3 \mathbf{k}_5}{(2\pi)^3} \delta_D(\mathbf{k}_4 + \mathbf{k}_5 - \mathbf{k}_1) (2\pi)^6 \delta_D(\mathbf{k}_4 + \mathbf{k}_2) \delta_D(\mathbf{k}_5 + \mathbf{k}_3) P_\phi(\mathbf{k}_2) P_\phi(\mathbf{k}_3) + \right. \\
 & + 2 \int \frac{d^3 \mathbf{k}_6 d^3 \mathbf{k}_7}{(2\pi)^3} \delta_D(\mathbf{k}_6 + \mathbf{k}_7 - \mathbf{k}_2) (2\pi)^6 \delta_D(\mathbf{k}_6 + \mathbf{k}_1) \delta_D(\mathbf{k}_7 + \mathbf{k}_3) P_\phi(\mathbf{k}_1) P_\phi(\mathbf{k}_3) + \\
 & \left. + 2 \int \frac{d^3 \mathbf{k}_8 d^3 \mathbf{k}_9}{(2\pi)^3} \delta_D(\mathbf{k}_8 + \mathbf{k}_9 - \mathbf{k}_3) (2\pi)^6 \delta_D(\mathbf{k}_8 + \mathbf{k}_1) \delta_D(\mathbf{k}_9 + \mathbf{k}_2) P_\phi(\mathbf{k}_1) P_\phi(\mathbf{k}_2) \right\}, \tag{2.63}
 \end{aligned}$$

which shows that the three terms are all the same integrals

$$\begin{aligned}
 & \langle \delta_1(\mathbf{x}_0) \delta_1(\mathbf{x}_0) \delta_1(\mathbf{x}_0) \rangle = \\
 &= \frac{6\beta^3 f_{NL}}{c^2} \int \prod_{i=1}^3 \frac{d^3 \mathbf{k}_i}{(2\pi)^3} \mathbf{k}_i^2 \mathbb{T}(\mathbf{k}_i) \mathbb{W}(\mathbf{k}_i, R) \delta_D(\mathbf{k}_3 + \mathbf{k}_2 - \mathbf{k}_1) P(\mathbf{k}_2) P(\mathbf{k}_3) (2\pi)^3 \\
 &= \frac{6\beta^3 f_{NL}}{c^2 (2\pi)^6} \int d^3 \mathbf{k}_2 d^3 \mathbf{k}_3 \mathbf{k}_2^2 \mathbf{k}_3^2 (\mathbf{k}_2 + \mathbf{k}_3)^2 \mathbb{T}_{\mathbf{k}_2} \mathbb{T}_{\mathbf{k}_3} \mathbb{T}_{\mathbf{k}_3 + \mathbf{k}_2} \mathbb{W}_{\mathbf{k}_2} \mathbb{W}_{\mathbf{k}_3} \mathbb{W}_{\mathbf{k}_3 + \mathbf{k}_2} P_{\mathbf{k}_2}^\phi P_{\mathbf{k}_3}^\phi. \tag{2.64}
 \end{aligned}$$

In order to integrate the angular part we use the approximation introduced by [Scoccimarro et al. \(2004\)](#)

$$\int \frac{d\Omega_{12}}{(4\pi)^2} \mathbb{T}_{\mathbf{k}_3+\mathbf{k}_2} W_{\mathbf{k}_3+\mathbf{k}_2} (\mathbf{k}_2 + \mathbf{k}_3)^2 \approx k_3^2 \mathbb{T}_{k_3} W_{k_3} \left( W_{k_2} + \frac{k_2 R}{3} W'_{k_2} \right) + k_3 \leftrightarrow k_2, \quad (2.65)$$

which, substituted in Equation 2.64 and noticing that the two symmetric terms are equivalent once integrated over all the  $k$ 's for both  $k_1$  and  $k_2$ , gives:

$$\begin{aligned} \langle \delta_1(\mathbf{x}_0) \delta_1(\mathbf{x}_0) \delta_1(\mathbf{x}_0) \rangle &= \\ &= \frac{48\beta^3 f_{NL}}{c^2 (2\pi)^4} \int dk_2 dk_3 k_2^6 k_3^4 \mathbb{T}_{k_2}^2 \mathbb{T}_{k_3} W_{k_2}^2 \left[ W_{k_3}^2 + \frac{k_3 R}{3} W_{k_3} W'_{k_3} \right] P_\phi(k_2) P_\phi(k_3). \end{aligned} \quad (2.66)$$

Using the standard relation from [Dodelson \(2003\)](#) to relate the primordial power spectrum with the matter one

$$P_m(k) = \beta^2 k^4 \mathbb{T}_k^2 P_\phi(k), \quad (2.67)$$

Equation 2.65 can be written as

$$\begin{aligned} \langle \delta_1(\mathbf{x}_0) \delta_1(\mathbf{x}_0) \delta_1(\mathbf{x}_0) \rangle &= \frac{5\Omega_m H_0^2 f_{NL}}{c^2 D_+ \pi^4} \int dk_2 dk_3 k_2^2 \mathbb{T}_{k_3}^{-1} \times \\ &\times W_{k_2}^2 \left[ W_{k_3}^2 + \frac{k_3 R}{3} W_{k_3} W'_{k_3} \right] P_m(k_2) P_m(k_3). \end{aligned} \quad (2.68)$$

For second term of the three-point auto-correlation function, following the same procedure one obtains

$$\begin{aligned}
 & 3\langle \delta_1(\mathbf{x}_0)\delta_1(\mathbf{x}_0)\delta_2(\mathbf{x}_0) \rangle = \\
 & = 3 \int \prod_{i=1}^3 d^3 \mathbf{x}_i W(\mathbf{x}_i - \mathbf{x}_0) \langle \delta_1(\mathbf{x}_1)\delta_1(\mathbf{x}_2)\delta_2(\mathbf{x}_3) \rangle \\
 & = 3 \int \prod_{i=1}^3 d^3 \mathbf{x}_i W(\mathbf{x}_i - \mathbf{x}_0) \int \prod_{i=1}^3 \frac{d^3 \mathbf{k}_i}{(2\pi)^3} e^{-i\mathbf{k}_i \cdot \mathbf{x}_i} \times \\
 & \times \int \frac{d^3 \mathbf{k}_4}{(2\pi)^3} d^3 \mathbf{k}_5 \delta_D(\mathbf{k}_4 + \mathbf{k}_5 - \mathbf{k}_3) F_2[\mathbf{k}_4, \mathbf{k}_5] \langle \delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_4)\delta(\mathbf{k}_5) \rangle \\
 & = 3 \int \prod_{i=1}^4 \frac{d^3 \mathbf{k}_i}{(2\pi)^3} W_{\mathbf{k}_1} W_{\mathbf{k}_2} W_{\mathbf{k}_3+\mathbf{k}_4} F_2[\mathbf{k}_3, \mathbf{k}_4] \langle \delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_3)\delta(\mathbf{k}_4) \rangle \\
 & = 3\beta^3 \int \prod_{i=1}^4 \frac{d^3 \mathbf{k}_i}{(2\pi)^3} \mathbf{k}_i^2 \Gamma(\mathbf{k}_i) W_{\mathbf{k}_1} W_{\mathbf{k}_2} W_{\mathbf{k}_3+\mathbf{k}_4} F_2[\mathbf{k}_3, \mathbf{k}_4] \langle \Phi_{\mathbf{k}_1}^P \Phi_{\mathbf{k}_2}^P \Phi_{\mathbf{k}_3}^P \Phi_{\mathbf{k}_4}^P \rangle \\
 & = 3\beta^3 \int \prod_{i=1}^4 \frac{d^3 \mathbf{k}_i}{(2\pi)^3} \mathbf{k}_i^2 \Gamma(\mathbf{k}_i) W_{\mathbf{k}_1} W_{\mathbf{k}_2} W_{\mathbf{k}_3+\mathbf{k}_4} F_2[\mathbf{k}_3, \mathbf{k}_4] \times \\
 & \times \left\langle \left[ \phi_{\mathbf{k}_1} + \frac{f_{NL}}{c^2} \int \frac{d^3 \mathbf{k}_5 d^3 \mathbf{k}_6}{(2\pi)^3} \delta_D(\mathbf{k}_5 + \mathbf{k}_6 - \mathbf{k}_1) \phi_{\mathbf{k}_5} \phi_{\mathbf{k}_6} - \frac{f_{NL}}{c^2} \delta_D(\mathbf{k}_1) \langle \phi^2 \rangle \right] \right. \\
 & \times \left[ \phi_{\mathbf{k}_2} + \frac{f_{NL}}{c^2} \int \frac{d^3 \mathbf{k}_7 d^3 \mathbf{k}_8}{(2\pi)^3} \delta_D(\mathbf{k}_7 + \mathbf{k}_8 - \mathbf{k}_2) \phi_{\mathbf{k}_7} \phi_{\mathbf{k}_8} - \frac{f_{NL}}{c^2} \delta_D(\mathbf{k}_2) \langle \phi^2 \rangle \right] \\
 & \times \left[ \phi_{\mathbf{k}_3} + \frac{f_{NL}}{c^2} \int \frac{d^3 \mathbf{k}_9 d^3 \mathbf{k}_{10}}{(2\pi)^3} \delta_D(\mathbf{k}_9 + \mathbf{k}_{10} - \mathbf{k}_3) \phi_{\mathbf{k}_9} \phi_{\mathbf{k}_{10}} - \frac{f_{NL}}{c^2} \delta_D(\mathbf{k}_3) \langle \phi^2 \rangle \right] \\
 & \left. \times \left[ \phi_{\mathbf{k}_4} + \frac{f_{NL}}{c^2} \int \frac{d^3 \mathbf{k}_{11} d^3 \mathbf{k}_{12}}{(2\pi)^3} \delta_D(\mathbf{k}_{11} + \mathbf{k}_{12} - \mathbf{k}_4) \phi_{\mathbf{k}_{11}} \phi_{\mathbf{k}_{12}} - \frac{f_{NL}}{c^2} \delta_D(\mathbf{k}_4) \langle \phi^2 \rangle \right] \right\rangle, \tag{2.69}
 \end{aligned}$$

where the second-order perturbation theory kernel (Fry, 1984; Peebles, 1980) is

$$F_2[\mathbf{k}_a, \mathbf{k}_b] = \left[ \frac{5}{7} + \frac{\mathbf{k}_a \cdot \mathbf{k}_b}{k_a^2} + \frac{2(\mathbf{k}_a \cdot \mathbf{k}_b)^2}{7 k_a^2 k_b^2} \right]. \tag{2.70}$$

Notice from the terms in the last expression in Equation 2.69:

- the only term without  $f_{NL}$  gives the standard first-order term in the 3pt auto-correlation function which is also the numerator of  $S_3$ ;
- all the terms  $\propto f_{NL}$  are formed by odd moments of a Gaussian random variable ( $\phi$ ) and hence vanish;
- the first terms non-vanishing are of order  $f_{NL}^2$  which are beyond the purpose of this work.

### 2.4.2 Two-point contribution

In order to compute later the influence of PNG on the  $S_3$  parameter, we must know its contribution to the variance. We proceed in the same way as in the three-point case

$$\begin{aligned} \langle \delta(\mathbf{x}_1)\delta(\mathbf{x}_2) \rangle &= \langle (\delta_1(\mathbf{x}_1) + \delta_2(\mathbf{x}_1))(\delta_1(\mathbf{x}_2) + \delta_2(\mathbf{x}_2)) \rangle = \\ &= \langle \delta_1(\mathbf{x}_1)\delta_1(\mathbf{x}_2) \rangle + 2\langle \delta_1(\mathbf{x}_1)\delta_2(\mathbf{x}_2) \rangle + \langle \delta_2(\mathbf{x}_1)\delta_2(\mathbf{x}_2) \rangle + \text{high. ord.s.} \end{aligned} \quad (2.71)$$

The first term, when PNG is not considered would give the standard variance, but with PNG the expansion is

$$\begin{aligned} \langle \delta_1(\mathbf{x}_0)\delta_1(\mathbf{x}_0) \rangle &= \\ &= \int \prod_{i=1}^2 d^3\mathbf{x}_i W(\mathbf{x}_i - \mathbf{x}_0) \langle \delta_1(\mathbf{x}_1)\delta_1(\mathbf{x}_2) \rangle = \int \prod_{i=1}^2 \frac{d^3\mathbf{k}_i}{(2\pi)^3} W_{\mathbf{k}_1} W_{\mathbf{k}_2} \langle \delta(\mathbf{k}_1)\delta(\mathbf{k}_2) \rangle \\ &= \beta^2 \int \prod_{i=1}^2 \frac{d^3\mathbf{k}_i}{(2\pi)^3} \mathbf{k}_i^2 T(\mathbf{k}_i) W_{\mathbf{k}_i} \langle \Phi_{\mathbf{k}_1}^p \Phi_{\mathbf{k}_2}^p \rangle \\ &= \beta^2 \int \prod_{i=1}^2 \frac{d^3\mathbf{k}_i}{(2\pi)^3} \mathbf{k}_i^2 T(\mathbf{k}_i) W_{\mathbf{k}_i} \times \\ &\times \left\langle \left[ \phi_{\mathbf{k}_1} + \frac{f_{NL}}{c^2} \int \frac{d^3\mathbf{k}_3 d^3\mathbf{k}_4}{(2\pi)^3} \delta_D(\mathbf{k}_3 + \mathbf{k}_4 - \mathbf{k}_1) \phi_{\mathbf{k}_3} \phi_{\mathbf{k}_4} - \frac{f_{NL}}{c^2} \delta_D(\mathbf{k}_1) \langle \phi^2 \rangle \right] \right. \\ &\times \left. \left[ \phi_{\mathbf{k}_2} + \frac{f_{NL}}{c^2} \int \frac{d^3\mathbf{k}_5 d^3\mathbf{k}_6}{(2\pi)^3} \delta_D(\mathbf{k}_5 + \mathbf{k}_6 - \mathbf{k}_2) \phi_{\mathbf{k}_5} \phi_{\mathbf{k}_6} - \frac{f_{NL}}{c^2} \delta_D(\mathbf{k}_2) \langle \phi^2 \rangle \right] \right\rangle, \end{aligned} \quad (2.72)$$

from which:

- the term without  $f_{NL}$  gives the standard variance;
- all the terms  $\propto f_{NL}$  are odd moments of a Gaussian distribution and therefore they vanish;
- the first non-linear term to not vanish is of order  $f_{NL}^2$ .

Simplifying as before, the term proportional to  $f_{NL}^2$  is

$$\langle \delta_1(\mathbf{x}_0)\delta_1(\mathbf{x}_0) \rangle = 2 \left( \frac{f_{NL}\beta}{c^2} \right)^2 \int \frac{d^3\mathbf{k}_3 d^3\mathbf{k}_4}{(2\pi)^6} W_{|\mathbf{k}_3+\mathbf{k}_4|}^2 T_{|\mathbf{k}_3+\mathbf{k}_4|}^2 (\mathbf{k}_3 + \mathbf{k}_4)^4 P_\phi(\mathbf{k}_3) P_\phi(\mathbf{k}_3), \quad (2.73)$$

which can be easily converted to the matter power spectrum.

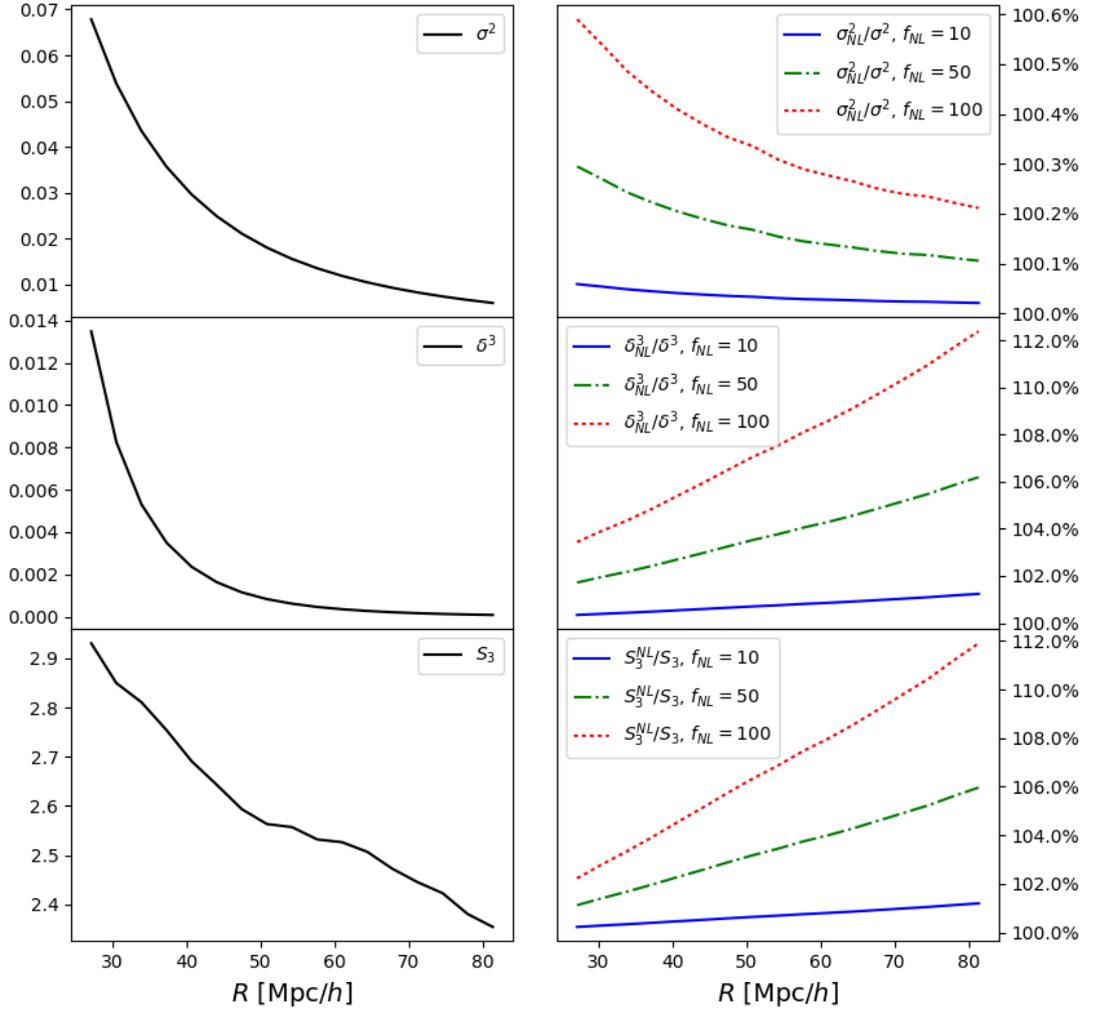


Figure 2.1:

**left column:** standard quantities  $\sigma^2$ ,  $\delta^3$  and  $S_3$  computed in the case of  $f_{NL} = 0$  as a function of the window function radius  $R$  using Equations 2.36 for a standard  $\Lambda$ CDM cosmology.

**right column:** Ratio between the quantities  $\sigma_{NL}^2$ ,  $\delta_{NL}^3$  and  $S_3^{NL}$  and the standard  $\sigma^2$ ,  $\delta^3$ ,  $S_3$  (computed assuming  $f_{NL} = 0$ ) for different values of  $f_{NL}$ , using the expression derived in Equations 2.68, 2.76 and 2.77.

Next we analyse the second term of the expansion in Equation 2.71.

$$\begin{aligned}
 & 2\langle\delta_1(\mathbf{x}_0)\delta_2(\mathbf{x}_0)\rangle = \\
 & = 2 \int \prod_{i=1}^2 d^3\mathbf{x}_i W(\mathbf{x}_i - \mathbf{x}_0) \int \prod_{i=1}^2 \frac{d^3\mathbf{k}_i}{(2\pi)^3} e^{-i\mathbf{k}_i\mathbf{x}_i} \times \\
 & \times \int \frac{d^3\mathbf{k}_3}{(2\pi)^3} d^3\mathbf{k}_4 \delta_D(\mathbf{k}_3 + \mathbf{k}_4 - \mathbf{k}_2) F_2[\mathbf{k}_3, \mathbf{k}_4] \langle\delta(\mathbf{k}_1)\delta(\mathbf{k}_3)\delta(\mathbf{k}_4)\rangle \\
 & = 2\beta^3 \int \prod_{i=1}^3 \frac{d^3\mathbf{k}_i}{(2\pi)^3} \mathbf{k}_i^2 T(\mathbf{k}_i) W_{\mathbf{k}_1} W_{\mathbf{k}_2+\mathbf{k}_3} F_2[\mathbf{k}_2, \mathbf{k}_3] \langle\Phi_{\mathbf{k}_1}^p \Phi_{\mathbf{k}_2}^p \Phi_{\mathbf{k}_3}^p\rangle \\
 & = 2\beta^3 \int \prod_{i=1}^3 \frac{d^3\mathbf{k}_i}{(2\pi)^3} \mathbf{k}_i^2 T(\mathbf{k}_i) W_{\mathbf{k}_1} W_{\mathbf{k}_2+\mathbf{k}_3} F_2[\mathbf{k}_2, \mathbf{k}_3] \times \\
 & \times \left\langle \left[ \phi_{\mathbf{k}_1} + \frac{f_{NL}}{c^2} \int \frac{d^3\mathbf{k}_4 d^3\mathbf{k}_5}{(2\pi)^3} \delta_D(\mathbf{k}_4 + \mathbf{k}_5 - \mathbf{k}_1) \phi_{\mathbf{k}_4} \phi_{\mathbf{k}_5} - \frac{f_{NL}}{c^2} \delta_D(\mathbf{k}_1) \langle\phi^2\rangle \right] \right. \\
 & \times \left[ \phi_{\mathbf{k}_2} + \frac{f_{NL}}{c^2} \int \frac{d^3\mathbf{k}_6 d^3\mathbf{k}_7}{(2\pi)^3} \delta_D(\mathbf{k}_6 + \mathbf{k}_7 - \mathbf{k}_2) \phi_{\mathbf{k}_6} \phi_{\mathbf{k}_7} - \frac{f_{NL}}{c^2} \delta_D(\mathbf{k}_2) \langle\phi^2\rangle \right] \\
 & \times \left. \left[ \phi_{\mathbf{k}_3} + \frac{f_{NL}}{c^2} \int \frac{d^3\mathbf{k}_8 d^3\mathbf{k}_9}{(2\pi)^3} \delta_D(\mathbf{k}_8 + \mathbf{k}_9 - \mathbf{k}_3) \phi_{\mathbf{k}_8} \phi_{\mathbf{k}_9} - \frac{f_{NL}}{c^2} \delta_D(\mathbf{k}_3) \langle\phi^2\rangle \right] \right\rangle \\
 & \tag{2.74}
 \end{aligned}$$

In the same fashion of the three-point case, from Equation 2.74 it is clear that:

- the term without  $f_{NL}$  that goes as  $\propto \langle\tilde{\phi}(\mathbf{k}_1)\tilde{\phi}(\mathbf{k}_2)\tilde{\phi}(\mathbf{k}_3)\rangle$  vanishes always because it is an odd moment of a gaussian field;
- terms proportional to  $f_{NL}^2$  vanish for the same reason;
- the three terms proportional to  $f_{NL}$  do not vanish and can be divided in two cases, depending on the relation between the  $\mathbf{k}$ 's inside the two kernels in the integral.

In particular, using Wick's theorem to simplify the resulting expression, one derives two terms

$$\begin{aligned}
 I_a & = 2\beta^3 \frac{f_{NL}}{c^2} \int \prod_{i=1}^3 \frac{d^3\mathbf{k}_i}{(2\pi)^3} \mathbf{k}_i^2 T(\mathbf{k}_i) W_{\mathbf{k}_1} W_{\mathbf{k}_2+\mathbf{k}_3} F_2[\mathbf{k}_2, \mathbf{k}_3] \times \\
 & \times \int \frac{d^3\mathbf{k}_4 d^3\mathbf{k}_5}{(2\pi)^3} \delta_D(\mathbf{k}_4 + \mathbf{k}_5 - \mathbf{k}_1) \times 2(2\pi)^6 \delta_D(\mathbf{k}_4 + \mathbf{k}_2) \delta_D(\mathbf{k}_5 + \mathbf{k}_3) P_\phi(k_2) P_\phi(k_3) \\
 I_b & = 4\beta^3 \frac{f_{NL}}{c^2} \int \prod_{i=1}^3 \frac{d^3\mathbf{k}_i}{(2\pi)^3} \mathbf{k}_i^2 T(\mathbf{k}_i) W_{\mathbf{k}_1} W_{\mathbf{k}_2+\mathbf{k}_3} F_2[\mathbf{k}_2, \mathbf{k}_3] \times \\
 & \times \int \frac{d^3\mathbf{k}_6 d^3\mathbf{k}_7}{(2\pi)^3} \delta_D(\mathbf{k}_6 + \mathbf{k}_7 - \mathbf{k}_2) \times 2(2\pi)^6 \delta_D(\mathbf{k}_6 + \mathbf{k}_1) \delta_D(\mathbf{k}_7 + \mathbf{k}_3) P_\phi(k_1) P_\phi(k_3). \\
 & \tag{2.75}
 \end{aligned}$$

Proceeding as in the three-point case, for  $I_a$  one obtains

$$I_a = \frac{4\beta^3 f_{NL}}{c^2 (2\pi)^6} \int d^3 \mathbf{k}_2 d^3 \mathbf{k}_3 \mathbf{k}_2^2 \mathbf{k}_3^2 (\mathbf{k}_2 + \mathbf{k}_3)^2 T_{\mathbf{k}_2} T_{\mathbf{k}_3} T_{|\mathbf{k}_2 + \mathbf{k}_3|} W_{|\mathbf{k}_2 + \mathbf{k}_3|}^2 F_2[\mathbf{k}_2, \mathbf{k}_3] P_\phi(k_2) P_\phi(k_3), \quad (2.76)$$

which can also be easily converted to the matter power spectrum expression.

Similarly

$$\begin{aligned} I_b &= \frac{8\beta^3 f_{NL}}{c^2 (2\pi)^6} \int \prod_{i=1}^3 d^3 \mathbf{k}_i \mathbf{k}_i^2 T(\mathbf{k}_i) W_{\mathbf{k}_i} W_{\mathbf{k}_2 + \mathbf{k}_3} F_2[\mathbf{k}_2, \mathbf{k}_3] \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) P_\phi(k_1) P_\phi(k_3) \\ &= \frac{8\beta^3 f_{NL}}{c^2 (2\pi)^6} \int d^3 \mathbf{k}_2 d^3 \mathbf{k}_3 \mathbf{k}_2^2 \mathbf{k}_3^2 (\mathbf{k}_2 + \mathbf{k}_3)^2 T_{\mathbf{k}_2} T_{\mathbf{k}_3} T_{|\mathbf{k}_2 + \mathbf{k}_3|} W_{|\mathbf{k}_2 + \mathbf{k}_3|}^2 \\ &\quad \times F_2[\mathbf{k}_2, \mathbf{k}_3] P_\phi(|\mathbf{k}_2 + \mathbf{k}_3|) P_\phi(k_3), \end{aligned} \quad (2.77)$$

which is very similar to what was obtained for  $I_a$ . Both expressions can be numerically integrated after being converted to the form with the matter power spectrum. The skewness parameter including PNG contributions can then be written as

$$S_3^{NL} = \frac{\langle \delta_{\text{std.}}^3 \rangle + \langle \delta_{NL}^3 \rangle}{(\langle \sigma_{\text{std.}}^2 \rangle + \langle \sigma_{NL}^2 \rangle)^2}, \quad (2.78)$$

and an interesting check is the ratio between the skewness obtained with the contribution given by PNG in Equation 2.78 and the standard one:

$$\text{ratio} = \frac{S_3^{NL}}{S_3}, \quad (2.79)$$

shown in Figure 2.1 for different values of  $f_{NL}$ . Another option to account for the impossibility of simplifying  $I_a$  and  $I_b$  is to assume that the contributions given by the  $f_{NL}$  terms are much smaller with respect to the standard ones, since once squared they go as  $\propto f_{NL}^2/c^4$ . In this case we can Taylor-expand the above expression

$$\begin{aligned} S_3^{NL} &= \frac{\langle \delta^3 \rangle + \langle \delta_{NL}^3 \rangle}{\langle \sigma^2 \rangle^2 \left( 1 + \frac{\langle \sigma_{NL}^2 \rangle}{\langle \sigma^2 \rangle} \right)^2} \\ &\simeq \frac{\langle \delta^3 \rangle + \langle \delta_{NL}^3 \rangle}{\langle \sigma^2 \rangle^2} \times \left( 1 - 2 \frac{\langle \sigma_{NL}^2 \rangle}{\langle \sigma^2 \rangle} + 3 \left( \frac{\langle \sigma_{NL}^2 \rangle}{\langle \sigma^2 \rangle} \right)^2 \right). \end{aligned} \quad (2.80)$$

In this way it is possible to consider only the first term of Equation 2.80, ignoring the non-linear contribution to the variance.

Comparing the theoretical predictions shown in Figure 2.1 with the work done by Scoccimarro et al. (2004) and Mao (2014) we find good agreement especially

when looking at the enhancement of the skewness signal in the case of PNG parametrised by  $f_{NL} = 100$ .

Currently, the best constraints on primordial non-Gaussianity are given by CMB experiments and in particular *Planck* (Planck Collaboration et al., 2016b) which constrained the local type of PNG up to  $f_{NL}^{\text{loc}} = 0.8 \pm 5.0$  combining both temperature and polarization data. The ratio in Equation 2.79 for  $f_{NL} = 1$  differs from unity by less than 0.5% up to the maximum scales shown in Figure 2.1.

## 2.5 $S_3$ from data

In this section are presented the cosmological parameter constraints derived from the measurement of the 2pt and 3pt auto-correlation functions on the SDSS DR11 BOSS CMASS north sample data. In order to obtain 1 and 2D confidence intervals, the galaxy mocks produced by Manera et al. (2013); Manera et al. (2015) have been used to estimate the covariance matrix.

The main reason for these measurements was originally to find experimental evidence of the result presented by Juszkiewicz (2013) relative to possibility of observing a BAO signal in the skewness. Indeed even if the signal would have been much smaller and less peaked than in the 2pt correlation function, from the theoretical point of view it would have been almost completely independent from the galaxy bias parametrisation.

In order to check the presence of BAO, Juszkiewicz takes the ratio of an  $S_3$  with a physical power spectrum to an  $S_3^{\text{smooth}}$  obtained by using an analytic expression for a no-wiggle power spectrum (Eisenstein and Hu, 1999). In this way for a radius of  $\sim 55-60$  Mpc/ $h$  of the top-hat window function, the ratio plot has the center of an oscillation around unity with an amplitude of approximately the 3% as shown in Figure 2.2 from Juszkiewicz (2013).

In our project, we planned to use a slightly different function in order to check for the presence of BAO in three-point cumulative statistics. In particular it has been considered an approximation of the derivative of  $S_3$ ,  $\Delta S_3$  which given a set of data points for different radii of the top hat window function  $S_3(R_i)$  it is defined using the Richardson's extrapolation formula as

$$\Delta S_3^i = \frac{-S_3^{i+2} + 8S_3^{i+1} - 8S_3^{i-1} + S_3^{i-2}}{\Delta R}, \quad (2.81)$$

where  $\Delta R = R_{i+1} - R_i$  and it is constant. This quantity has been introduced for two main reasons. First, it is experimentally easier to observe and measure

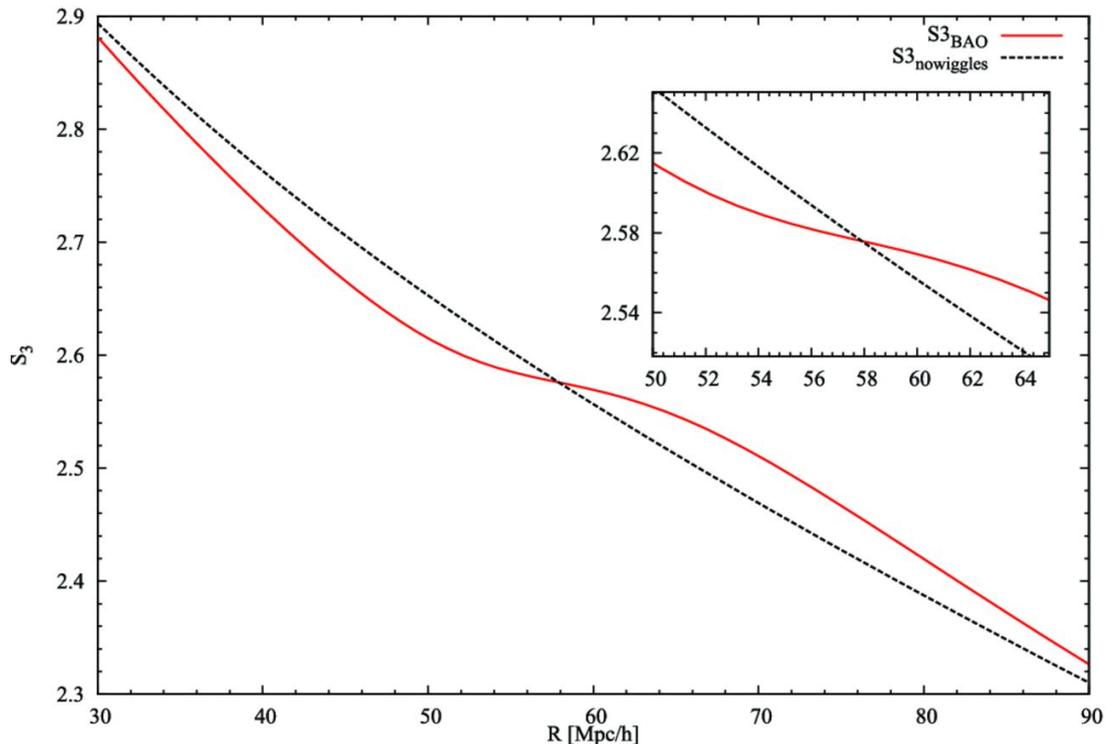


Figure 2.2: Plot from [Juszkiewicz \(2013\)](#) showing the skewness BAO feature. The red solid line represents the volume-averaged skewness as function of top-hat scale. It shows the characteristic skewness BAO feature, a mild shoulder around the characteristic crossing scale. For comparison, the dotted line represents the skewness for a "no-wiggle" power spectrum. The inset zooms in on the shoulder at the crossing scale.

a peak as the one expected in  $\Delta S_3$  than the crossing point expected in the  $S_3$  ratio.

Secondly, since it is not possible to directly probe the matter field distribution. We measure the skewness of the galaxy field  $S_3^g$ , from the mocks and the BOSS data, assuming that the galaxies act as tracers of the matter field. The bias relation linking  $S_3^g$  with  $S_3^m$  ([Fry and Gaztanaga, 1993](#)) is

$$S_3^g = \frac{1}{b} (S_3^m + 3c_2), \quad (2.82)$$

where  $b$  and  $c_2$  are the two relevant bias parameters.

From this definition and from Equation 2.82 it is immediate to see the independence of  $\Delta S_3$  from the bias parameter  $c_2$ . A total independence from the galaxy-matter bias is then gained by studying the ratio between  $\Delta S_3$  obtained from the data and the one obtained by smoothing the data. This is analogous to the  $S_3$  ratio used by [Juszkiewicz \(2013\)](#) in his paper, with the advantage that

in this case there is theoretically no dependence from the bias

$$R_{\Delta S_3} = \frac{\Delta S_3^m}{\Delta S_3^{m,sm.}} = \frac{\Delta S_3^g}{\Delta S_3^{g,sm.}}, \quad (2.83)$$

where "m", "g" and "sm." stand for "matter", "galaxies" and "smoothed", respectively.

Unfortunately, as soon as we measured  $\delta^2$  and  $\delta^3$  from both data and galaxy mocks, it was evident that the cosmic variance error component was too large for a detection of the BAO feature to be possible. We used the measurements of the 2pt and 3pt correlation functions joint data vector to constrain model parameters like the linear galaxy bias  $b_1$ , the normalisation of the dark matter oscillation amplitude  $\sigma_8$  and the primordial non Gaussianities parameter  $f_{NL}$ .

### 2.5.1 BOSS DR11 data and mocks

In this work we use the DR11 CMASS north sample of the SDSS Baryon Oscillation Spectroscopic Survey, containing 579,461 observed galaxies covering a total area of 6,769 deg<sup>2</sup>. The redshift range is  $0.43 < z < 0.7$ .

In order to compute the covariance matrix for our analysis we have used 600 galaxy mocks and 600 random mocks for DR11 created by [Manera et al. \(2013\)](#). The mocks for the galaxies contained also the "true" redshift of the galaxies together with the observed one so that it has been possible to study the effect of the redshift-space distortions of the used statistics. The galaxy mocks have been constructed by placing dark matter halos in a 2LPT field and populating them with galaxies. The cosmology of the mocks is compatible with WMAP5-7:  $\Omega_m = 0.274$ ,  $\Omega_\Lambda = 0.726$ ,  $\sigma_8 = 0.8$ ,  $n_s = 0.95$ ,  $h = 0.7$ ,  $\Omega_b \Delta h^2 = 0.0224$ . The galaxy mocks also have a weighting system that corrects for the incompleteness of a real data-set:  $w_{boss}$  for galaxies/randoms reduced by completeness  $C_{boss}$ ,  $w_{cp}$  close pairs weight,  $w_{red}$  redshift failure weight. The system consists in using only the galaxies with all three weights  $> 0$  and applying to each one the weight:  $w_{tot} = w_{cp} + w_{red} - 1$ . Therefore the total weight for each galaxy is increased if the closest galaxy had a redshift failure or if the redshift was not measured since both galaxies formed a close pair. The mocks have a survey geometry corresponding to the BOSS DR11.

### 2.5.2 Statistical estimators

We measured separately  $\langle \delta_g^3 \rangle$  and  $\langle \delta_g^2 \rangle$  by randomly arranging an arbitrary number  $n_{sph.}$  of spheres inside the survey volume on a regular three-dimensional grid, and then counting the number of galaxies inside each sphere of radius  $R$ . The same is done for a contrast field catalog with approximately one hundred times

the density, distributing the points using a Poisson process. This allows us to define a  $\delta_g(R)$  for the galaxy field as a function of the sphere's radius:

$$\delta_g(R) = \frac{n_g(R)}{n_r(R)} F - 1 \quad \text{and} \quad F = \frac{n_r^{\text{tot}}}{n_g}, \quad (2.84)$$

where  $F$  is the renormalisation factor. From now on unless specified, all quantities refer to the galaxy field. Since we are dealing with a discrete distribution it is necessary to use estimators for the continuous galaxy distribution field derived from the discrete counting. What it is actually measured from counting  $n_g(R)$  and  $n_r(R)$  inside the a sphere of radius  $R$  are the quantities:

$$\begin{aligned} k_2 &= \frac{1}{n_{\text{sph.}}} \sum \left( \frac{n_g(R)}{n_r(R)} F - 1 \right)^2 \\ k_3 &= \frac{1}{n_{\text{sph.}}} \sum \left( \frac{n_g(R)}{n_r(R)} F - 1 \right)^3. \end{aligned} \quad (2.85)$$

In order to relate these quantities with the continuous cumulative two- and three-point functions for the continuous galaxy field  $\langle \delta^2 \rangle$  and  $\langle \delta^3 \rangle$  we need to average the quantities in Equation 2.85 both over the galaxy Poisson sampling and the cosmic variance. Averaging over cosmic variance means to average over different realizations, in our case this is done by averaging over different mocks. For the two-point expression with same redshift but different sampling densities (the density of the points, galaxies or randoms, varies with the redshift  $z$ ) we have that the second moment of the discrete density distribution  $\langle k_2 \rangle$  is given by

$$\langle k_2 \rangle = \langle \delta^2 \rangle + \frac{1}{n_{\text{sph.}}} \sum \frac{1}{n_r(\mathbf{x})}, \quad (2.86)$$

where  $n_r(\mathbf{x})$  is the number of random points (not clustered) inside a sphere of radius  $R$  at a position  $\mathbf{x}$  inside the survey. Analogously the third moment of discrete density distribution  $\langle k_3 \rangle$  is related to the continuous one by

$$\langle k_3 \rangle = \langle \delta^3 \rangle + \frac{\langle \delta^2 \rangle}{n_{\text{sph.}}} \sum \frac{3}{n_r(\mathbf{x})} + \frac{1}{n_{\text{sph.}}} \sum \frac{1}{n_r(\mathbf{x})^2}. \quad (2.87)$$

Both expressions agree with Gaztanaga (1994). Since by construction the galaxy number density depends only on the redshift, in all the previous expressions can be rewritten with  $n_r(\mathbf{x}) = n_r(z)$ .

## 2.6 Treatment of noise sources

The fact that the survey has a finite volume, which has also masked zones, is a source of noise in the measurement of the cumulative moments. Indeed one can ask what happens when part of the sphere used to estimate the moments of the distribution falls outside the survey volume. Since the perturbation to the average density  $\delta$  is computed by taking the ratio of the counted number of galaxies with the counted number of randoms inside the sphere, its statistic is independent of whether the sphere partially falls out of the survey. The problem is to associate this statistic to an effective sphere radius  $R_{eff}$  relative to a sphere which would have the same statistic if its volume would be completely inside the survey.

In order to avoid this complicated problem, in this work the measurements have been considered in the analysis only up to the maximum radius for which spheres still fully lie inside the survey volume. Hence each sphere contributed to the statistics up to the scale corresponding to their own maximum radius. In this way all the volume of the survey is probed even if not uniformly for all the scales: for larger window function radii  $R$ , fewer spheres were included in the survey volume.

## 2.7 Redshift-space distortions effect

The galaxy mocks used for the analysis have both the true and the observed redshift for the galaxies. It has therefore been possible to check its effect on  $\delta^2$ ,  $\delta^3$  and  $S_3$ . The effect of redshift-space distortions in galaxy clustering analysis was studied using Lagrangian perturbation theory by [Hivon et al. \(1995\)](#). In [Figure 2.3](#) we show a plot of the observed effect in our measurements. In order to empirically check for the statistical effect of redshift-space distortions on the measurements, we run the pipeline on each of the 600 galaxy mocks once using the "observed"  $z$  including RSD and once using the "true"  $z$ . In [Figure 2.3](#) it has been plotted the means of 2pt and 3pt auto-correlation functions,  $\langle\delta^2(R)\rangle$  and  $\langle\delta^3(R)\rangle$ , together with the skewness  $S_3(R)$  for both the "true" and "obs" redshift cases with the shaded regions indicating the errorbars given by the mocks.

## 2.8 RSD-corrected expressions for $\langle\delta^2\rangle$ and $\langle\delta^3\rangle$

[Figure 2.3](#) shows that there is a substantial difference between configuration (without RSD) and redshift-space results ("TRUE Z" and "OBS. Z" in the Figure, respectively), The parameter constraints obtained by using the real data would

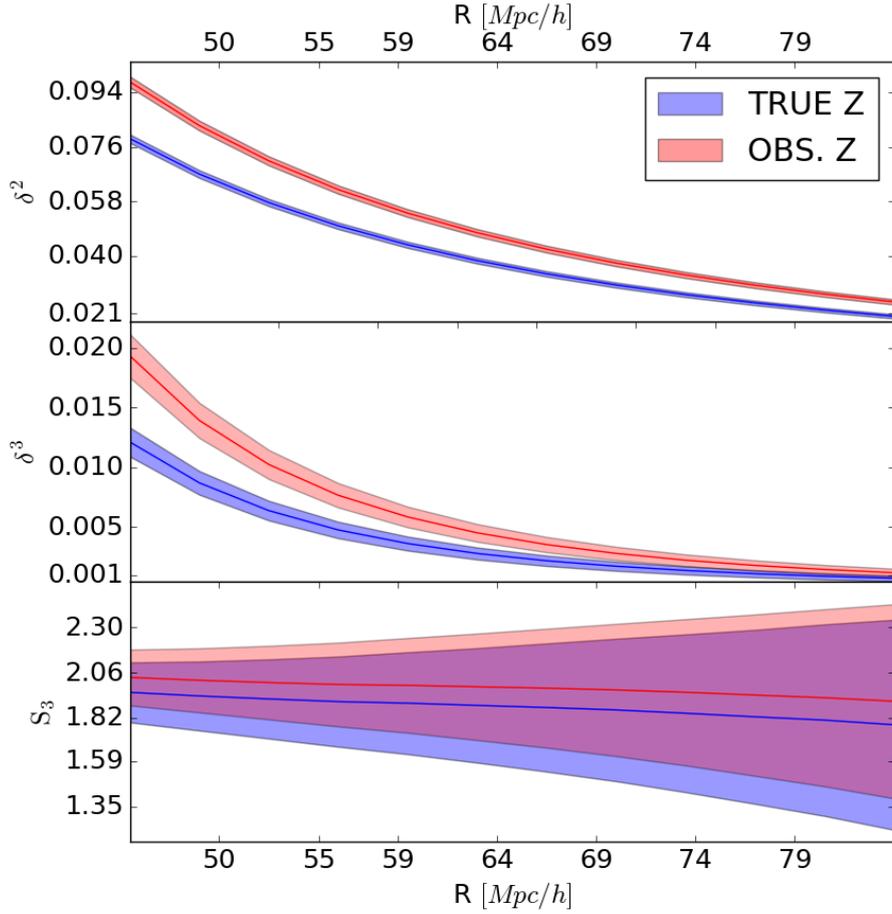


Figure 2.3: From top to bottom:  $\langle\delta^2(R)\rangle$ ,  $\langle\delta^3(R)\rangle$  and  $S_3$  measured from DR11 mocks [Manera et al., 2015](#) once using the 'true' redshift (without RSD) of the galaxies (blue) and the 'observed' one (with RSD, in red). The shaded area corresponds to the error-bars given by the sample covariance matrices.

result biased if RSD were not taken into account. Using the relations presented in [Heavens et al. \(1998\)](#) to relate the galaxy perturbation density function in redshift-space to the matter one in real space, it is possible to compute the corrected theoretical model for the two- and three-point correlation functions. The perturbation theory derived kernels corrected for RSD used below have been previously described in Section 1.5.

The statistics measured from data depend on the over-density of the galaxy field in redshift-space  $\langle\delta_s^g\rangle$ . Repeating Equation 1.65,  $\langle\delta_s^g\rangle$  is related to the linear perturbation variable for the matter density field  $\langle\delta_m\rangle$  by

$$\begin{aligned}
 \delta_{s\mathbf{k}}^g &= F_s^{(1)}(\mathbf{k}) \delta_{\mathbf{k}} + \frac{1}{(2\pi)^3} \int d^3 \mathbf{k}_1 d^3 \mathbf{k}_2 \delta_D(\mathbf{k} - \mathbf{k}_2 - \mathbf{k}_1) F_s^{(2)}(\mathbf{k}_1, \mathbf{k}_2) \delta_{\mathbf{k}_1} \delta_{\mathbf{k}_2} \\
 &\quad + \frac{1}{(2\pi)^6} \int d^3 \mathbf{k}_1 d^3 \mathbf{k}_2 d^3 \mathbf{k}_3 \delta_D(\mathbf{k} - \mathbf{k}_3 - \mathbf{k}_2 - \mathbf{k}_1) \\
 &\quad \times F_s^{(3)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) \delta_{\mathbf{k}_1} \delta_{\mathbf{k}_2} \delta_{\mathbf{k}_3} .
 \end{aligned} \tag{2.88}$$

Therefore starting with the 3pt auto-correlation function

$$\begin{aligned}
 \langle \delta_s^g(\mathbf{x}_0)^3 \rangle &= \int \prod_{i=1}^3 d^3 \mathbf{x}_i W_{\mathbf{x}-\mathbf{x}_0} \frac{d^3 \mathbf{k}_i}{(2\pi)^3} e^{-i\mathbf{k}_i \mathbf{x}_i} \langle \delta_s^g(\mathbf{k}_1) \delta_s^g(\mathbf{k}_2) \delta_s^g(\mathbf{k}_3) \rangle \\
 &= 3 \int \prod_{i=1}^3 \frac{d^3 \mathbf{k}_i}{(2\pi)^3} W_{\mathbf{k}_i} \int \frac{d^3 \mathbf{k}_4 d^3 \mathbf{k}_5}{(2\pi)^3} F_s^{(1)}(\mathbf{k}_1) F_s^{(1)}(\mathbf{k}_2) \\
 &\quad \times F_s^{(2)}[\mathbf{k}_1, \mathbf{k}_2] \delta_D(\mathbf{k}_3 - \mathbf{k}_4 - \mathbf{k}_5) \langle \delta_{\mathbf{k}_1}^m \delta_{\mathbf{k}_2}^m \delta_{\mathbf{k}_3}^m \delta_{\mathbf{k}_4}^m \rangle \\
 &= \frac{6}{(2\pi)^6} \int d^3 \mathbf{k}_1 d^3 \mathbf{k}_2 W_{\mathbf{k}_1} W_{\mathbf{k}_2} W_{|\mathbf{k}_1+\mathbf{k}_2|} \\
 &\quad \times F_s^{(1)}(\mathbf{k}_1) F_s^{(1)}(\mathbf{k}_2) F_s^{(2)}[\mathbf{k}_1, \mathbf{k}_2] P(k_1) P(k_2) ,
 \end{aligned} \tag{2.89}$$

where  $\langle \delta_m \rangle$  is a Gaussian variable. The first term of the expansion in Equation 2.89 vanishes giving

$$\langle \delta_s^g(\mathbf{k}_1) \delta_s^g(\mathbf{k}_2) \delta_s^g(\mathbf{k}_3) \rangle = F_s^{(1)3} \langle \delta_{\mathbf{k}_1}^m \delta_{\mathbf{k}_2}^m \delta_{\mathbf{k}_3}^m \rangle + 3F_s^{(1)2} F_s^{(2)} \langle \delta_{\mathbf{k}_1}^m \delta_{\mathbf{k}_2}^m \delta_{\mathbf{k}_3}^m \delta_{\mathbf{k}_4}^m \rangle + O(\delta^6) . \tag{2.90}$$

For the variance, proceeding in the same way it is possible to obtain:

$$\begin{aligned}
 \langle \delta_s^g(\mathbf{x}_0)^2 \rangle &= \\
 &= \prod_{i=2}^2 \int \frac{d^3 \mathbf{k}_i}{(2\pi)^3} W_{\mathbf{k}_i} \langle \delta_s^g(\mathbf{k}_1) \delta_s^g(\mathbf{k}_2) \rangle \\
 &= \prod_{i=2}^2 \int \frac{d^3 \mathbf{k}_i}{(2\pi)^3} W_{\mathbf{k}_i} \left\{ F_s^{(1)}(\mathbf{k}_1) F_s^{(1)}(\mathbf{k}_2) (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2) P(k_1) \right. \\
 &\quad + \frac{1}{(2\pi)^6} \int d^3 \mathbf{k}_a d^3 \mathbf{k}_b d^3 \mathbf{k}_c d^3 \mathbf{k}_d \delta_D(\mathbf{k}_1 - \mathbf{k}_a - \mathbf{k}_b) \delta_D(\mathbf{k}_2 - \mathbf{k}_c - \mathbf{k}_d) \\
 &\quad \times F_s^{(2)}[\mathbf{k}_a, \mathbf{k}_b] F_s^{(2)}[\mathbf{k}_c, \mathbf{k}_d] \langle \delta_{\mathbf{k}_a}^m \delta_{\mathbf{k}_b}^m \delta_{\mathbf{k}_c}^m \delta_{\mathbf{k}_d}^m \rangle \\
 &\quad \left. + \frac{2}{(2\pi)^6} F_s^{(1)}(\mathbf{k}_1) \int d^3 \mathbf{k}_e d^3 \mathbf{k}_f d^3 \mathbf{k}_g \delta_D(\mathbf{k}_2 - \mathbf{k}_e - \mathbf{k}_f - \mathbf{k}_g) F_s^{(3)}[\mathbf{k}_e, \mathbf{k}_f, \mathbf{k}_g] \langle \delta_{\mathbf{k}_1}^m \delta_{\mathbf{k}_e}^m \delta_{\mathbf{k}_f}^m \delta_{\mathbf{k}_g}^m \rangle \right\} \\
 &= \prod_{i=2}^2 \int \frac{d^3 \mathbf{k}_i}{(2\pi)^3} W_{\mathbf{k}_i} \left\{ F_s^{(1)}(\mathbf{k}_1) F_s^{(1)}(\mathbf{k}_2) (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2) P(k_1) \right. \\
 &\quad + \frac{2}{(2\pi)^6} \int d^3 \mathbf{k}_a d^3 \mathbf{k}_b F_s^{(2)}[\mathbf{k}_a, \mathbf{k}_b]^2 \delta_D(\mathbf{k}_1 - \mathbf{k}_a - \mathbf{k}_b) \delta_D(\mathbf{k}_2 - \mathbf{k}_a - \mathbf{k}_b) (2\pi)^6 P(k_a) P(k_b) \\
 &\quad + \frac{6}{(2\pi)^6} F_s^{(1)}(\mathbf{k}_1) \int d^3 \mathbf{k}_e d^3 \mathbf{k}_f d^3 \mathbf{k}_g \delta_D(\mathbf{k}_2 - \mathbf{k}_e - \mathbf{k}_f - \mathbf{k}_g) F_s^{(3)}[\mathbf{k}_e, \mathbf{k}_f, \mathbf{k}_g] \\
 &\quad \left. \times (2\pi)^6 \delta_D(\mathbf{k}_1 + \mathbf{k}_e) \delta_D(\mathbf{k}_g + \mathbf{k}_f) P(k_1) P(k_f) \right\} \\
 &= \int \frac{d^3 \mathbf{k}_1}{(2\pi)^3} W_{\mathbf{k}_1}^2 F_s^{(1)^2}(\mathbf{k}_1) P(k_1) + \frac{2}{(2\pi)^6} \int d^3 \mathbf{k}_1 d^3 \mathbf{k}_2 W_{|\mathbf{k}_1 + \mathbf{k}_2|}^2 F_s^{(2)^2}[\mathbf{k}_1, \mathbf{k}_2] P(k_1) P(k_2) \\
 &\quad + \frac{6}{(2\pi)^6} \int d^3 \mathbf{k}_1 d^3 \mathbf{k}_2 W_{\mathbf{k}_1}^2 F_s^{(1)}(\mathbf{k}_1) F_s^{(3)}[\mathbf{k}_1, \mathbf{k}_2, -\mathbf{k}_2] P(k_1) P(k_2). \tag{2.91}
 \end{aligned}$$

## 2.9 RSD-corrected expressions for $\langle \delta^2 \rangle$ and $\langle \delta^3 \rangle$ : $\mathbf{f}_{\text{NL}}$ terms

It is also possible to compute the  $f_{\text{NL}}$  contributions to the redshift-space galaxy-field quantities. Similar expressions have been derived in the literature (Scoccimarro et al., 2004) but the ones in Section 2.9 are original for this work. Starting for example from the 3pt auto-correlation function, the first term of the expansion in Equation 2.90 in the case of primordial non-Gaussianities does not vanish and using the result obtained in Equation 2.64:

$$\begin{aligned}
 \langle \delta_1(\mathbf{x}_0) \delta_1(\mathbf{x}_0) \delta_1(\mathbf{x}_0) \rangle &= \\
 &= \frac{6\beta^3 f_{\text{NL}}}{c^2} \int \prod_{i=1}^3 \frac{d^3 \mathbf{k}_i}{(2\pi)^3} \mathbf{k}_i^2 \mathbb{T}(\mathbf{k}_i) \mathbb{W}(\mathbf{k}_i, R) \delta_D(\mathbf{k}_3 + \mathbf{k}_2 - \mathbf{k}_1) \mathbb{P}_{\mathbf{k}_2}^\phi \mathbb{P}_{\mathbf{k}_3}^\phi (2\pi)^3 \\
 &= \frac{6\beta^3 f_{\text{NL}}}{c^2 (2\pi)^6} \int d^3 \mathbf{k}_2 d^3 \mathbf{k}_3 \mathbf{k}_2^2 \mathbf{k}_3^2 (\mathbf{k}_2 + \mathbf{k}_3)^2 \mathbb{T}_{\mathbf{k}_2} \mathbb{T}_{\mathbf{k}_3} \mathbb{T}_{\mathbf{k}_3 + \mathbf{k}_2} \mathbb{W}_{\mathbf{k}_2} \mathbb{W}_{\mathbf{k}_3} \mathbb{W}_{\mathbf{k}_3 + \mathbf{k}_2} \mathbb{P}_{\mathbf{k}_2}^\phi \mathbb{P}_{\mathbf{k}_3}^\phi.
 \end{aligned} \tag{2.92}$$

The 3pt expression in redshift-space for the galaxy field is then

$$\begin{aligned}
 \langle \delta_s^{\text{g}}(\mathbf{x}_0)^3 \rangle_{f_{\text{NL}}} &= \frac{6\beta^3 f_{\text{NL}}}{c^2} \int \prod_{i=1}^3 \frac{d^3 \mathbf{k}_i}{(2\pi)^3} \mathbb{W}_{\mathbf{k}_i} F_{s,i}^{(1)} \mathbf{k}_i^2 \mathbb{T}_{\mathbf{k}_i} \delta_D(\mathbf{k}_3 + \mathbf{k}_2 - \mathbf{k}_1) \mathbb{P}_{\mathbf{k}_2}^\phi \mathbb{P}_{\mathbf{k}_3}^\phi (2\pi)^3 \\
 &= \frac{6\beta^3 f_{\text{NL}}}{c^2} \int \frac{d^3 \mathbf{k}_2 d^3 \mathbf{k}_3}{(2\pi)^6} \mathbf{k}_2^2 \mathbf{k}_3^2 (\mathbf{k}_2 + \mathbf{k}_3)^2 \mathbb{T}_{\mathbf{k}_2} \mathbb{T}_{\mathbf{k}_3} \mathbb{T}_{\mathbf{k}_3 + \mathbf{k}_2} \mathbb{W}_{\mathbf{k}_2} \mathbb{W}_{\mathbf{k}_3} \mathbb{W}_{\mathbf{k}_3 + \mathbf{k}_2} \\
 &\quad \times F_{s,\mathbf{k}_2}^{(1)} F_{s,\mathbf{k}_3}^{(1)} F_{s,\mathbf{k}_3 + \mathbf{k}_2}^{(1)} \mathbb{P}_{\mathbf{k}_2}^\phi \mathbb{P}_{\mathbf{k}_3}^\phi,
 \end{aligned} \tag{2.93}$$

which can be easily converted to the linear matter power spectrum using the relation  $\mathbb{P}^\phi(k) = \mathbb{P}^{\text{m}}(k) \beta^{-2} k^{-4} \mathbb{T}_k^{-2}$ . Using cylindrical coordinates along the line of sight and decomposing the wave vectors in terms of parallel and perpendicular components to the line of sight with the addition of the azimuthal angle, the six-dimensional integral in Equation 2.93 can be reduced to a five-dimensional one. In particular we use as coordinates the difference and the sum between the two azimuthal angles of the two wave-vectors (the sum is integrated out):

$$\begin{aligned}
 \langle \delta_s^{\text{g}}(\mathbf{x}_0)^3 \rangle_{f_{\text{NL}}} &= \frac{6f_{\text{NL}}}{\beta c^2 (2\pi)^5} \int dk_2^\perp dk_2^\parallel dk_3^\perp dk_3^\parallel d\theta_d k_2^\perp k_3^\perp \mathbb{T}_{\mathbf{k}_2}^{-1} \mathbb{T}_{\mathbf{k}_3}^{-1} \mathbb{T}_{\mathbf{k}_3 + \mathbf{k}_2} \mathbb{W}_{\mathbf{k}_2} \mathbb{W}_{\mathbf{k}_3} \mathbb{W}_{\mathbf{k}_3 + \mathbf{k}_2} \\
 &\quad \times k_2^{-2} k_3^{-2} (\mathbf{k}_2 + \mathbf{k}_3)^2 F_{s,\mathbf{k}_2}^{(1)} F_{s,\mathbf{k}_3}^{(1)} F_{s,\mathbf{k}_3 + \mathbf{k}_2}^{(1)} \mathbb{P}_{\mathbf{k}_2}^{\text{m}} \mathbb{P}_{\mathbf{k}_3}^{\text{m}}.
 \end{aligned} \tag{2.94}$$

For the 2pt contributions, starting from Equation 2.91, the standard terms of the expansion of  $\langle \delta_s^{\text{g}2} \rangle$  do not give any contribution at order  $f_{\text{NL}}$ . Only the term that, without primordial non-Gaussianities would be zero due to linearity, gives a contribution

$$\begin{aligned}
 \langle \delta_s^{\text{g}}(\mathbf{x}_0)^2 \rangle_{f_{\text{NL}}} &= \\
 &= \prod_{i=2}^2 \int \frac{d^3 \mathbf{k}_i}{(2\pi)^3} \mathbb{W}_{\mathbf{k}_i} \langle \delta_s^{\text{g}}(\mathbf{k}_1) \delta_s^{\text{g}}(\mathbf{k}_2) \rangle \\
 &= \prod_{i=2}^2 \int \frac{d^3 \mathbf{k}_i}{(2\pi)^3} \mathbb{W}_{\mathbf{k}_i} \left\{ 2F_{s,\mathbf{k}_1}^{(1)} \frac{1}{(2\pi)^3} \int d^3 \mathbf{k}_3 d^3 \mathbf{k}_4 \delta_D(\mathbf{k}_2 - \mathbf{k}_3 - \mathbf{k}_4) F_s^{(2)}[\mathbf{k}_3, \mathbf{k}_4] \langle \delta_{\mathbf{k}_1}^{\text{m}} \delta_{\mathbf{k}_3}^{\text{m}} \delta_{\mathbf{k}_4}^{\text{m}} \rangle \right\}.
 \end{aligned} \tag{2.95}$$

Equation 2.95 gives the term already computed in the matter 2pt case corrected for RSD

$$\begin{aligned}
 I_a &= \frac{4\beta^3 f_{NL}}{c^2(2\pi)^6} \int d^3\mathbf{k}_2 d^3\mathbf{k}_3 \mathbf{k}_2^2 \mathbf{k}_3^2 (\mathbf{k}_2 + \mathbf{k}_3)^2 T_{\mathbf{k}_2} T_{\mathbf{k}_3} T_{|\mathbf{k}_2+\mathbf{k}_3|} W_{|\mathbf{k}_2+\mathbf{k}_3|}^2 \\
 &\quad \times F_s^{(1)}(\mathbf{k}_2 + \mathbf{k}_3) F_s^{(2)}[\mathbf{k}_2, \mathbf{k}_3] P_\phi(k_2) P_\phi(k_3) \\
 I_b &= \frac{8\beta^3 f_{NL}}{c^2(2\pi)^6} \int d^3\mathbf{k}_2 d^3\mathbf{k}_3 \mathbf{k}_2^2 \mathbf{k}_3^2 (\mathbf{k}_2 + \mathbf{k}_3)^2 T_{\mathbf{k}_2} T_{\mathbf{k}_3} T_{|\mathbf{k}_2+\mathbf{k}_3|} W_{|\mathbf{k}_2+\mathbf{k}_3|}^2 \\
 &\quad \times F_s^{(1)}(\mathbf{k}_2 + \mathbf{k}_3) F_s^{(2)}[\mathbf{k}_2, \mathbf{k}_3] P_\phi(|\mathbf{k}_2 + \mathbf{k}_3|) P_\phi(k_3). \tag{2.96}
 \end{aligned}$$

Finally we can convert them into the matter power spectrum form

$$\begin{aligned}
 I_a &= \frac{4f_{NL}}{\beta c^2(2\pi)^6} \int d^3\mathbf{k}_2 d^3\mathbf{k}_3 \mathbf{k}_2^{-2} \mathbf{k}_3^{-2} (\mathbf{k}_2 + \mathbf{k}_3)^2 T_{\mathbf{k}_2}^{-1} T_{\mathbf{k}_3}^{-1} T_{|\mathbf{k}_2+\mathbf{k}_3|} W_{|\mathbf{k}_2+\mathbf{k}_3|}^2 \\
 &\quad \times F_s^{(1)}(\mathbf{k}_2 + \mathbf{k}_3) F_s^{(2)}[\mathbf{k}_2, \mathbf{k}_3] P_m(k_2) P_m(k_3) \\
 I_b &= \frac{8f_{NL}}{\beta c^2(2\pi)^6} \int d^3\mathbf{k}_2 d^3\mathbf{k}_3 \mathbf{k}_2^{-2} \mathbf{k}_3^{-2} (\mathbf{k}_2 + \mathbf{k}_3)^2 T_{\mathbf{k}_2}^{-1} T_{\mathbf{k}_3}^{-1} T_{|\mathbf{k}_2+\mathbf{k}_3|} W_{|\mathbf{k}_2+\mathbf{k}_3|}^2 \\
 &\quad \times F_s^{(1)}(\mathbf{k}_2 + \mathbf{k}_3) F_s^{(2)}[\mathbf{k}_2, \mathbf{k}_3] P_m(|\mathbf{k}_2 + \mathbf{k}_3|) P_m(k_3). \tag{2.97}
 \end{aligned}$$

Equation 2.97, as for the 3pt case, can be simplified and integrated using cylindrical coordinates.

## 2.10 Model parameters constraints from the joint data-vector [ $\langle\delta^2\rangle$ ; $\langle\delta^3\rangle$ ]

Bayesian statistics is often used to infer the model parameters best fit the data in Astrophysics. Bayesian statistics is based on Bayes's Theorem, which is

$$\mathcal{P}(\mathcal{M}|\mathcal{D}) = \frac{\mathcal{P}(\mathcal{D}|\mathcal{M})\mathcal{P}(\mathcal{M})}{\mathcal{P}(\mathcal{D})}. \tag{2.98}$$

Bayes's Theorem in Equation 2.98 states that the probability of the model given the data ( $\mathcal{P}(\mathcal{M}|\mathcal{D})$  also known as *posterior*) is equal to the product of the probability of the data given the model ( $\mathcal{P}(\mathcal{D}|\mathcal{M})$  also known as *likelihood*) and the probability (or better an assumed a-priori) of having that particular model ( $\mathcal{P}(\mathcal{M})$  also known as *prior*) divided by probability of having the data given all the possible models made available by the theory ( $\mathcal{P}(\mathcal{D})$  also known as *evidence*). If no selection between different models is involved, the evidence can then be safely treated as a normalisation constant and therefore ignored. For a very good review of these concepts see [Heavens \(2009\)](#).

In this Section we assumed that the likelihood distribution in our case could be approximated by a Gaussian multivariate distribution:

$$\mathcal{L} = \mathcal{P}(\mathcal{D}|\mathcal{M}) = \frac{1}{\sqrt{(2\pi)^n \det \text{Cov}}} \exp \left[ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \text{Cov}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right], \quad (2.99)$$

where  $n$  is the number of model parameter (dimension of the multivariate Gaussian distribution),  $\mathbf{x}$  is the data-vector model given by a certain set of parameters,  $\boldsymbol{\mu}$  is the data-vector measured from data and Cov is the covariance matrix which could be either computed analytically or numerically estimated from simulations.

In our case to estimate the covariance matrix we used galaxy mocks. In this case the covariance matrix element  $\text{Cov}_{\mu\nu}$  is given by

$$\text{Cov}_{\mu\nu} = \frac{1}{N_{\text{mocks}} - 1} \sum_{i=1}^{N_{\text{mocks}}} (x_{\mu}^i - \langle x_{\mu} \rangle) \times (x_{\nu}^i - \langle x_{\nu} \rangle), \quad (2.100)$$

where  $x_{\mu}^i$  is the  $\mu$ -element of the data-vector  $\mathbf{x}$  measured on the  $i$ -th galaxy mock.  $\langle \mathbf{x} \rangle$  is the average of the data vector  $\mathbf{x}$  measured over all the galaxy mocks.

For the joint data-vector [ $\delta^2$ ;  $\delta^3$ ] measured on the BOSS CMASS NGC sample we constrained three parameters of interest: the linear galaxy bias  $b_1$ , the normalisation of the dark matter oscillations amplitude  $\sigma_8$  and the primordial non-Gaussianity parameter  $f_{NL}$ .

In order to compute the linear matter power spectrum we fixed a fiducial  $\Lambda$ CDM-like cosmology with model parameters values close to the *Planck* cosmology: ( $h_0 = 0.677$ ,  $\Omega_m = 0.307$ ,  $\Omega_b = 0.048$ ,  $\Omega_{\Lambda} = 1 - \Omega_m$ ,  $n_s = 0.96$ ,  $A_s = 2.9 \times 10^{-9}$ ).

We used 5 millions spheres randomly distributed inside the survey 3D volume. The range of scales considered (radius of the spheres range) was between  $r_{\text{min.}} = 40 \text{ Mpc}/h$  and  $r_{\text{max.}} = 90 \text{ Mpc}/h$ . The upper limit was chosen such that cosmic variance was not dominant at the largest scales. We used flat uninformative priors for the cosmological parameters considered. The second-order bias parameter was fixed at  $b_2 = 0.50$ , a similar value to the one estimated in [Gil-Marín et al. \(2015\)](#). For the growth rate  $f$  we used the approximation valid in GR:  $f(z) \simeq \Omega_m^{\gamma}(z)$  where  $\Omega_m(z)$  is the mass density parameter function at a given redshift  $z$ , the growth index  $\gamma \simeq 0.55$  for a standard cosmology ([Peebles, 1980](#); [Lahav et al., 1991](#); [Linder, 2005](#)).

The best-fit parameter values obtained are in agreement for  $b_1$  and  $\sigma_8$  with those from the BOSS analysis ([Gil-Marín et al., 2015](#)). For PNG our constraints show that it is not possible to detect non Gaussianities parametrised by an  $f_{NL} < 100 - 200$  with the auto-correlation function. However the potential of

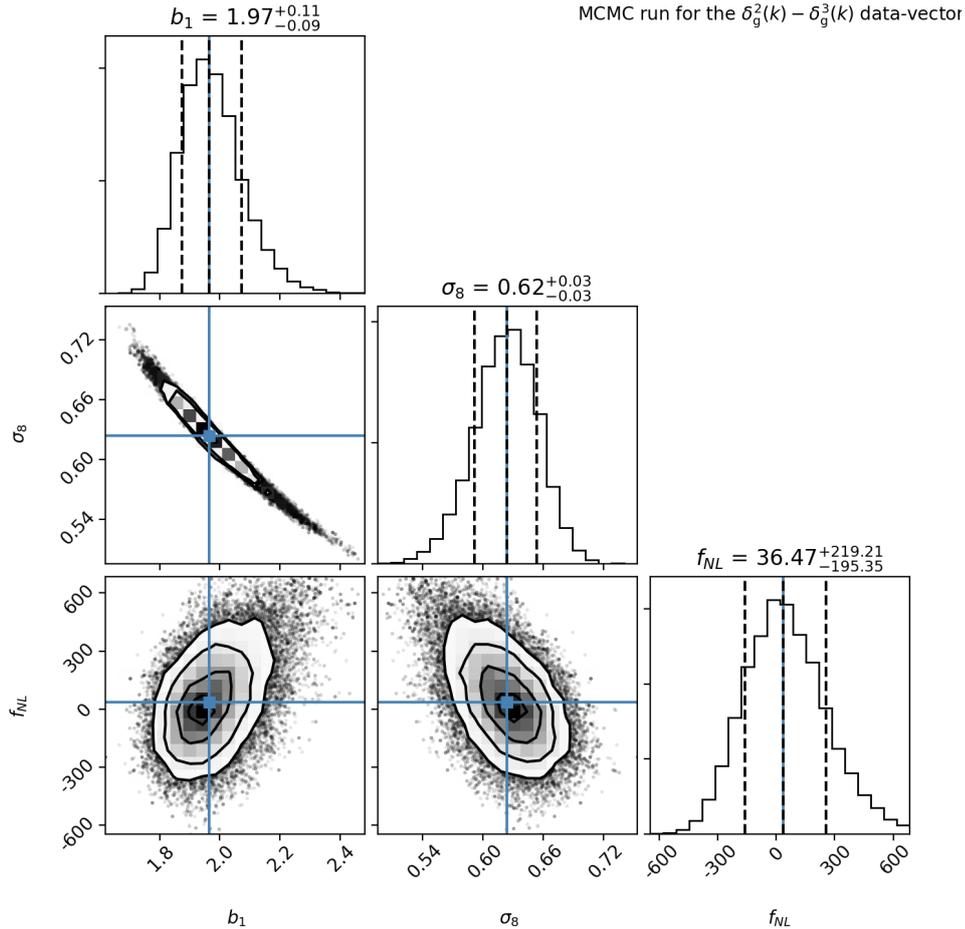


Figure 2.4: Joint data vector  $[\langle \delta^2(R) \rangle, \langle \delta^3(R) \rangle]$ . Constraints on the model parameters  $(b_1, \sigma_8, f_{NL})$  from the statistics measured from DR11 BOSS CMASS NGC sample. The vertical dashed lines show the 1D 68% confidence intervals; the precise values of intervals are reported on top of each 1D marginal posterior distribution together with the mode of the distribution.

these statistics could improve in future surveys like DESI, Euclid and PFS for which a much larger volume will be available, reducing the influence of cosmic variance especially at large scales. Our constraints on  $f_{NL}$  are similar to the ones derived by measuring the moments of the density field on simulations by [Mao \(2014\)](#).

## 2.11 Chapter recap

In this Chapter we derived analytic expressions for the skewness quantity and its constituents, the 2pt and 3pt autocorrelation functions. Terms including PNG were also computed and their magnitude has been shown in Figure 2.1. We measured the 2pt and 3pt auto-correlation functions with the intent of detecting BAO ([Eisenstein et al., 2005](#)) as forecasted by [Juszkiewicz \(2013\)](#). Unfortunately

cosmic variance dominates the signal at the relevant scales. Therefore it seems that in order to detect BAO the full 3pt correlation function must be used, as done by [Slepian et al. \(2017b\)](#).

We then used the measurements of the 2pt and 3pt auto-correlation function to constrain the combination of cosmological parameters  $(b_1, \sigma_8, f_{NL})$ . The 1 and 2D posterior distributions obtained are shown in [Figure 2.4](#).

# Appendix

## 2.A Transfer function from the matter power spectrum

In order to obtain the transfer function  $T(k)$  from the matter power spectrum  $P_m(k)$  we start from the Poisson Equation (2.55):

$$\nabla^2 \phi^{\text{l.t.}}(\vec{x}, a) = \frac{3 \Omega_m H_0^2}{2 a} \delta(\vec{x}, a) \quad (2.101)$$

Substituting into the last expression Equation (2.52) in order to pass to the primordial perturbation field  $\phi_{\text{prim}}$  and switching to Fourier space we get:

$$-k^2 \frac{9}{10} \frac{D_+}{a} T(k) \phi_k^{\text{prim}} = \frac{3 \Omega_m H_0}{2 a} \delta_k \quad (2.102)$$

which squared gives:

$$\frac{9}{25} k^4 D_+^2 T(k)^2 P_\phi(k) = (\Omega_m H_0)^2 P_m(k) \quad (2.103)$$

Knowing that the primordial power spectrum can be defined as  $P_\phi = A k^{(n_s-1)}$  where  $A = \text{const}$  is a normalization constant, the previous expression becomes:

$$\begin{aligned} \frac{9}{25} D_+^2 T(k)^2 A k^{(3+n_s)} &= (\Omega_m H_0)^2 P_m(k) \\ A T(k)^2 &= \frac{25 (\Omega_m H_0)^2}{9 D_+^2} k^{-(3+n_s)} P_m(k) \end{aligned} \quad (2.104)$$

Since we are only interested in the  $k$ 's dependence of the transfer function, it is possible to define a second normalization constant:

$$B = A \left[ \frac{5 \Omega_m H_0}{3 D_+} \right]^{-2} \quad (2.105)$$

Therefore:

$$T(k) = \sqrt{\frac{k^{-(3+n_s)} P_m(k)}{B}} \quad (2.106)$$

In order to fix  $B$  in the code, it has been imposed in the last expression  $T(k) = 1$  for the lowest  $k$  in the data-vector, since it corresponds to the largest scales available.

*“I have passed through fire and deep water, since we parted. I have forgotten much that I thought I knew, and learned again much that I had forgotten.”*

- J.R.R. Tolkien, The Lord of the Rings

# 3 | Maximal compression of the redshift space galaxy power spectrum and bispectrum

## 3.1 Abstract

We explore two methods of compressing the redshift space galaxy power spectrum and bispectrum with respect to a chosen set of cosmological parameters. Both methods involve reducing the dimension of the original data-vector ( e.g. 1000 elements ) to the number of cosmological parameters considered ( e.g. seven ) using the MOPED algorithm. In the first case, we run MCMC sampling on the compressed data-vector in order to recover the one-dimensional (1D) and two-dimensional (2D) posterior distributions. The second option, approximately 2000 times faster, works by orthogonalising the parameter space through diagonalisation of the Fisher information matrix before the compression, obtaining the posterior distributions without the need of MCMC sampling. Using these methods for future spectroscopic redshift surveys like DESI, EUCLID and PFS would drastically reduce the number of simulations needed to compute accurate covariance matrices with minimal loss of constraining power. We consider a redshift bin of a DESI-like experiment. Using the power spectrum combined with the bispectrum as a data-vector, both compression methods on average recover the 68% credible regions to within 0.7% and 2% of those resulting from standard MCMC sampling respectively. These confidence intervals are also smaller than the ones obtained using only the power spectrum by (81%, 80%, 82%) respectively for the bias parameter  $b_1$ , the growth rate  $f$  and the scalar amplitude parameter  $A_s$ .

## 3.2 Introduction

In recent years the number of available large data sets to be used for cosmological analysis has drastically increased (*PLANCK*, <sup>1</sup> Ade et al., 2014 ; Sloan Digital Sky Survey <sup>2</sup>, Eisenstein et al., 2011; DES, The Dark Energy Survey Collaboration, 2005 <sup>3</sup>) and will do even more so in the near future (DESI<sup>4</sup>, Levi et al., 2013; EUCLID <sup>5</sup>, Laureijs et al., 2011; PFS <sup>6</sup>, Takada et al., 2014). Improving current analysis techniques to extract as much information as possible from these catalogues has become highly relevant. Up to now most work has been done using two points statistics (2pt) like the two points correlation function or its Fourier transform, the power spectrum. However, gravity increases the level of non-linearity in the matter distribution field, creating non-Gaussian features which are not constrained by the sole use of 2pt statistics. Higher-order statistics like the three points (3pt) correlation function or its Fourier transform, the bispectrum, have already been studied in the past in order to capture the non-Gaussian part of the statistical information contained in the large scale structure galaxy field. The first measurements of the 3pt correlation function and the bispectrum on a galaxy catalogue were performed by Peebles and Groth (1975), Groth and Peebles (1977) and Fry and Seldner (1982). Fry (1984) studied the relation between the cosmological and bias parameters, modelling the relation between luminous and dark matter, which affects the amplitude and the shape of the bispectrum. Integrated 3pt statistics like the skewness were introduced in order to reduce the complexity and number of modelling parameters (Peebles, 1980; Fry and Scherrer, 1994; Bernardeau, 1994; Juszkiewicz, 2013). The modelling of redshift-space distortions into the 3pt statistics was later introduced and studied by Matarrese et al. (1997), Verde et al. (1998), Heavens et al. (1998), Scoccimarro et al. (1998), Scoccimarro (2000). Different 3pt statistics have also been proposed as useful tools to quantify deviations from GR (Borisov and Jain, 2009; Bernardeau and Brax, 2011) and to measure primordial non-Gaussianities (Fry and Scherrer, 1994; Gangui et al., 1994; Verde et al., 2000; Liguori et al., 2010; Tellarini et al., 2016). In the last twenty years, these statistics have received increasing attention especially for the benefit of lifting degeneracies present in 2pt statistics between cosmological and nuisance parameters in datasets produced by spectroscopic redshift surveys like BOSS. The most recent measurement on this

---

<sup>1</sup><http://sci.esa.int/planck/>

<sup>2</sup><http://www.sdss3.org/surveys/boss.php>

<sup>3</sup><https://www.darkenergysurvey.org>

<sup>4</sup><http://desi.lbl.gov>

<sup>5</sup><http://sci.esa.int/euclid/>

<sup>6</sup><http://pfs.ipmu.jp>

dataset for the bispectrum was made by [Gil-Marín et al. \(2017\)](#) and for the 3pt correlation function by [Slepian et al. \(2015\)](#).

When studying the power spectrum and the bispectrum, it is necessary to subdivide the range of possible  $\mathbf{k}$ -vectors in Fourier space into bins. By doing this, one defines the number of possible elements for both the power spectrum and bispectrum data-vectors. While in Fourier space the number of possible elements for the power spectrum data-vector is the same as the number of bins, for the bispectrum it corresponds to the number of triangles that can be formed by triplets of the available  $\mathbf{k}$ -vectors (less than  $\sim n_{\text{bins}}^3$ ). This difference becomes even larger when redshift-space distortions ([Kaiser, 1987](#)), described in section [3.3.2](#), are included in the analysis. In addition, as explained in [Appendix 3.A](#), not only the modulus of the three  $\mathbf{k}$ -vectors but also the orientation of the triangle with respect to the line of sight becomes relevant. Consequently, a fine binning in  $\mathbf{k}$ -space corresponds to a very large number of possible triangles. While for the power spectrum in redshift space there are  $\sim n_{\text{bins}}^2$  possible data-vector elements, for the bispectrum there are slightly less than  $\sim n_{\text{bins}}^5$  possible triplets since the vectors need to satisfy the triangle condition.

The problem is that when doing parameter estimation using a data-vector for a given statistic, the corresponding covariance matrix is needed. This can either be computed analytically, which requires the evaluation of several multidimensional integral expressions, or it can be estimated from simulations. However, in order to obtain a precise and accurate estimate of the covariance using numerical simulations, the number of realisations must be larger than the number of elements of the data-vector ([Hartlap et al., 2007](#); [Taylor and Joachimi, 2014](#)). Therefore, numerically estimating the covariance matrix of a 3pt statistic from simulations could become very expensive in terms of simulations required to predict it accurately. Indeed, one usually sacrifices the stronger constraining power achievable by considering a larger number of triangle configurations for a more accurate estimate of the covariance matrix obtained using only a subset of all the possible triangles. Compressing the original data-vector is an efficient method to avoid losing access to part of the constraining power contained in the 3pt statistics. Alternatives to the full bispectrum have been proposed in the last few years. For example [Schmittfull et al. \(2015\)](#) substituted to the tree level matter bispectrum, functions of the cross-power spectrum between  $\delta^2(x)$  and  $\delta$ . The performances of some of them, including the modal decomposition proposed by [Schmittfull et al. \(2013\)](#), have been recently studied by [Byun et al. \(2017\)](#). We will compare their results with ours in the conclusions.

Therefore one can reduce the original number of data-vector elements either by performing a selection of triangles based on some criteria or by compressing

the data-vector. This paper is about this second option, where we present two variations of the method presented in [Heavens et al. \(2000\)](#) and named MOPED, which achieves maximal compression of the original data-vector by extending to the multiple parameters case the algorithm introduced in [Tegmark et al. \(1997\)](#). The underlying principle is to assign a vector, such that the weights are proportional to the sensitivity of each element to the variation of a model parameter. When the covariance matrix can be assumed to be parameter independent, the dimension of the compressed data-vector corresponds to the number of considered model parameters.

We use this prescription in two ways to do parameter inference, which are summarised in [Figure 3.1](#). In our first method, we run an MCMC sampling for the compressed data-vector (hereafter MCMC + MP). This has the immediate and appealing benefit for a cosmological survey of requiring fewer simulations to estimate precisely the covariance matrix in the case of a long data-vector like the bispectrum (with  $\gtrsim 10^3$  triangles) to just 1 data-vector element for each cosmological parameter. Indeed, in order to have a reliable estimate of the covariance matrix, one needs a significantly larger number of mocks than of data-vector elements. For example, in the DR12 BOSS paper studying the bispectrum ([Gil-Marín et al., 2017](#)), the number of triangles used was limited to 825 out of the 6391 possible ones, since only 2048 mocks were available ([Kitaura et al., 2016](#)).

In our second method, before compressing, we orthogonalise the parameter space by diagonalising the Fisher information matrix (principal component analysis, hereafter PCA + MP). This enables sampling from 1D posterior distributions of the new set of parameters, to recover the full multi-dimensional posterior distribution for the original physical set of parameters without the need of an MCMC sampling. This proves to be  $\sim 10^3$  times faster than an MCMC run, in particular because the linear matter power spectrum needs to be recomputed far fewer times.

In addition to being competitive with the standard likelihood approach, this method could be also very useful for future cosmological surveys in order to test and forecast the constraining power of the expected data set for several different combinations of cosmological and nuisance (e.g. bias) parameters. We apply our analysis to the redshift space galaxy bispectrum and to the joint data-vector formed by the galaxy power spectrum and bispectrum.

The paper is structured as follows: [section 3.3](#) describes the basis of perturbation theory upon which the data-vector estimators and covariance matrix terms are computed. In [section 3.4](#) we present the covariance terms. [Section 3.5](#) contains the specifics of the analysis performed. In [section 3.6](#) the compression

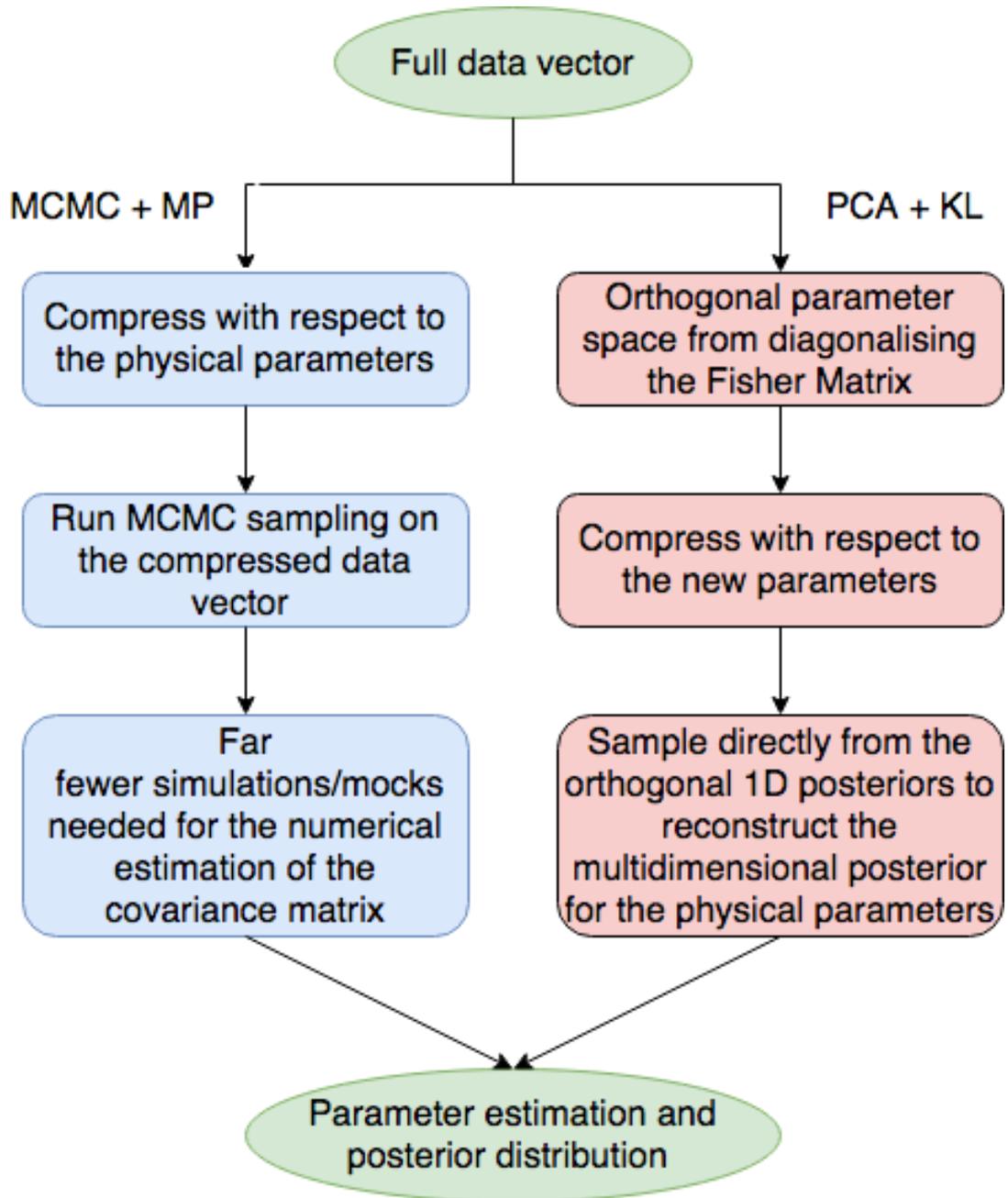


Figure 3.1: Diagram highlighting the two compression methods presented in this work, including the respective advantages with respect to the uncompressed data-vector.

formalism is presented. In section 3.7 we present the results of applying the MCMC + MP method. Section 3.8 describes the performance of the PCA + MP compression in obtaining constraints on the cosmological parameters. In section 3.9 we comment on the added value given by jointly using the power spectrum to the bispectrum. We conclude in section 3.10. All detailed derivations are deferred to the appendices 3.A (estimator definition), 3.B (covariance matrix terms), 3.C (compressed covariance formalism), 3.D (weights orthogonalisation).

## 3.3 Perturbation theory with redshift-space distortions

### 3.3.1 Bias model

Since luminous objects like galaxies are not exact tracers of the underlying dark matter distribution, it is necessary to model their relationship. This is commonly referred to as ‘galaxy biasing’; the relationship could be linear/non-linear, deterministic/stochastic, local/non-local, and a function of scale and cosmic time. Understanding biasing is important in its own right as a probe of galaxy formation and evolution. Galaxy biasing was recognised when it was noticed that different populations of galaxies (e.g. spirals, ellipticals) have different clustering strengths.

A physical mechanism for galaxy biasing was suggested by Kaiser (1984b) and developed by Bardeen et al. (1986b), namely that galaxies would tend to form in peaks in the matter density distribution thus being more clustered than the underlying matter distribution. In this model, more massive (and thus rarer) tracers are naturally more highly biased. Biasing can be non-linear (McDonald and Roy, 2009b) and stochastic (Dekel and Lahav, 1999b). Galaxy biasing also evolves with redshift (Clerkin et al., 2015 and references therein), being larger at higher redshift. It is also scale dependent at small physical scales where the non-linear effects of galaxy formation are important, although there is almost no scale dependence above  $20 - 40 \text{ Mpc } h^{-1}$  (Manera and Gaztañaga, 2011; Crocce et al., 2015). Another popular approach, known as the Halo Model, is to parametrise the relationship between galaxies and the dark matter distribution by assuming that all galaxies reside in dark matter haloes or sub-haloes, discussed e.g. in Tinker, 2010 and references therein. We note that biasing and the relationship between galaxies and haloes can also be constrained through combinations of various observables, e.g. galaxy positions and weak gravitational lensing.

Here we adopt a biasing model which is a Taylor expansion of the galaxy density fluctuations  $\delta_g$  in terms of the matter fluctuations  $\delta_m$  (Fry and Gaztanaga, 1993):

$$\delta_g = \sum_{i=0}^{\infty} \frac{b_i}{i!} \delta_{i,m}^{\text{n.l.}} \quad (3.1)$$

where  $\delta_{i,m}^{\text{n.l.}}$  is the non-linear matter density fluctuation.

In this work, only the first two terms of the above expansion are considered,  $b_1$  and  $b_2$ . The above bias model considers only the local relationship between  $\delta_g$  and  $\delta_m$ ; non-local bias terms may be included in a more accurate modelling (Chan et al., 2012; Baldauf et al., 2012; Bel et al., 2015).

### 3.3.2 Redshift space formalism

The conversion of the galaxy redshifts from surveys like BOSS or DESI to proper distances is a cosmological model-dependent operation. In addition, the local gravitational field influences the peculiar velocities of galaxies producing redshift-space distortions (Kaiser, 1987) which affect the observed redshift. Heavens et al. (1998) have shown how to express the Fourier transform of the redshift space galaxy overdensity fluctuation, measured in cosmological surveys,  $\delta_g^s(\mathbf{k})$  in relation to the real space linear matter fluctuation  $\delta_m(\mathbf{k})$

$$\begin{aligned} \delta_g^s[\mathbf{k}] &= F_s^{(1)}[\mathbf{k}] \delta_m(\mathbf{k}) \\ &+ \frac{1}{(2\pi)^3} \int d^3k_1 d^3k_2 \delta_D(\mathbf{k} - \mathbf{k}_2 - \mathbf{k}_1) \\ &\times F_s^{(2)}[\mathbf{k}_1, \mathbf{k}_2] \delta_m(\mathbf{k}_1) \delta_m(\mathbf{k}_2) \\ &+ \frac{1}{(2\pi)^6} \int d^3k_1 d^3k_2 d^3k_3 \delta_D(\mathbf{k} - \mathbf{k}_3 - \mathbf{k}_2 - \mathbf{k}_1) \\ &\times F_s^{(3)}[\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3] \delta_m(\mathbf{k}_1) \delta_m(\mathbf{k}_2) \delta_m(\mathbf{k}_3), \end{aligned} \quad (3.2)$$

where the redshift-space distortion kernels  $F_s^{(1,2,3)}$  are given in the Section 1.5. In this work all  $\mathbf{k}$ -vectors are described in terms of their components parallel  $k^{\parallel}$  and perpendicular  $k^{\perp}$  to the line of sight. We define  $\mu = k_i^{\parallel}/k_i$ . The galaxy redshift space power spectrum is defined as

$$\langle \delta_g^s(\mathbf{k}_1) \delta_g^s(\mathbf{k}_2) \rangle = (2\pi)^3 P_g^s(\mathbf{k}_1) \delta_D(\mathbf{k}_1 + \mathbf{k}_2). \quad (3.3)$$

Substituting Equation 3.2 into Equation 3.3 and applying Wick's theorem assuming that the initial perturbations are Gaussian the power spectrum is given by:

$$\begin{aligned}
 P_g^s(\mathbf{k}) &\equiv P_g^{s,(1)} + P_g^{s,(2)} + P_g^{s,(3)} \\
 &= (b_1 + f\mu^2)^2 P^m(k) \\
 &+ 2 \int \frac{d^3q}{(2\pi)^3} P^m(q) P^m(|\mathbf{k} - \mathbf{q}|) \left( F_s^{(2)}[\mathbf{q}, \mathbf{k} - \mathbf{q}] \right)^2 \\
 &+ 6 (b_1 + f\mu^2) P^m(k) \int \frac{d^3q}{(2\pi)^3} P^m(q) F_s^{(3)}[\mathbf{q}, -\mathbf{q}, \mathbf{k}], \tag{3.4}
 \end{aligned}$$

where the growth rate is defined and parametrised as  $f(z) \equiv d \ln D / d \ln a \simeq \Omega_m^\gamma(z)$  where  $\Omega_m(z)$  is the mass density parameter function at a given redshift  $z$ , the growth index  $\gamma \simeq 0.55$  for a standard cosmology (Peebles, 1980; Lahav et al., 1991; Linder, 2005),  $D(a)$  is the growing mode of the amplitude fluctuation, and scale factor  $a = (1 + z)^{-1}$ .  $P^m(k)$  is the linear matter power spectrum defined analogously to Equation 3.3. In this work when considering the power spectrum data-vector we stop at tree level using only  $P_g^s = P_g^{s,(1)}$ . This choice is consistent with previous analyses (Scoccimarro et al., 1999; Sefusatti et al., 2006; Song, Yong-Seon and Taruya, Atsushi and Oka, Akira, 2015; Gagrani and Samushia, 2017). The error arising from excluding the 1loop terms (Jain and Bertschinger, 1994; Bernardeau et al., 2002; Taruya et al., 2008; Lazanu et al., 2016) is less than 10% up to  $k_{\max} = 0.2h\text{Mpc}^{-1}$  increasing to 30% for  $k_{\max} = 0.3h\text{Mpc}^{-1}$ . An accurate modelling of the redshift space galaxy power spectrum was introduced by Taruya et al. (2010). This translates approximately into an error on the diagonal of the power spectrum covariance matrix computed in section 3.4.1 of  $\sim 1\%$  up to  $k_{\max} = 0.2h\text{Mpc}^{-1}$  and  $\sim 9\%$  for  $k_{\max} = 0.3h\text{Mpc}^{-1}$  when including cosmic variance terms. For brevity, sometimes we write  $P_g^s(k_i^\perp, k_i^\parallel) \equiv P_g^s(\mathbf{k}_i)$ , keeping in mind that the galaxy power spectrum depends on the parallel and perpendicular components of the wave-vector. Therefore  $P_g^s$  will not depend on the second order bias parameter  $b_2$ .

Analogously, the bispectrum is defined as (Fry, 1984):

$$\begin{aligned}
 \langle \delta_g^s(\mathbf{k}_1) \delta_g^s(\mathbf{k}_2) \delta_g^s(\mathbf{k}_3) \rangle &= (2\pi)^3 B_g^s(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) \\
 &\times \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3). \tag{3.5}
 \end{aligned}$$

The relation to the linear matter density power spectrum is at first order:

$$\begin{aligned}
 B_g^s(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) &= 2P^m(k_1) P^m(k_2) F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] F_s^{(2)}[\mathbf{k}_1, \mathbf{k}_2] \\
 &+ \text{two cyclic terms.} \tag{3.6}
 \end{aligned}$$

For the redshift space galaxy bispectrum, the most recent and accurate models (up to the mild non-linear regime) were introduced by Hashimoto et al. (2017) and Bertacca et al. (2017).

### 3.4 Covariance

In this section we summarise the covariance terms at tree level for the power spectrum and the bispectrum, including also the cross covariance between the two. Here are written only the final analytical expressions, while we show the full derivations in Appendix 3.B.

#### 3.4.1 $C^{\text{PP}}$ : power spectrum covariance matrix

The power spectrum covariance matrix is given by two terms,

$$C^{\text{Ps}} = C_{\text{P}}^{\text{PP}} + C_{\text{T}}^{\text{PP}}, \quad (3.7)$$

where  $C_{\text{P}}^{\text{PP}}$  is proportional to the square of the power spectrum and  $C_{\text{T}}^{\text{PP}}$  is proportional to the trispectrum. The first term is given by:

$$\begin{aligned} C_{\text{P}}^{\text{PP}}(\bar{\mathbf{k}}_1; \bar{\mathbf{k}}_3) &= \frac{1}{V_s^2 V_1^c V_3^c} \prod_{i=0}^4 \int dV_i^c \delta_{\text{D}}(\mathbf{k}_1 + \mathbf{k}_2) \delta_{\text{D}}(\mathbf{k}_3 + \mathbf{k}_4) \\ &\quad \times \langle \delta_{\text{g}}^{\text{s}}(\mathbf{k}_1) \delta_{\text{g}}^{\text{s}}(\mathbf{k}_3) \rangle \langle \delta_{\text{g}}^{\text{s}}(\mathbf{k}_2) \delta_{\text{g}}^{\text{s}}(\mathbf{k}_4) \rangle + 1\text{p}. \\ &\approx \frac{2\pi}{V_s \bar{k}_1^{\perp} \Delta k^2} D_{1234}^{\text{P}} P_{\text{g}}^{\text{s}}(\bar{\mathbf{k}}_1)^2 = \frac{4\pi}{V_s \bar{k}_1^{\perp} \Delta k^2} \delta_{13}^{\text{K}} P_{\text{g}}^{\text{s}}(\bar{\mathbf{k}}_1)^2, \end{aligned} \quad (3.8)$$

where the integrals are over a cylindrical shell  $V_i^c$  centered at each  $\bar{\mathbf{k}}_i$  with integral limits on:  $\bar{k}_i^{\parallel} - \Delta k^{\parallel}/2 \leq k_i^{\parallel} \leq \bar{k}_i^{\parallel} + \Delta k^{\parallel}/2$ ,  $\bar{k}_i^{\perp} - \Delta k^{\perp}/2 \leq k_i^{\perp} \leq \bar{k}_i^{\perp} + \Delta k^{\perp}/2$  and  $0 \leq \phi \leq 2\pi$ . From the definition of the power spectrum estimator given in Appendix 3.A, the cylindrical bins for  $\mathbf{k}_2$  and  $\mathbf{k}_4$  are centered respectively on  $\bar{\mathbf{k}}_1$  and  $\bar{\mathbf{k}}_3$ . When a Dirac delta  $\delta_{\text{D}}(\mathbf{k}_i + \mathbf{k}_j)$  is used to simplify one of the integrals over the cylindrical shells, it produces a Kronecker delta  $\delta_{ij}^{\text{K}}$  which is equal to one and not to zero only when  $k_i^{\parallel} = k_j^{\parallel}$  and  $k_i^{\perp} = k_j^{\perp}$ .

We have defined a combination of Kronecker deltas  $D_{1234}^{\text{P}} = \delta_{13}^{\text{K}} \delta_{24}^{\text{K}} + \delta_{14}^{\text{K}} \delta_{23}^{\text{K}}$  in order to take into account the additional permutation, which, given the initial Dirac deltas conditions, reduces to  $D_{1234}^{\text{P}} = 2 \times \delta_{13}^{\text{K}}$ .  $V_s$  is the survey volume. The second term in Equation 3.7 is proportional to the trispectrum and is given by

$$C_{\text{T}}^{\text{PP}}(\bar{\mathbf{k}}_1; \bar{\mathbf{k}}_3) \approx \frac{1}{2\pi V_s} \int_0^{2\pi} d\phi_{13} T_{\text{g}}^{\text{s}}(\bar{k}_1^{\perp}, \bar{k}_1^{\parallel}, \bar{k}_3^{\perp}, \bar{k}_3^{\parallel}, \phi_{13}). \quad (3.9)$$

where  $\phi_{13}$  is the difference between the azimuthal angles of  $\bar{\mathbf{k}}_1$  and  $\bar{\mathbf{k}}_3$ . The full expanded expression is written in Appendix 3.B, while the trispectrum definition is in Appendix 3.A.

### 3.4.2 $C^{\text{BB}}$ : bispectrum covariance matrix

For the bispectrum covariance matrix, at leading order, only the diagonal term proportional to the product of three power spectra is required:

$$C_{\text{P}}^{\text{BB}}(\bar{\mathbf{k}}_1, \bar{\mathbf{k}}_2, \bar{\mathbf{k}}_3; \bar{\mathbf{k}}_4, \bar{\mathbf{k}}_5, \bar{\mathbf{k}}_6) = \frac{(2\pi)^5 \Delta k_3^{\parallel}}{V_s \bar{k}_1^{\perp} \bar{k}_2^{\perp} \bar{k}_3^{\perp} \Delta k^6} \Lambda_{123}^{-1} \times D_{123456} \prod_{i=1}^3 P_{\text{g}}^{\text{s}}(\bar{\mathbf{k}}_i). \quad (3.10)$$

where  $\Lambda$  is a function defined in Appendix 3.A which is related to the fraction of wave-vector triplets allowed by the triangle condition such that the bispectrum estimator 3.28 is unbiased.  $D_{123456}$  is a shorthand notation for the sum of all the possible permutations of pairings of  $\mathbf{k}$ -vectors between the first and second triplets, encoded in Kronecker deltas, e.g.  $\delta_{14}^{\text{K}} \delta_{25}^{\text{K}} \delta_{36}^{\text{K}}$ . This is the symmetry factor which is equal respectively to 1, 2 and 6 in the case of scalene, isosceles and equilateral triangles. Full computations can be found in Appendix 3.B.

### 3.4.3 $C^{\text{BP}}$ : cross-variance matrix

The cross-variance term is also given by the sum of two parts. The first part is proportional to the product between the power spectrum and bispectrum. The second part is proportional to the tetraspectrum

$$C^{\text{BP}} = C_{\text{m1}}^{\text{BP}} + C_{\text{Te}}^{\text{BP}}. \quad (3.11)$$

The expression of the first term as derived in Appendix 3.B is:

$$C_{\text{m1}}^{\text{BP}}(\bar{\mathbf{k}}_1; \bar{\mathbf{k}}_3, \bar{\mathbf{k}}_4, \bar{\mathbf{k}}_5) = 2 \times \frac{(2\pi)^2}{V_s V_1^c} \Lambda_{345}^{-1} (\delta_{13}^{\text{K}} + \delta_{14}^{\text{K}} + \delta_{15}^{\text{K}}) \times P_{\text{g}}^{\text{s}}(\bar{\mathbf{k}}_1) B_{\text{g}}^{\text{s}}(\bar{\mathbf{k}}_3, \bar{\mathbf{k}}_4, \bar{\mathbf{k}}_5), \quad (3.12)$$

Where the semicolon separates the wave vector relative to the power spectrum from the ones of the bispectrum.

The last non-vanishing term is the one proportional to the tetraspectrum  $T_{\text{e,g}}^{\text{s}}$  given by the connected part of the five-points correlator

$$\begin{aligned}
 C_{\text{Te}}^{\text{BP}}(\bar{\mathbf{k}}_1; \bar{\mathbf{k}}_3, \bar{\mathbf{k}}_4, \bar{\mathbf{k}}_5) &= \frac{(2\pi)^2 \Delta k_5^{\parallel}}{V_s} \Lambda_{345}^{-1} \prod_{i=1,3,4,5} \int \frac{dV_i^c}{V_i^c} \\
 &\times \delta_{\text{D}}(\mathbf{k}_3 + \mathbf{k}_4 + \mathbf{k}_5) T_{\text{e,g}}^{\text{s}}(\mathbf{k}_1, \mathbf{k}_3, \mathbf{k}_4, \mathbf{k}_5). \quad (3.13)
 \end{aligned}$$

The definition and analytical expression for the tetraspectrum are in Appendix 3.A. In the Appendix it is shown that at leading order the tetraspectrum is proportional to the fourth power of the linear matter power spectrum. Being this one order higher than all the terms considered in the paper,  $C_{\text{Te}}^{\text{BP}}$  has not been included in the numerical computations.

All the terms rederived above agree with the ones derived in the literature when considering redshift-space distortions (Sefusatti et al., 2006; Scoccimarro et al., 1998).

#### 3.4.4 Shot noise contribution

All terms of the covariance matrix have been corrected in order to account for the shot noise, which is given by the average number density of galaxies. In our analysis we consider a DESI-like ELGs sample with  $\bar{n}_g \approx 1.175 \times 10^{-3} h^3 \text{Mpc}^{-3}$  and a survey volume  $V_s \approx 12.3 h^{-3} \text{Gpc}^3$  (Levi et al., 2013; Song, Yong-Seon and Taruya, Atsushi and Oka, Akira, 2015; DESI Collaboration and others, 2016). The only correction corresponds to substituting the galaxy power spectrum with:

$$P_{\text{g}}^{\text{s}}(\mathbf{k}) \longrightarrow P_{\text{g}}^{\text{s}}(\mathbf{k}) + \frac{1}{\bar{n}_g}. \quad (3.14)$$

We made the assumption that the shot noise is well approximated by a Gaussian distribution (which is reasonable if the galaxy number density is fairly high). In that case only second moments exist, and thus contributions appear only in the 2pt correlators, and those yield power spectrum contributions to the cosmic variance. If the noise were Poisson-distributed, the corrected expressions would become much more complicated (Matarrese et al., 1997).

### 3.5 Analysis setup

In this section we describe the pipeline and codes for the compression and the MCMC analyses. We compressed the power spectrum and the bispectrum with respect to the set of parameters:  $\{b_1, b_2, f, \Omega_{\text{m}}, \Omega_{\text{b}}, A_{\text{s}}, n_{\text{s}}\}$  where the first two are galaxy bias parameters,  $f$  is the linear growth rate,  $\Omega_{\text{m}}$  and  $\Omega_{\text{b}}$  are the density parameters of total matter and baryonic matter,  $A_{\text{s}}$  is the scalar amplitude of

the primordial perturbations and  $n_s$  is the scalar spectral index. For both cases the fiducial cosmology was fixed to  $b_1 = 1.90$ ,  $b_2 = 0.20$ ,  $f = \Omega_m(z = 0.81)^{0.55}$ ,  $\Omega_m(z = 0) = 0.307$ ,  $\Omega_b(z = 0) = 0.0482$  and  $A_s = 2.9 \times 10^{-9}$  similar to the one used in [Kitauro et al. \(2016\)](#) and close to the one reported in the *PLANCK* 15 results ([Ade et al., 2016](#)). The redshift used is that of the effective redshift-bin ( $0.6 < z < 1.0$ ) of a ELG sample of a DESI-like survey with  $z_{\text{eff}} = 0.81$ .

The  $k$ -range chosen is  $0.01 < k < 0.2 \text{ Mpc}^{-1} h$ , 12 bins for both parallel (linear binning) and perpendicular (logarithmic binning) to the line of sight components. We have adopted a logarithmic binning for the perpendicular components of the wave-vectors in order to better capture the different features at different scales. A linear binnings has been used for parallel to the line of sight. For the perpendicular component,  $\Delta \log_{10} k^\perp = 0.11827$  while for the parallel one  $\Delta k^\parallel = 0.0182 \text{ Mpc}^{-1} h$ . With these settings, the resulting number of configurations satisfying the triangle condition is  $N_{\text{tr}} = 1333$  for the bispectrum and  $N_{\text{pairs}} = 132$  combinations of parallel and perpendicular components for the power spectrum .

The triangle configurations for the bispectrum are generated by a five dimensional loop choosing first the three perpendicular components of the sides of the triangle and secondly two of the parallel ones. The third parallel component is chosen such that the final triangle satisfies the triangle condition. All sides and projections must be in the range given above.

For the MCMC sampling we have used 64 samplers together with the same number of Xeon E5-2650 processors connected through MPI ([Gabriel et al., 2004](#)), each with 2000 burn-in steps followed by 10000 steps for the actual posterior sampling. The Python package Emcee was used as MCMC sampler ([Foreman-Mackey et al., 2013](#)). Uninformative flat priors have been used both in the compression code and MCMC sampler. We employed the CAMB code ([Lewis et al., 2000](#)) in order to generate the linear matter power spectrum for different cosmological parameters.

## 3.6 Compression Formalism

### 3.6.1 Fisher information matrix

The log-likelihood  $\mathcal{L}$  for a Gaussian probability distribution relative to an  $n$ -dimensional data-vector  $\mathbf{x}$ , can be written as

$$-2\mathcal{L} = n \ln 2\pi + \ln \det \mathbf{Cov} + (\mathbf{x} - \langle \mathbf{x} \rangle)^\top \mathbf{Cov}^{-1} (\mathbf{x} - \langle \mathbf{x} \rangle), \quad (3.15)$$

where  $\mathbf{Cov} = \langle (\mathbf{x} - \langle \mathbf{x} \rangle) (\mathbf{x} - \langle \mathbf{x} \rangle)^\top \rangle$  is the covariance matrix and  $\langle \mathbf{x} \rangle$  is the mean of the data-vector. From this quantity the Fisher information matrix can be defined as

$$F_{ij} = - \left\langle \frac{\partial^2 \mathcal{L}}{\partial \theta_i \partial \theta_j} \right\rangle \Big|_{\boldsymbol{\theta}_{ML}} \equiv - \langle \mathcal{L}_{,ij} \rangle, \quad (3.16)$$

which is a measure of the curvature around the maximum likelihood point  $\boldsymbol{\theta}_{ML} = (\theta_{ML}^1, \theta_{ML}^2, \dots, \theta_{ML}^m)$  where the  $\theta$ 's are the  $m$  model parameters and the comma notation indicates the derivatives with respect to them. In the case of a Gaussian likelihood, the Fisher matrix can be expressed as

$$F_{ij} = \frac{1}{2} \text{Tr} [\mathbf{A}_i \mathbf{A}_j + \mathbf{Cov}^{-1} \mathbf{M}_{ij}], \quad (3.17)$$

where the matrices  $\mathbf{A}_i$  and  $\mathbf{M}_{ij}$  are defined as  $\mathbf{A}_i \equiv \mathbf{Cov}^{-1} \mathbf{Cov}_{,i}$  and  $\mathbf{M}_{ij} \equiv \langle \mathbf{x} \rangle_{,i} \langle \mathbf{x} \rangle_{,j}^\top + \langle \mathbf{x} \rangle_{,i}^\top \langle \mathbf{x} \rangle_{,j}$ .

The diagonal entries of the Fisher matrix are related to the minimum error attainable in estimating a parameter  $\theta_i$ . In particular, in the case of single parameter estimation the minimum attainable error is  $\Delta \theta_i^{\min} = 1/(F_{ii})^{\frac{1}{2}}$ . When more than one parameter is considered, the full Fisher matrix is needed to compute the minimum marginalised error for each parameter, which is given by  $\Delta \theta_i^{\min} = (F_{ii}^{-1})^{\frac{1}{2}}$ . The target of the compression is to obtain a new data vector such that for each parameter  $\theta_i$ ,  $\Delta \theta_i^{\min}$  is minimised. In other words, the compression algorithm will be obtained by maximising  $F_{ii}$  for each parameter  $\theta_i$ .

### 3.6.2 Compression Algorithm

A general linear transformation of the data-vector  $\mathbf{x}$  with a transformation matrix  $\mathbf{B}$  is given by

$$\mathbf{y} = \mathbf{B} \mathbf{x}. \quad (3.18)$$

The mean and the covariance matrix for  $\mathbf{y}$  become respectively  $\langle \mathbf{y} \rangle = \mathbf{B} \langle \mathbf{x} \rangle$  and  $\mathbf{Cov}_{\mathbf{y}} = \mathbf{B}^\top \mathbf{Cov}_{\mathbf{x}} \mathbf{B}$ . In the case in which only one of the linear combinations of the data is considered,  $\mathbf{B}$  has only one row,  $\mathbf{B} = \mathbf{b}^\top$ . Therefore the diagonal entries of the Fisher matrix are given by:

$$F_{ii} = \frac{1}{2} \left( \frac{\mathbf{b}^\top \mathbf{Cov}_{,i} \mathbf{b}}{\mathbf{b}^\top \mathbf{Cov} \mathbf{b}} \right)^2 + \frac{(\mathbf{b}^\top \langle \mathbf{x} \rangle_{,i})^2}{(\mathbf{b}^\top \mathbf{Cov} \mathbf{b})}, \quad (3.19)$$

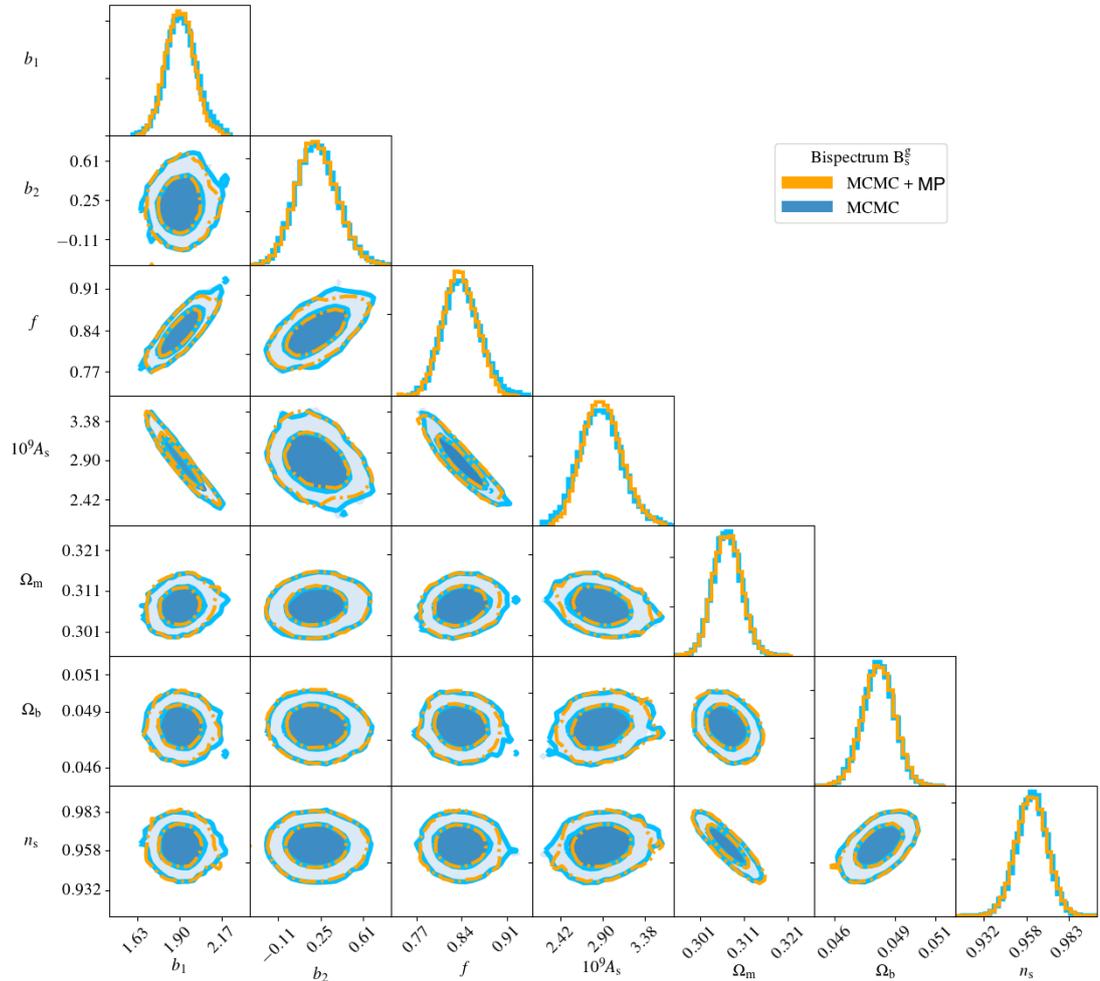


Figure 3.2: Bispectrum case: 1D and 2D posterior distributions for the MCMC sampling done using the full data-vector (blue) and the compressed data-vector obtained through the MP compression (orange). The contours correspond to the 68% and 95% confidence intervals. We see that there is no substantial loss of information despite the presence of strong degeneracies between the parameters; this is quantified in Figure 3.7 and Table 3.1. The 2D posterior distributions have been smoothed using the Gaussian kernel density estimation package provided by Scipy.

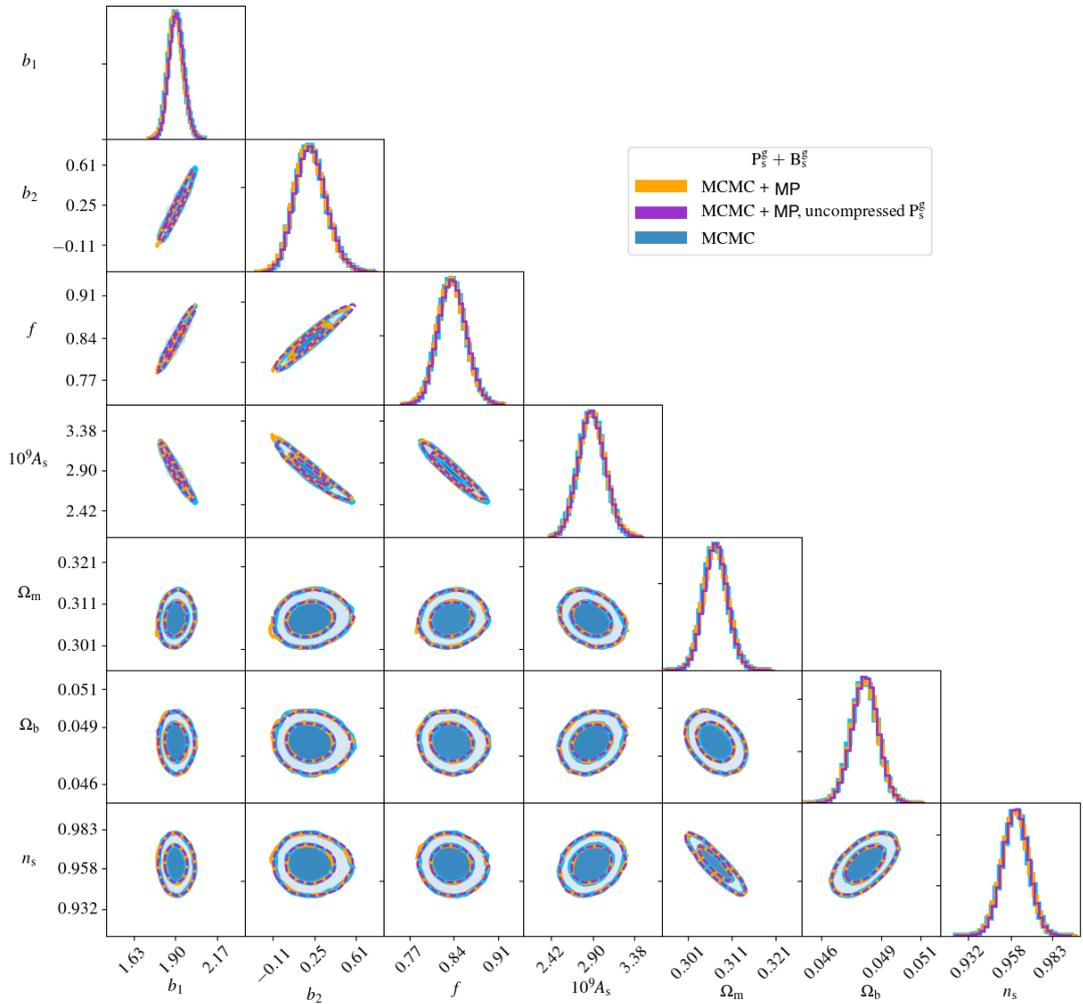


Figure 3.3: Joint data-vector  $[P_g^s, B_g^s]$  posteriors: the colours for the 1D and 2D posterior distributions are: the full data-vector (blue), the compressed data-vector (orange) and the uncompressed power spectrum plus the compressed bispectrum (purple). The contours correspond to the 68% and 95% confidence intervals. There is no substantial loss of information due to compression, which again is quantified in Figure 3.7 and Table 3.1

where again the comma notation followed by the index  $i$  stands for the derivative with respect to the model parameter  $\theta_i$ . As explained by [Heavens et al. \(2000\)](#), who also derive weights based on minimizing the Fisher matrix diagonal elements, it is a very complex problem to find an analytical solution for  $\mathbf{b}$  from the full Equation 3.19. Only by ignoring then the first term of Equation 3.19 it is possible to find an analytic solution for  $\mathbf{b}$ . This implies assuming that the derivatives of the covariance matrix with respect to model parameters are negligible when compared to the data vector ones. Recently [Heavens et al., 2017](#) presented a method to relax this approximation.

For both the power spectrum and bispectrum we numerically checked that it is reasonable to assume it to be valid, by taking the ratio between the diagonal elements of  $\mathbf{Cov}_{,i}/\sqrt{\mathbf{Cov}}$  with  $\langle \mathbf{x} \rangle_{,i}$  (corresponding to the ratio of the first and second term in Equation 3.19, respectively). This ratio for the bispectrum case results to be on average  $\leq 10\%$  for the considered parameters. There are peaks for few triangle configurations ( $\leq 10$  for more than a thousand triangles) for which the ratio is around 70 – 80% while the minima are around 2%.

By maximising  $F_{ii}$  as described in [Tegmark et al. \(1997\)](#) using a Lagrange multiplier, it follows that the compressed data-vector is given by a scalar

$$y_i = \langle \mathbf{x} \rangle_{,i}^T \mathbf{Cov}^{-1} \mathbf{x} \equiv \mathbf{b}_i^T \mathbf{x}, \quad (3.20)$$

where we have defined the weighting vector  $\mathbf{b}_i = \mathbf{Cov}^{-1} \langle \mathbf{x} \rangle_{,i}$ .

For the compression it is acceptable to use an analytic approximation of the covariance matrix (in our case as it is described in section 3.4), as any systematic error in the covariance would mostly degrade the compression; it would not bias the parameter inference. The compression enables then to use an accurate simulated covariance for the actual inference. This has the advantage of being able to compute weights for a data vector with an arbitrary large dimension (e.g. number of triangles). The Fisher matrix diagonal element can be rewritten as

$$F_{ii} = \langle \mathbf{x} \rangle_{,i}^T \mathbf{Cov}^{-1} \langle \mathbf{x} \rangle_{,i}. \quad (3.21)$$

In order to apply the MP compression method, one needs to choose a fiducial set of parameters at which to compute (analytically or numerically) the derivatives of the mean. In our case the fiducial values are reported in section 3.5. We compute the numerical derivatives using the five point method for the first derivative in one dimension ([Abramowitz, 1974](#)).

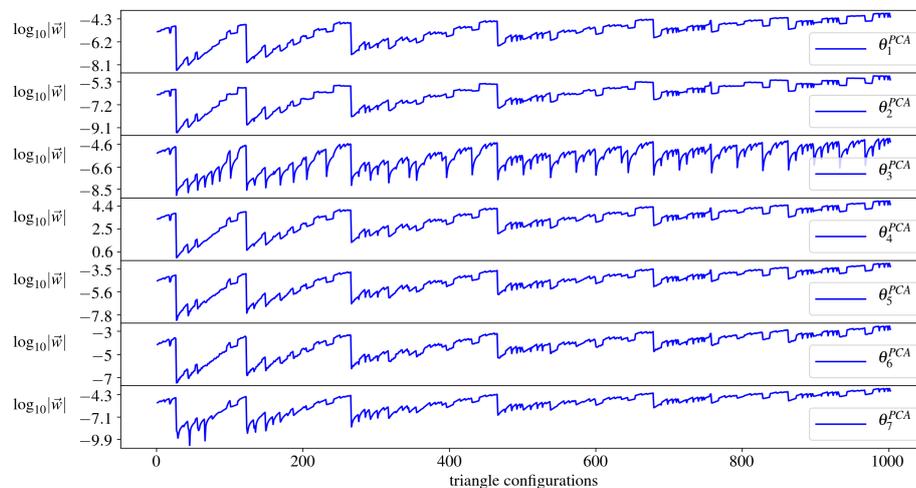


Figure 3.4: Logarithm of the absolute value of the weights for all the triangle configurations used in the bispectrum data-vector. Each row corresponds to the weights for the bispectrum  $B_g^s$  with respect to a specific linear combination  $\theta_{\text{PCA}}$  of the original cosmological parameters obtained by diagonalising the Fisher information matrix. The discontinuities observed reflect the five loops used to produce the set of triangles. From left to right the average size of the sides of the triangles increases, from the smallest triangle up to the largest. The amplitude of the weights slightly depends on the size of the triangle (increasing from left to right). The fluctuations are far more influenced by the shape and orientation of the triangles (which can be seen within each loop).

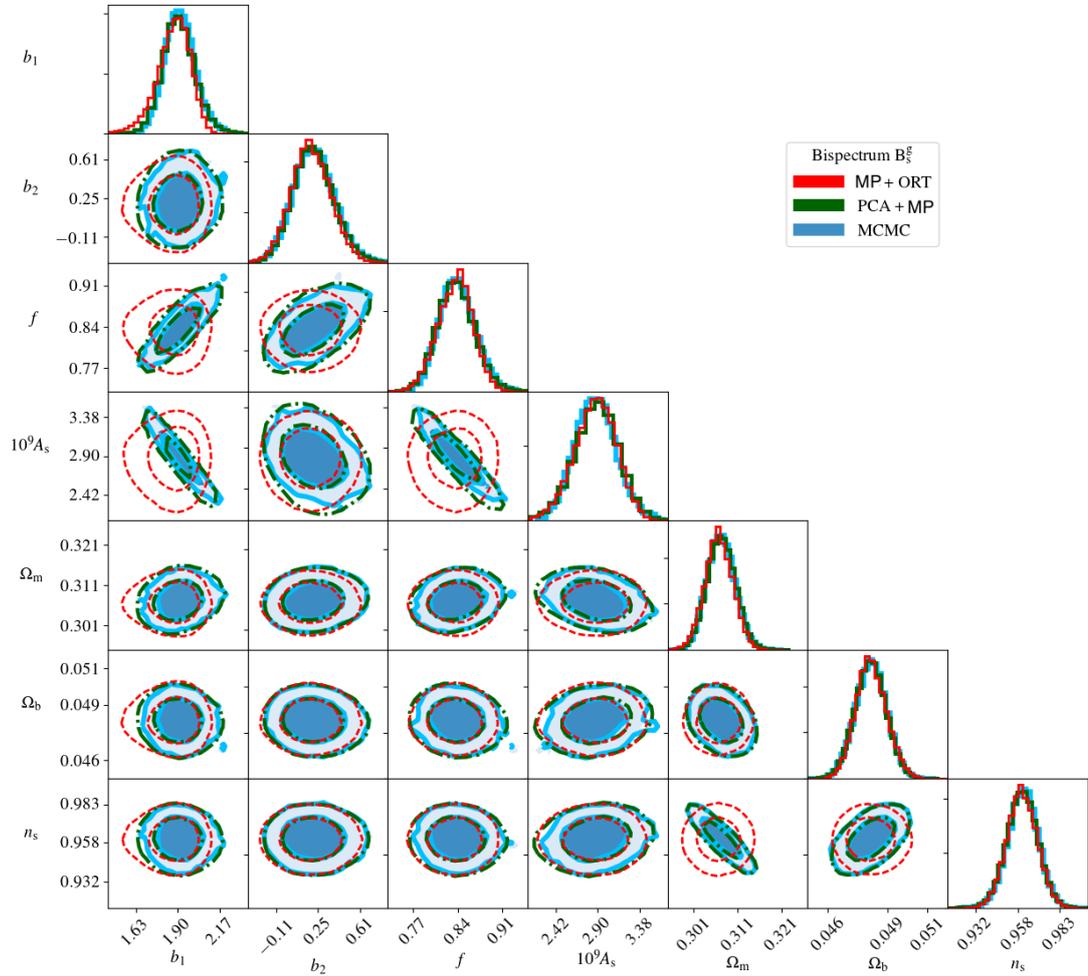


Figure 3.5: Comparison between the MCMC-derived posteriors and the ones obtained using only the compression. The 1D and 2D posterior distributions are relative to the data-vector  $B_{g,2}^s$ , for the three different cases: MCMC (blue), MP + ORT compression (red), PCA + MP compression (green). The contours correspond to the 68% and 95% confidence intervals (Figure 3.7 and Table 3.1 for numerical values). The method MP + ORT proposed by [Zablocki and Dodelson \(2016\)](#), fails to recover accurate 2D posterior distributions. Therefore taking the outer product of 1D posteriors distributions for the data-vector, compressed by using an orthogonalised set of weights, does not recover the multidimensional posterior beside its 1D marginalisations.

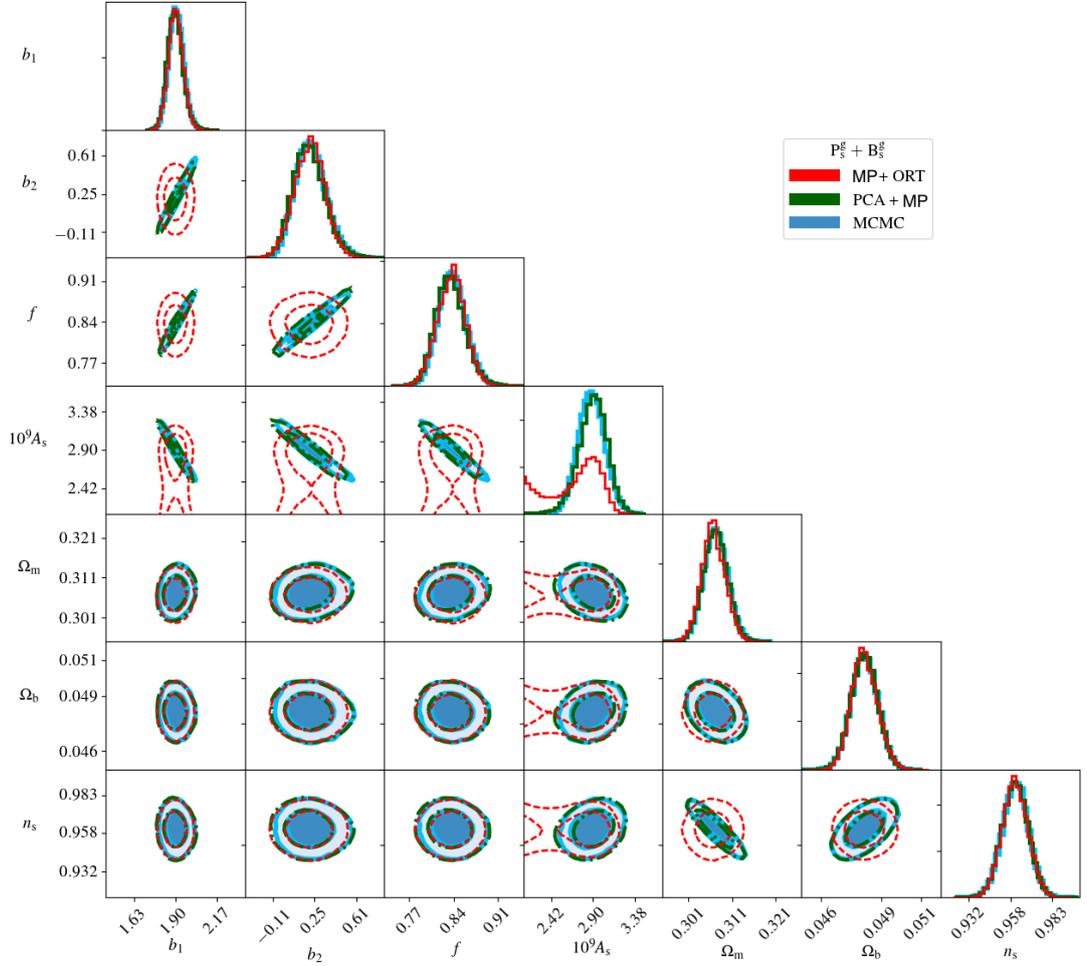


Figure 3.6: 1D and 2D posterior distributions using as data-vector  $[\mathbf{P}_g^s, \mathbf{B}_g^s]$ , for the three different cases: MCMC (blue), MP + ORT compression (red), PCA + MP compression (green). The contours correspond to the 68% and 95% confidence intervals (Figure 3.7 and Table 3.1 for numerical values). The MP + ORT method fails to reproduce the correct contours, both 1D and 2D, since for the used value of  $k_{\max}$  it fails to work, producing an unphysical secondary peak for the parameter  $A_s$ . On the contrary the PCA + MP method recovers with very good agreement the MCMC contours.

### 3.7 MCMC of Compressed data-vectors

In this section we compare the results obtained by running the MCMC algorithm for both uncompressed and compressed data-vectors. The goal is to check whether it would be possible to substitute the original data-vector for its compressed version. Even if this would not bring any relevant advantage in terms of speed when computing the data-vector from the theoretical model, in the case of a real cosmological survey, it will be much easier and less expensive in terms of required simulations / mock catalogues to estimate the covariance matrix of the compressed data-vector rather than the full data-vector. Moreover, in the case of the bispectrum, a much larger number of triangles could be used for the original data-vector, allowing more information to be captured, since the dimension of the compressed covariance matrix would be reduced to the number of parameters considered in the analysis. In order to use the compression, it is necessary to convert the covariance matrix for the full data-vector to the one for the compressed data-vector. This is shown in Appendix 3.C.

In this paper, we consider two cases for the data-vector: the galaxy bispectrum  $B_g^s$  and the joint data-vector  $[P_g^s, B_g^s]$ . For the latter, we include two further cases depending on whether or not we compress the power spectrum  $P_g^s$ .

Figure 3.2 shows the 1D and 2D marginalised posterior distributions when only the bispectrum data-vector is considered. For the considered set of parameters there is no substantial loss of information when the compressed data-vector is used, even if some degeneracies are present. In this case, the compressed vector has seven elements instead of the  $\sim 1000$  triangles for the uncompressed bispectrum. The 1D and 2D posteriors have been smoothed using a Gaussian kernel density estimation procedure, for clearer visualisation. Figure 3.3 shows the 1D and 2D marginalised posterior distributions when we consider both the power spectrum and the bispectrum. As can be seen, there is no qualitative or quantitative difference on the posterior distributions between compressing or not compressing the power spectrum together with the bispectrum. The precise numbers can be found in the summarising Table 3.1.

### 3.8 Posterior distributions directly from compressed data-vectors

It is possible to compute the 1D likelihood for each linear combination  $y_{\theta_i}$  of the original data-vector obtained compressing with respect to the parameter  $\theta_i$  as done in Zablocki and Dodelson (2016):

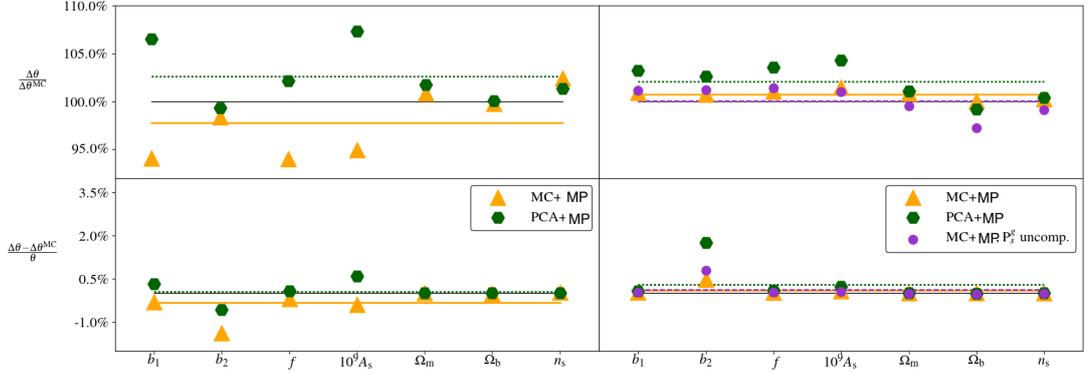


Figure 3.7: Ratio (first panel) and relative difference (second panel) between the 68% confidence intervals of the 1D marginalised posteriors for all the cosmological parameters obtained using either the compression plus the MCMC sampling (MCMC + MP, orange triangles) or just the compression (PCA + MP, green hexagons) with the ones obtained running the MCMC on the full data-vector. The black lines in all plots represent the reference values given by running a MCMC sampling using the full data-vector. The ratios  $\Delta\theta/\Delta\theta^{\text{MC}}$  give an idea of how much the 68% confidence intervals obtained through compression differ from the one given when using the full data-vector. The relative difference  $(\Delta\theta - \Delta\theta^{\text{MC}})/\theta$  is helpful since it scales the difference between the compressed and full data-vector results with respect to the value of the chosen parameters. The first column shows the results when just the bispectrum  $B_g^s$  is considered, where the largest discrepancy happens for the parameters most degenerate between each others ( $b_1$ ,  $f$ ,  $A_s$ ). Using the MCMC + MP method on the bispectrum proves to be with negligible loss of information. In the second column the ratios and relative differences in the case of the joint data-vector  $[P_g^s + B_g^s]$  are shown. In all the subplots, the horizontal lines show the averages of the ratios and relative differences of the same colour. For  $[P_g^s + B_g^s]$  the compression is optimal, both using MCMC + MP and PCA + MP. Compressing or not the power spectrum together with the bispectrum seems to produce no relevant statistical difference. The bottom right subplot  $b_2$  appears to be the parameters whose divergence from the MCMC result is greater with respect to its fiducial value.

$$\ln \mathcal{L} = -\frac{(y_{\theta_i} - \bar{y}_{\theta_i})^2}{2\sigma_{\langle y_{\theta_i} \rangle}^2} \quad \text{with} \quad \sigma_{\langle y_{\theta_i} \rangle}^2 = \sum_{j=1}^{n_x} b_j^2 \mathbf{Cov}_{jj}, \quad (3.22)$$

where  $\bar{y}_{\theta_i} \equiv \langle \bar{y}_{\theta_i} \rangle$  is the mean of the compressed linear combination. Nevertheless, these 1D likelihoods would not be realistic since they are obtained using the compressed data-scalar  $y_{\theta_i}$  by varying only one parameter at a time. In reality  $y_{\theta_i}$  is sensitive to all the other parameters.

In order to account for this fact, we transform the set of  $y_{\theta_i}$  scalars in such a way that, at linear order, they are sensitive only to the parameter with respect to which the original data-vector was compressed. One method to do so is to orthogonalise the parameter space by diagonalising the Fisher Information matrix. This has the advantage of obtaining realistic multidimensional posteriors directly from the 1D posteriors of the new set of model parameters (linear combinations of the original, physical parameters).

For comparison we include another method presented in [Zablocki and Dodelson \(2016\)](#) which consists of orthogonalising the weights in such a way that the Fisher matrix for the compressed data-vector would become diagonal; we call this weights-orthogonalisation and it is described in [Appendix 3.D](#). The idea behind these two slightly different orthogonalisation approaches is the same, namely to diagonalise the Fisher Information matrix. The difference is that while the diagonalisation in the first method is the starting point independently from compressing (or not compressing) the data-vector later, in the second method the diagonalisation is a consequence of the procedure used.

Both methods are approximations at linear level. Therefore as anticipated earlier they both fail (even if at different levels as it is shown later) when non-linear degeneracies are present. For example, cases where the 2D-posterior distribution of a pair of parameters can no longer be approximated by an ellipse but has instead a ‘banana’-shape. This break-down of the above procedures is in agreement with the fact that the compression method relies on the assumption that the multidimensional posteriors are Gaussian.

### 3.8.1 Parameter space orthogonalisation - PCA

As anticipated above, the compression returns only 1D posterior distributions for each one of the parameters. Therefore an additional step is required in order to be able to assume that these distributions correspond to 1D marginalisations from the original multidimensional distribution. This is because the MP compression with respect one model parameter returns a linear combination of the original data vector which is still sensitive to the variation of the other model parameters.

Zablocki and Dodelson (2016) obtained marginalised 1D posterior distributions by orthogonalising the weighting vectors for all the model parameters through a Gram-Schmidt like procedure described in Appendix 3.D. We label this method as MP + ORT.

In order to test whether the compression results match the ones obtained running a MCMC sampling algorithm, we also compare the 2D posterior distributions for different pairs of parameters. In the case of MP + ORT the only possibility to reconstruct the multidimensional posterior is to take the outer product of the 1D marginalised posteriors. However doing so would be misleading: it would return ellipses with axes oriented at different angles to the ones given by the MCMC sampling as it can be seen in Figures 3.5 and 3.6 looking at the difference between the 2D red (MP + ORT) and blue (MCMC) contours. These axes orientation reflects the different degree of degeneracy between the parameters.

Therefore in order to avoid this difference, instead of orthogonalising the weights we perform a principal component analysis (PCA) transformation of our parameter space before applying the MP compression. This is done by diagonalising the Fisher information matrix using the eigenvalue decompositions

$$\mathbf{F}_{\boldsymbol{\theta}_{\text{phys.}}} = \mathbf{P} \mathbf{F}_{\boldsymbol{\theta}_{\text{PCA}}} \mathbf{P}^\top \quad \text{where} \quad \boldsymbol{\theta}_{\text{PCA}} = \mathbf{P}^\top \boldsymbol{\theta}_{\text{phys.}}, \quad (3.23)$$

where  $\mathbf{P}$  is the linear transformation matrix. After having diagonalised the Fisher matrix we compress the data-vector with respect to this new set of parameters  $\boldsymbol{\theta}_{\text{PCA}}$ . The weights obtained doing so are displayed in Figure 3.4 for the case of the bispectrum as data-vector. The effect of a PCA decomposition is to rotate the parameter space to the axes corresponding to the degeneracies between the original set of parameters. Therefore taking the outer product of the 1D posteriors of the parameters  $\boldsymbol{\theta}_{\text{PCA}}$  in order to get the multidimensional posterior distribution should return with good approximation the one sampled by the MCMC code. Once the multidimensional posterior has been reconstructed by taking the outer product of the 1D posterior distributions for the  $\boldsymbol{\theta}_{\text{PCA}}$  set of parameters, this can be randomly sampled in terms of the physical parameters  $\boldsymbol{\theta}_{\text{phys.}}$  using the rotation matrix  $\mathbf{P}$  in order to get the 1D and 2D marginalised posterior distributions for the  $\boldsymbol{\theta}_{\text{phys.}}$  parameters. The results for the PCA + MP method just described can be seen looking at the green (PCA + MP) and blue (MCMC) 2D contours still in Figures 3.5 and 3.6. The PCA + MP method recovers tilted ellipses in good agreement with the MCMC ones. On the contrary the method MP + ORT proposed by Zablocki and Dodelson (2016), fails to recover accurate 2D posterior distributions. Therefore taking the outer product of 1D posteriors distributions

for the data-vector, compressed by using an orthogonalised set of weights, cannot recover the multidimensional posterior beside its 1D marginalisations unless the physical parameters are completely uncorrelated.

### 3.8.2 Comparison with MCMC sampling

Figure 3.7 shows the ratio and relative difference of the 68% confidence intervals of the 1D marginalised posteriors between the compression methods MCMC + MP and PCA + MP and the standard MCMC sampling. In the bispectrum case, while MCMC + MP tends to underestimate the 68% confidence intervals obtained by the MCMC, PCA + MP tends to overestimate them by approximately the same amount ( $\sim 2.5\%$ ). For the joint data-vector, MCMC + MP returns equivalent confidence intervals to the MCMC's ones while PCA + MP overestimates them in average by  $\sim 2.5\%$  (first panel) but still less than  $\sim 1\%$  in terms of relative difference (second panel). We consider the averages of these ratios since at the same time the compression methods overestimate the 68% confidence intervals for some parameters while underestimating them for others. In terms of the individual parameters 68% confidence intervals, PCA + MP diverges at most from the MCMC's ones respectively by  $\sim 7\%$  and  $\sim 6\%$  in the case of the bispectrum and the joint data-vector. MCMC + MP diverges at most respectively by  $\sim 6\%$  and  $\sim 4\%$  in the case of the bispectrum and the joint data-vector.

Both in the case of the bispectrum and in the case of the joint data-vector  $[\mathbf{P}_g^s, \mathbf{B}_g^s]$  the compression for both MCMC + MP (Figures 3.2 and 3.3) and PCA + MP (Figures 3.5 and 3.6) methods well matches the 1-2D contours derived from the MCMC sampling. This shows that in the bispectrum case there is no relevant difference between MCMC + MP and PCA + MP methods while there is a very small one in the case of the joint data-vector.

In the bispectrum case (Figure 3.5) when the 2D posterior distributions are considered, only the contours derived by orthogonalising the parameter space before the compression (PCA + MP) have elliptical shapes with the right inclination. In the case of the joint data-vector instead (Figure 3.6), the MP method alone fails to recover even the 1D posterior for the parameter  $A_s$ , producing an unphysical secondary peak, while using PCA + MP returns 1D and 2D contours equivalent to the MCMC's ones. This shows that the orthogonalisation of the parameter space (PCA step) performs better than the weights orthogonalisation for combinations of parameters presenting strong non-linear degeneracies. As described in the following subsection, the compression methods PCA + MP and MP + ORT break down when the  $k_{\max}$  used drops below a certain threshold

value. For the MP + ORT method this is  $\sim 0.2 \text{ Mpc}^{-1}h$  as shown in Figure 3.6, while for PCA + MP the threshold value is lower. Applying the compression up to those mildly non-linear scales on real data like the BOSS DR12 requires to modify the model in order to take into account non-local bias terms. However this would not require the introduction of additional bias parameters. This is because assuming that even if the galaxy bias is non-local in Eulerian space, it is in Lagrangian space, and hence the additional non-local bias terms can be related at first order to the linear bias  $b_1$  as done in Gil-Marín et al. (2015).

The parameter set used in our analysis has been chosen to have strong degeneracies in order to test the applicability of the MCMC + MP and PCA + MP methods. When working with real data the parameter set is usually designed to be less degenerate, for example considering only  $\sigma_8$  instead of  $\Omega_m$  and  $A_s$  or also combining  $\sigma_8$  with  $b_1$ ,  $b_2$  and  $f$ .

In absence of non-linear (banana shape) degeneracies, the compression PCA + MP can be a valid and much faster substitute to the standard MCMC sampling since it returns realistic multidimensional posterior distributions. In particular, compression could be used to accurately forecast the constraints for different sets of parameters. It would also be possible to qualitatively study the degeneracies present in each set.

Running both compression algorithms (MP or PCA) takes approximately 20 minutes for seven parameters, with the time depending on the number of intervals for each parameter range and therefore how many times the linear matter power spectrum is computed. As a rough estimate, if we consider 100 intervals for each of the seven parameter ranges,  $P^m$  will be recomputed 300 times (no need to recompute the matter power spectrum when varying only one of the parameters  $b_1$ ,  $b_2$ ,  $f$  or  $A_s$ ). For a standard MCMC with 64 samplers and  $10^4$  steps instead of 300 times  $P^m$  needs to be recomputed  $64 \times 10^4$  times. Therefore the compression is more than  $\sim 2100$  times faster than the MCMC.

The compression PCA + MP is also much less demanding in terms of computing resources. It can be run in few minutes on a laptop with a single Intel i7 processor while for the MCMC we have used 64 threads working in parallel on 64 Xeon E5-2650 processors connected using MPI for  $\sim 72$  hours.

### 3.8.3 Limitations of the compression

Unfortunately the orthogonalisation prescriptions do not work when "strong" non-linear degeneracies are present in the parameter space and the Fisher matrix can no longer be diagonalised in practice. This happens because the multidimensional posterior distribution can no longer be approximated by a multi-variate

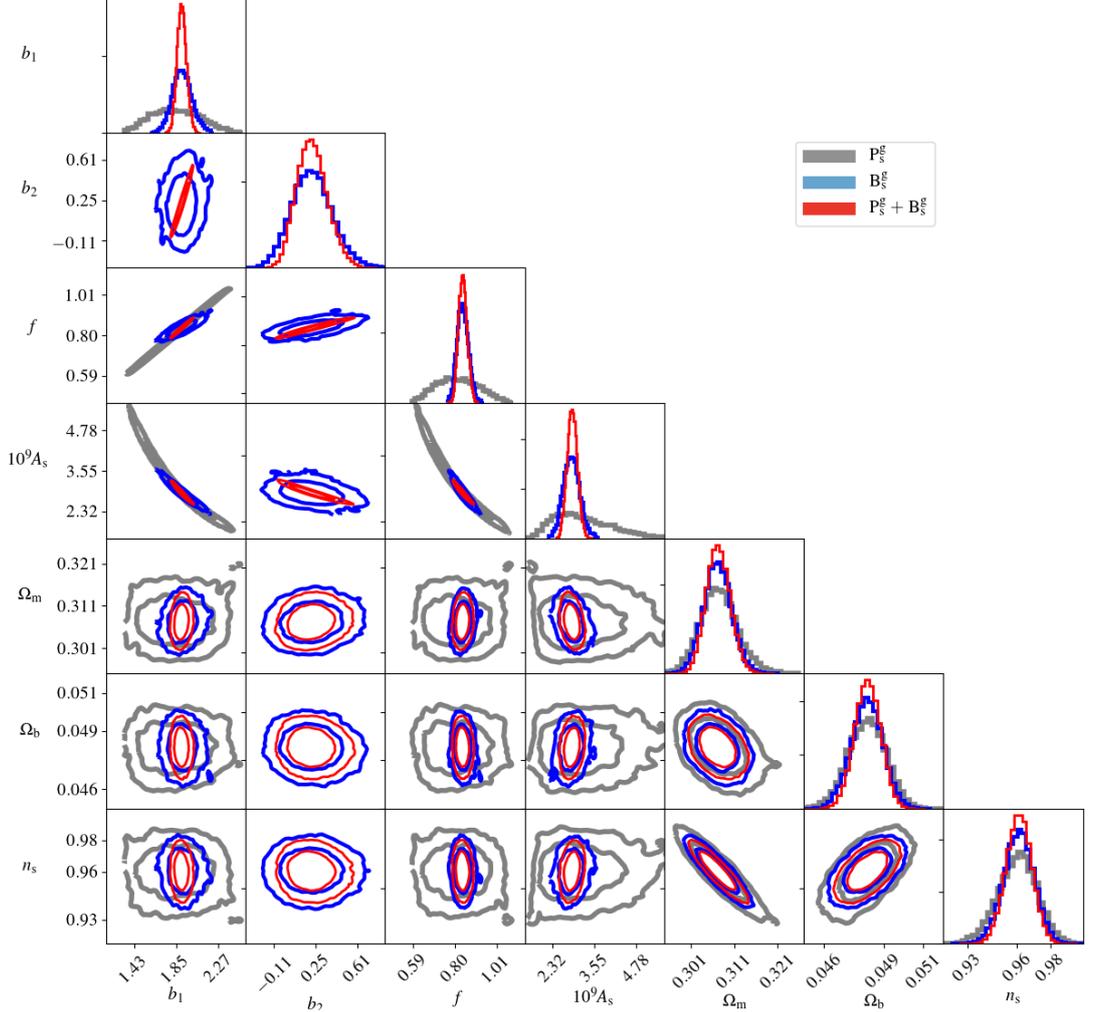


Figure 3.8: Forecasted posteriors for the redshift bin of a DESI-like survey for what concerns the combined two and three points statistics. 1D and 2D contours are derived from MCMC sampling using for the three data-vectors:  $P_g^s$  (grey),  $B_g^s$  (blue) and  $[P_g^s + B_g^s]$  (red). Combining two and three points statistics allows to drastically reduce and break the degeneracies in the parameter space, in particular the ones between  $b_1$ ,  $A_s$  and  $b_2$ . This improvement obtained using the joint data-vector is therefore particularly evident for all these amplitude-like parameters (Table 3.1). In particular it's qualitatively evident from this plot the benefit of adding the bispectrum to the power spectrum. It is important to notice that for the adopted model the galaxy power spectrum does not depend on the parameter  $b_2$ . Even with one parameter less to constrain than the bispectrum, as expected, the MCMC for the galaxy power spectrum alone does not well converge for most of the degenerate parameters.

Gaussian distribution. In our work, this happens when  $k_{\max}$  is lowered, reducing the information accessible through the power spectrum and the bispectrum.

This failure of the method manifests itself with the appearance of unphysical secondary peaks in the posterior distributions. For example, when the MCMC (or MCMC + MP) returns 2D posterior contours with degeneracies that are banana-shaped, like in the case of the joint data-vector  $[\mathbf{P}_g^s, \mathbf{B}_g^s]$  for the parameters  $(b_1, b_2, f, \sigma_8)$ , the compression alone (MP + ORT or PCA + MP) fails to recover the 1D and 2D posteriors. In these cases both the MP + ORT and PCA + MP produce unphysical secondary peaks in the posterior distribution. However, as shown in Figure 3.6 PCA + MP still works considering only larger scales ( $k_{\max}$  lower), where standard perturbation theory gives more accurate predictions, than when using only the MP + ORT method.

In the case of the bispectrum the breakdown of the PCA + MP method happens for  $k_{\max} < 0.18 \text{ Mpc}^{-1}h$  while for the joint data-vector  $[\mathbf{P}_g^s, \mathbf{B}_g^s]$  for  $k_{\max} < 0.17 \text{ Mpc}^{-1}h$ . A potential solution to this limitation could be the application of the Gaussianisation method proposed by [Schuhmann et al. \(2016\)](#) which we plan to include in a future work. This failure also happens when  $V_s$  or  $n_g$  are one order of magnitude smaller. However, these scenarios are below the specifications of current and future cosmological surveys. Another solution is to consider less parameters by either fixing to a fiducial value one or more of the most degenerate ones or by rewriting them in terms of combinations which absorb the degeneracies (e.g.  $f\sigma_8$ ). It is important to note that in this paper we have on purpose considered an extreme case of degenerate parameter space in order to test the method applicability. In a realistic case a much less degenerate parameter space is usually considered when analysing data. In the considered parameter space the main cause of parameters degeneracy is due to having both  $A_s$ ,  $\Omega_m$  and  $f$  instead of using only  $\sigma_8$  and  $f$ . It is then reasonable to assume that the method would still work when a more complex biasing model is considered by adding a further bias parameter  $b_3$  or the tidal bias  $b_{s_2}$ , provided that  $A_s$  and  $\Omega_m$  are substituted with  $\sigma_8$ . A more accurate biasing model would be needed in an application to real data in order to describe the bias non local nature observed in simulations ([Sheth et al., 2013](#); [Modi et al., 2017](#)).

	$\Delta\theta = \frac{\theta_{\max}^{68\%} - \theta_{\min}^{68\%}}{2}$									
	$\Delta b_1$	$\Delta b_2$	$\Delta f$	$\Delta 10^9 A_s$	$\Delta\Omega_m$	$\Delta\Omega_b$	$\Delta n_s$			
$P_g^s$	MCMC	0.251	-	0.111	0.856	0.0041	0.0009	0.0116		
$B_g^s$	MCMC	0.095	0.174	0.028	0.232	0.0031	0.0008	0.0091		
	MCMC + MP	0.090	0.172	0.027	0.220	0.0032	0.0008	0.0093		
	PCA + MP	0.101	0.173	0.029	0.249	0.0032	0.0008	0.0092		
$[P_g^s, B_g^s]$	MCMC	0.047	0.132	0.022	0.151	0.0027	0.0007	0.0079		
	MCMC + MP	0.047	0.134	0.022	0.153	0.0027	0.0007	0.0079		
	" (P <sub>g</sub> <sup>s</sup> uncomp.)	0.048	0.134	0.022	0.152	0.0027	0.0006	0.0078		
	PCA + MP	0.049	0.136	0.022	0.157	0.0027	0.0007	0.0079		
$\% (\Delta\theta^B - \Delta\theta^{P+B}) / \Delta\theta^B$	MCMC	51%	24%	23%	35%	13%	15%	13%		
	MCMC + MP	47%	22%	17%	31%	13%	14%	15%		
	" (P <sub>g</sub> <sup>s</sup> uncomp.)	47%	22%	17%	31%	14%	17%	16%		
	PCA + MP	52%	22%	22%	37%	14%	15%	14%		
$\% (\Delta\theta^P - \Delta\theta^{P+B}) / \Delta\theta^P$	MCMC	81%	-%	80%	82%	34%	27%	32%		
	MCMC + MP	81%	-%	80%	82%	33%	27%	31%		
	" (P <sub>g</sub> <sup>s</sup> uncomp.)	81%	-%	80%	82%	34%	29%	33%		
	PCA + MP	81%	-%	80%	82%	33%	27%	32%		

Table 3.1: The 68% confidence intervals of the 1D posteriors for the bispectrum and the joint data-vector  $[P_g^s, B_g^s]$  obtained from the full data-vectors and from the two possible compression applications, MP + MCMC sampling or PCA + MP compression. Since the marginalised 1D distributions are not perfectly Gaussian (symmetric with respect to the maximum likelihood point) in the Table is reported for each parameter the halved difference between the maximum and minimum values included in the 68% confidence interval. The four rows in the third panel show the improvement in percentage on the constraints gained by using the joint data-vector instead of the bispectrum alone for all the compression methods considered. The percentage shows indicates how much smaller are the constraints obtained using  $[P_g^s, B_g^s]$  as data-vector than the ones computed using just  $B_g^s$ . In the last panel the percentages are relative to the improvement obtained by adding the bispectrum to the power spectrum. The added value of the bispectrum with respect to power spectrum alone consists in 68% confidence intervals  $\sim 4$  times smaller for the amplitude-like parameters  $b_1$ ,  $f$  and  $A_s$ .

### 3.9 Joint data-vector added value

In both cases presented in this paper, either running an MCMC sampling on the compressed data-vector or doing the parameter estimation directly from the compression (MCMC + MP or PCA + MP), the added value of combining the power spectrum with the bispectrum is qualitatively (Figure 3.8) and quantitatively (Figure 3.7 and Table 3.1) evident. This improvement obtained using the joint data-vector is particularly relevant for all the degenerate amplitude-like parameters:  $b_1$ ,  $b_2$ ,  $f$  and  $A_s$ . For the redshift bin of a DESI-like survey considered in our analysis, the improvement for the above mentioned parameters obtained using the joint data-vector with respect using only the bispectrum is respectively of 52%, 22%, 22% and 37% (using PCA + MP). The improvement with respect to the power spectrum alone is even greater, in particular: 81%, 80% and 82% for  $b_1$ ,  $f$  and  $A_s$ .

Comparing our results to Sefusatti et al. (2006) and in particular their Table 8 for  $k_{\max} = 0.2\text{Mpc}^{-1}h$ , it is possible to see a similar effect due to including the bispectrum in the analysis together with the power spectrum. The similarity consists in the fact that the parameters that more benefit from this addition are the bias ones ( $b_1$ ,  $b_2$ ) and the amplitude ones ( $A_s$ ,  $\sigma_8$ ). The greater improvement that we find for certain parameters by adding the bispectrum to the power spectrum with respect to Sefusatti et al. can be explained by the fact that we consider redshift space distortions, which are encoded in the growth rate parameter  $f$ . This increases the degeneracies already present between  $b_1$ ,  $b_2$ ,  $\sigma_8$ ,  $A_s$  and  $\Omega_m$ . Therefore the added value of the bispectrum in lifting the degeneracies increases. Together with this, in Sefusatti et al.'s paper the covariance matrix of the bispectrum includes also off-diagonal terms, which decrease the constraining power of the bispectrum since these terms describe the correlation between different triangle configurations.

For what concerns the most recent bispectrum measurements from BOSS DR12 data by Gil-Marín et al. (2017), it is important to point out that the data vector considered is given by the power spectrum monopole and quadrupoles together with the bispectrum monopole. Using the bispectrum monopole implies washing out some of its information when integrating over the azimuthal and polar angles. This could explain the lower impact of the bispectrum monopole with respect to the power spectrum monopole and quadrupole in constraining the parameters. Moreover in this case the covariance matrix used to do parameter inference is estimated from galaxy mocks, therefore it includes not only off diagonal terms describing the correlation between different triangles, but also the noise due to the fact that it is an estimated quantity.

Finally in the recent paper by [Byun et al. \(2017\)](#) similar improvement to ours have been obtained in parameter constraints (Table 3), in particular for what concerns  $b_1$  and  $\sigma_8$  (proxy for  $A_s$ ).

The efficiency of the compression PCA + MP also drastically improves if ones considers the joint data-vector  $[P_g^s, B_g^s]$  as can be seen in Figures 3.3 and 3.6. Indeed, the combination of power spectrum and bispectrum reduces the degeneracies between the considered parameters and the results obtained by running approximately 20 minutes long compression-pipeline almost perfectly match the ones given by a three days of MCMC sampling run in parallel on 64 processors (right panel Figure 3.7).

As is well known in the literature, the degeneracy between the bias parameters (in particular at the linear order  $b_1$ ) with the amplitude of the dark matter perturbations  $\sigma_8$  or the primordial perturbations scalar amplitude  $A_s$  cannot be broken using only the power spectrum. These degeneracies are even larger when redshift-space distortions are considered. On the other hand, the bispectrum alone can (in theory) lift these degeneracies, even if it requires including the quadratic bias parameter  $b_2$ . In any case being a 3pt statistic, it is more difficult to measure and analyse from real surveys than 2pt statistics. Therefore combining power spectrum and bispectrum is of fundamental importance in order to obtain the best possible constraints, especially in light of the large data sets that are going to become available in the near future (DESI, Euclid, PFS, etc.).

## 3.10 Conclusions

In this paper we have shown that it is possible to compress the information about cosmological parameters contained in the galaxy power spectrum and bispectrum and to extract this information in an efficient way. In particular we have shown:

- Compressing the data-vector using the MOPED algorithm before running an MCMC sampling gives negligible loss of information in terms of parameters constraints (Figures 3.2, 3.3, 3.7). In particular running an MCMC sampling on the compressed data-vector (MCMC + MP) returns 68% confidence intervals less than 1% larger in terms of relative difference than the MCMC's ones. This happens in both the cases of the bispectrum and power spectrum plus bispectrum ( $B_g^s$  and  $[P_g^s, B_g^s]$ ). For real surveys this would allow us to drastically reduce the number of simulations needed to numerically estimate the covariance matrix. This is because the dimension of the compressed covariance matrix corresponds to the number of model parameters, not the number of the original data-vector elements.

- Orthogonalising the parameter space through the diagonalisation of the Fisher matrix before applying the compression (PCA + MP), proves to be competitive with the MCMC sampling with minimal loss of constraining power (Figures 3.5 and 3.6). PCA + MP returns for both  $B_g^s$  and  $[P_g^s, B_g^s]$  68% confidence intervals less than 1% different in terms of relative difference from the ones obtained running the MCMC sampling for the full data-vectors. This method cannot be applied when the parameter space presents non-linear degeneracies, since a multidimensional Gaussian posterior distribution is no longer a valid approximation. Using the joint data-vector  $[P_g^s, B_g^s]$  lowers the minimum  $k_{\max}$  necessary for the only-compression method (PCA + MP) to work to  $k_{\max} \simeq 0.17 \text{ Mpc}^{-1}h$  (while in case of just the bispectrum this threshold is  $k_{\max} \simeq 0.18 \text{ Mpc}^{-1}h$ ). These values have been obtained considering the redshift bin of a DESI-like survey. Moreover these values can be further lowered if the parameter space we considered is reduced to a less degenerate one, for example  $(b_1, b_2, f, \sigma_8)$ .
- [Byun et al. \(2017\)](#) reduced the covariance matrix dimension without significant loss of information by using a proxy that aggregates the matter bispectrum over a subset of Fourier configurations. In particular they obtained their best results in terms of constraints on  $\Lambda$ CDM parameters from combining the power spectrum with the modal decomposition of the bispectrum ([Fergusson et al., 2012](#); [Regan et al., 2012](#)). Their results demonstrate that the modal bispectrum performs as well as the Fourier bispectrum, even with considerably fewer modes (10) than Fourier configurations (95).

The main difference with the approach presented in this work is that we use the original full galaxy bispectrum data-vector in order to compress it. This does not have any limitations in terms of original size of the bispectrum data-vector. Therefore it allows us to access the full information content achievable through the bispectrum. Moreover the number of elements of the compressed data-vector in our case is independent of the number of elements of the original data-vector. In our case we considered  $\sim 1000$  triangle configurations and 7 parameters which gives a compression of at least two orders of magnitude. In [Byun et al. \(2017\)](#) it is left for future work to check whether the achieved compression of the bispectrum data vector by an order of magnitude would improve if more triangle configurations are taken into consideration.

- As already quantified using real data-sets like BOSS ([Gil-Marín et al., 2017](#)), the use of the joint data-vector  $[P_g^s, B_g^s]$  significantly increases the

constraining power on the cosmological parameters compared to using only the power spectrum or bispectrum individually. In the case of an ELGs sample for a DESI-like survey the improvement obtained by combining power spectrum and bispectrum is quantified and can be visualised in the second panel of Figure 3.7 and the second and third rows of Table 3.1. The constraints obtained considering power spectrum and bispectrum together are up to 52% smaller than the constraints obtained using only the bispectrum. The difference is even greater with respect to the power spectrum alone. Together with lifting the degeneracies between amplitude-like parameters as it can be seen in Figure 3.8, the 68% confidence intervals of the marginalised 1D posteriors for the joint data-vector are up to  $\sim 5$  times smaller than the power spectrum ones.

- Using the compression PCA + MP as analysis method is much faster than MCMC and less computationally demanding (few minutes on a single processor compared to days using several processors working in parallel). Since it is relatively easy to implement, it can be used to "sample" different sets of parameters and obtain reliable constraints for a given model without having to wait days for each one of them as in the case of the MCMC.
- The pre-compression PCA transformation allows us to better capture the nature of the degeneracies between the chosen parameters, returning realistic multidimensional posterior distributions that follow closely the MCMC ones. Also in the case of "strong" degeneracies, orthogonalising the parameter space before compressing returns contours qualitatively more realistic and closer to the MCMC than orthogonalising the weights after compressing.

Future work will include the study of how the compression method applied here is affected by the choice of the set of considered parameters. Finally we would like to test the methods presented here using BOSS data or simulations for the upcoming DESI, EUCLID and PFS surveys. Using MCMC + MP or PCA + MP compression methods has the potential of becoming a standard fast and reliable approach to adopt when dealing with large data-vector as in the case of higher-order statistics.

# Appendix

## 3.A Estimators definition and unbiasedness check

In this appendix we present the definition of the power spectrum and the bispectrum estimators, as well as the definition of the trispectrum and the tetraspectrum, which are necessary for the computation of the full covariance matrix for the joint data-vector  $[\mathbf{P}_g^s, \mathbf{B}_g^s]$ .

Including the effect of redshift-space distortions in our analysis requires enlarging the number of parameters needed to describe a particular configuration of the bispectrum. In this case, it is possible to see that, considering the distant observer approximation, which consists in assuming that all the line of sight vectors paired to each wave-vector are parallel between each others, the natural symmetry to exploit is the cylindrical one. In other words, what characterises the redshift effect on the bispectrum are the parallel to the line of sight components of the wave-vectors, from now on labeled as  $k^\parallel$ . Therefore the natural set of coordinates to describe each wave-vector is the cylindrical one:  $(k^\parallel, k^\perp, \phi)$  where  $k^\perp$  and  $\phi$  describe the component of the wave-vector laying on the perpendicular plane to the line of sight. Hence, for what concerns the bispectrum in redshift-space, from the original 9 degrees of freedom, 3 are canceled by translational invariance given by the closed triangle condition  $\delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3)$ .

Moreover, from the chosen coordinates above it is possible to see that there is a further symmetry which is the rotation along the line of sight of a particular triangle of wave-vectors. Different configurations given simply by rotating the same triangle around the line of sight give the same value for the bispectrum. The coordinate  $\phi$  describes these rotations and expresses this symmetry. As a consequence of the symmetries mentioned, we are left with five remaining degrees of freedom, describing all possible configurations, which are given by:  $\{k_1^\perp, k_2^\perp, k_3^\perp, k_1^\parallel, k_2^\parallel\}$ , from which we can derive all the other quantities (e.g.  $k_3^\parallel = -k_1^\parallel - k_2^\parallel$ ). It is necessary then to define a new type of bin for our estimator: a cylindrical anullii defined by  $\phi \in [0, 2\pi]$ ,  $\bar{k}^\perp \in [\bar{k}^\perp - \Delta k^\perp/2, \bar{k}^\perp + \Delta k^\perp/2]$  and  $\bar{k}^\parallel \in [\bar{k}^\parallel - \Delta k^\parallel/2, \bar{k}^\parallel + \Delta k^\parallel/2]$ . For thin anullii the surface area is given

by  $A_c = 2\pi\Delta k \bar{k}^\perp$  and hence the volume is given by  $V^c = 2\pi\Delta k^2 \bar{k}^\perp$ . The 3D Dirac's delta can be decomposed as

$$\delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) = \delta_D(k_1^\parallel + k_2^\parallel + k_3^\parallel) \delta_D(\mathbf{k}_1^\perp + \mathbf{k}_2^\perp + \mathbf{k}_3^\perp). \quad (3.24)$$

### 3.A.1 Power spectrum

A standard way to define an estimator for the power spectrum is the following (Peebles, 1980):

$$\hat{P}_g^s(\bar{\mathbf{k}}_1) = \frac{1}{V_s} \int_{V_1^c} \frac{dV_1^c}{V_1^c} \int_{V_2^c} dV_2^c \delta_D(\mathbf{k}_1 + \mathbf{k}_2) \delta_g^s(\mathbf{k}_1) \delta_g^s(\mathbf{k}_2), \quad (3.25)$$

where  $V_s$  is the survey volume. Notice that both cylindrical bins are centered on  $\bar{\mathbf{k}}_1$  and therefore by definition  $\bar{\mathbf{k}}_1 = -\bar{\mathbf{k}}_2$ . We specify that the power spectrum for the redshift galaxy field depends on the wave-vector, precisely on its perpendicular and parallel components to the line of sight, and not on its modulus. Following the definition given in Equation 3.25, it is straight forward to check whether the estimator is unbiased; one only needs to take the average

$$\begin{aligned} \langle \hat{P}_g^s(\bar{\mathbf{k}}_1) \rangle &= \frac{1}{V_s} \int_{V^c(\mathbf{k}_1)} \frac{dV_1^c}{V^c(\mathbf{k}_1)} \int_{V^c(\mathbf{k}_1)} dV_2^c \delta_D(\mathbf{k}_1 + \mathbf{k}_2) \langle \delta_g^s(\mathbf{k}_1) \delta_g^s(\mathbf{k}_2) \rangle \\ &= \frac{1}{V_s} \int_{V^c(\mathbf{k}_1)} \frac{dV_1^c}{V^c(\mathbf{k}_1)} \int_{V^c(\mathbf{k}_1)} dV_2^c \delta_D(\mathbf{k}_1 + \mathbf{k}_2)^2 (2\pi)^3 P_g^s(\mathbf{k}_1) \\ &= \int_{V^c(\mathbf{k}_1)} \frac{dV_1^c}{V^c(\mathbf{k}_1)} \int_{V^c(\mathbf{k}_1)} dV_2^c \delta_D(\mathbf{k}_1 + \mathbf{k}_2) P_g^s(\mathbf{k}_1) = \int_{V^c(\mathbf{k}_1)} \frac{dV_1^c}{V^c(\mathbf{k}_1)} P_g^s(\mathbf{k}_1) \\ &= \int_{\bar{k}_1^\perp - \frac{\Delta k}{2}}^{\bar{k}_1^\perp + \frac{\Delta k}{2}} \int_{\bar{k}_1^\parallel - \frac{\Delta k}{2}}^{\bar{k}_1^\parallel + \frac{\Delta k}{2}} \frac{dk_i^\perp dk_i^\parallel k_i^\perp}{\bar{k}_i^\perp \Delta k^2} P_g^s(\mathbf{k}_1) \approx P_g^s(\bar{\mathbf{k}}_1). \end{aligned} \quad (3.26)$$

where the standard expression for the galaxy power spectrum at leading order is given by

$$P_g^s(\mathbf{k}_1) = P^m(k_1) F_s^{(1)}(\mathbf{k}_1)^2. \quad (3.27)$$

### 3.A.2 Bispectrum estimator

Proceeding now as shown in Joachimi et al. (2009), we define the following estimator for the bispectrum

$$\begin{aligned} \hat{B}_g^s(\bar{k}_1^\perp, \bar{k}_2^\perp, \bar{k}_3^\perp, \bar{k}_1^\parallel, \bar{k}_2^\parallel) &= \\ &= \frac{(2\pi)^2 \Delta k_3^\parallel}{V} \Lambda^{-1} (\bar{k}_1^\perp, \bar{k}_2^\perp, \bar{k}_3^\perp, \bar{k}_1^\parallel, \bar{k}_2^\parallel) \prod_{i=1}^3 \int_{V_i^c} \frac{dV_i^c}{V_i^c} \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) \delta_g^s(\mathbf{k}_1) \delta_g^s(\mathbf{k}_2) \delta_g^s(\mathbf{k}_3), \end{aligned} \quad (3.28)$$

where  $\Lambda$  is a function related to the fraction of wave-vectors triplets allowed by the triangle condition and defined such that the bispectrum estimator 3.28 is unbiased, which can be checked by taking the average of the estimator, which also it is reported in the appendix. The expression for  $\Lambda$  is

$$\begin{aligned}\Lambda(k_1^\perp, k_2^\perp, k_3^\perp) &= 2\pi \int_0^\infty dr^\perp r^\perp \prod_{i=1}^3 J_0(k_i^\perp r^\perp) \\ &= \frac{4}{\sqrt{2k_1^{\perp 2} k_2^{\perp 2} + 2k_1^{\perp 2} k_3^{\perp 2} + 2k_2^{\perp 2} k_3^{\perp 2} - k_1^{\perp 4} - k_2^{\perp 4} - k_3^{\perp 4}}},\end{aligned}\quad (3.29)$$

if  $|k_1^\perp - k_2^\perp| < k_3^\perp < k_1^\perp + k_2^\perp$  or 0 otherwise.  $J_0$  is the zero-th order spherical Bessel function. It is also possible to check that for the bispectrum the estimator is unbiased

$$\begin{aligned}\langle \hat{B}_g^s(\bar{k}_1^\perp, \bar{k}_2^\perp, \bar{k}_3^\perp, \bar{k}_1^\parallel, \bar{k}_2^\parallel) \rangle &= \\ &= \frac{(2\pi)^2 \Delta k_3^\parallel}{V_s} \Lambda^{-1}(\bar{k}_1^\perp, \bar{k}_2^\perp, \bar{k}_3^\perp, \bar{k}_1^\parallel, \bar{k}_2^\parallel) \\ &\times \int_{V_1^c} \int_{V_2^c} \int_{V_3^c} (2\pi)^3 (\delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3))^2 B_g^s(k_1^\perp, k_2^\perp, k_3^\perp, k_1^\parallel, k_2^\parallel) \\ &= (2\pi)^2 \Delta k_3^\parallel \Lambda_{123}^{-1} \int_{V_1^c} \int_{V_2^c} \int_{V_3^c} \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) B_{g,123}^s,\end{aligned}\quad (3.30)$$

where a shorthand notation for the bispectrum has been introduced and where we have used the approximation  $\delta_D^2 \simeq V_s / (2\pi)^3 \delta_D$  from Joachimi et al. (2009). Since the bispectrum is invariant under rotation around the line of sight, we integrate now the angular part, namely over  $\phi$

$$\begin{aligned}\int_0^{2\pi} d\phi_1 \int_0^{2\pi} d\phi_2 \int_0^{2\pi} d\phi_3 \delta_D(\mathbf{k}_1^\perp + \mathbf{k}_2^\perp + \mathbf{k}_3^\perp) &= \int d\phi_1 d\phi_2 d\phi_3 \int \frac{d^2 \mathbf{r}^\perp}{(2\pi)^2} e^{i(\mathbf{k}_1^\perp + \mathbf{k}_2^\perp + \mathbf{k}_3^\perp) \mathbf{r}^\perp} \\ &= \int \frac{d^2 \mathbf{r}^\perp}{(2\pi)^2} \int_0^{2\pi} d\phi_1 e^{i k_1^\perp r^\perp \cos(\phi_1 - \phi_r)} \\ &\times \int_0^{2\pi} d\phi_2 e^{i k_2^\perp r^\perp \cos(\phi_2 - \phi_r)} \int_0^{2\pi} d\phi_3 e^{i k_3^\perp r^\perp \cos(\phi_3 - \phi_r)} \\ &= \int \frac{d^2 \mathbf{r}^\perp}{(2\pi)^2} (2\pi)^3 J_0(k_1^\perp r^\perp) J_0(k_2^\perp r^\perp) J_0(k_3^\perp r^\perp) \\ &= (2\pi)^2 \int_0^\infty dr^\perp r^\perp \prod_{i=1}^3 J_0(k_i^\perp r^\perp) = 2\pi \Lambda(k_1^\perp, k_2^\perp, k_3^\perp),\end{aligned}\quad (3.31)$$

from which it is possible to see that in this case  $\Lambda$  depends only on the perpendicular components of the wave-vectors. This agrees with the fact that on the orthogonal plane the wave-vectors components must form a closed triangle. In

one of the steps above we used the argument that integrating a trigonometric function over an interval of  $2\pi$  gives the same result no matter what are the extrema of integration as long as the interval is of  $2\pi$ . For completeness the Bessel functions of order 0 are defined as

$$J_0(x) = \int_0^{2\pi} \frac{d\phi}{2\pi} e^{ix \cos \phi}. \quad (3.32)$$

Then, inserting the expression for  $\Lambda$  3.31 into 3.30 it is possible to verify that our estimator is unbiased

$$\begin{aligned} \langle \hat{B}_{g,123}^s \rangle &= (2\pi)^2 \Delta k_3^\parallel \Lambda_{123}^{-1} \prod_{i=1}^3 \int_{V_i^c} \frac{d^2 \mathbf{k}_i^\perp dk_i^\parallel}{2\pi \bar{k}_i^\perp \Delta k^2} \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) B_g^s(k_1^\perp, k_2^\perp, k_3^\perp, k_1^\parallel, k_2^\parallel) \\ &= (2\pi)^2 \Delta k_3^\parallel \Lambda_{123}^{-1} \prod_{i=1}^3 \int_{\bar{k}_i^\perp - \frac{\Delta k}{2}}^{\bar{k}_i^\perp + \frac{\Delta k}{2}} \int_{\bar{k}_i^\parallel - \frac{\Delta k}{2}}^{\bar{k}_i^\parallel + \frac{\Delta k}{2}} \frac{dk_i^\perp dk_i^\parallel k_i^\perp}{2\pi \bar{k}_i^\perp \Delta k^2}, 2\pi \Lambda_{123} \delta_D(k_1^\parallel + k_2^\parallel + k_3^\parallel) B_{g,123}^s \\ &\approx \delta_{123}^K B_g^s(\bar{k}_1^\perp, \bar{k}_2^\perp, \bar{k}_3^\perp, \bar{k}_1^\parallel, \bar{k}_2^\parallel), \end{aligned} \quad (3.33)$$

where  $\delta^K$  is a Kronecker delta. In the last step the thin-shell approximation has been used in order to bring out the bispectrum from the integrals. The standard expression for the galaxy bispectrum is given by

$$B_g^s(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = 2 P^m(k_1) P^m(k_2) F_s^{(1)}(\mathbf{k}_1) F_s^{(1)}(\mathbf{k}_2) F_s^{(2)}[\mathbf{k}_1, \mathbf{k}_2] + 2 \text{p.} \quad (3.34)$$

### 3.A.3 Trispectrum definition

The trispectrum is defined as

$$\langle \delta_g^s(\mathbf{k}_1) \delta_g^s(\mathbf{k}_2) \delta_g^s(\mathbf{k}_3) \delta_g^s(\mathbf{k}_4) \rangle_c = (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4) T_g^s(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4), \quad (3.35)$$

where the subscript "c" indicates that trispectrum is the connected part of the four points correlation function. Therefore the leading order terms are of two types:  $T_g^{s(2)}$  and  $T_g^{s(3)}$  respectively characterised by the RSD perturbation kernels  $F_s^{(2)}$  and  $F_s^{(3)}$ . Starting from the first one we have that:

$$\begin{aligned}
 & (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4) T_g^{s(2)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) = \\
 & = \langle F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] \delta_{\mathbf{k}_1} \delta_{\mathbf{k}_2} \frac{1}{(2\pi)^3} \int d^3 k_a d^3 k_b \delta_D(\mathbf{k}_a + \mathbf{k}_b - \mathbf{k}_3) F_s^{(2)}[\mathbf{k}_a, \mathbf{k}_b] \delta_{\mathbf{k}_a} \delta_{\mathbf{k}_b} \\
 & \times \int d^3 k_c d^3 k_d \delta_D(\mathbf{k}_c + \mathbf{k}_d - \mathbf{k}_4) F_s^{(2)}[\mathbf{k}_c, \mathbf{k}_d] \delta_{\mathbf{k}_c} \delta_{\mathbf{k}_d} \rangle + 5 \text{ p.} \\
 & = \frac{1}{(2\pi)^6} F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] \int d^3 k_a d^3 k_b \delta_D(\mathbf{k}_a + \mathbf{k}_b - \mathbf{k}_3) F_s^{(2)}[\mathbf{k}_a, \mathbf{k}_b] \\
 & \times \int d^3 k_c d^3 k_d \delta_D(\mathbf{k}_c + \mathbf{k}_d - \mathbf{k}_4) F_s^{(2)}[\mathbf{k}_c, \mathbf{k}_d] \langle \delta_{\mathbf{k}_1} \delta_{\mathbf{k}_2} \delta_{\mathbf{k}_a} \delta_{\mathbf{k}_b} \delta_{\mathbf{k}_c} \delta_{\mathbf{k}_d} \rangle + 5 \text{ p.} \quad (3.36)
 \end{aligned}$$

From the last line of the previous expression it follows that, when Wick's theorem is applied, there are three different ways to pair the wave-vectors.

$$T_g^{s(2)} \begin{cases} T_g^{s(2a)} & \Leftrightarrow \langle \delta_{\mathbf{k}_1} \delta_{\mathbf{k}_2} \rangle \langle \delta_{\mathbf{k}_a} \delta_{\mathbf{k}_c} \rangle \langle \delta_{\mathbf{k}_b} \delta_{\mathbf{k}_d} \rangle \times 2 \quad (\text{sym.}) \\ T_g^{s(2b)} & \Leftrightarrow \langle \delta_{\mathbf{k}_1} \delta_{\mathbf{k}_a} \rangle \langle \delta_{\mathbf{k}_2} \delta_{\mathbf{k}_c} \rangle \langle \delta_{\mathbf{k}_b} \delta_{\mathbf{k}_d} \rangle \times 4 \quad (\text{sym.}) \\ T_g^{s(2c)} & \Leftrightarrow \langle \delta_{\mathbf{k}_1} \delta_{\mathbf{k}_c} \rangle \langle \delta_{\mathbf{k}_2} \delta_{\mathbf{k}_a} \rangle \langle \delta_{\mathbf{k}_b} \delta_{\mathbf{k}_d} \rangle \times 4 \quad (\text{sym.}) \end{cases}$$

The first term represents a non connected 1-loop correction to the power spectrum covariance matrix (Fry, 1984; Mohammed et al., 2017). For completeness, we just show below that the simplified expression makes explicit the fact that it is an unconnected term of the full 4-points correlator

$$\begin{aligned}
 & (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4) T_g^{s(2a)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) = \\
 & = \frac{2}{(2\pi)^6} F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] \int d^3 k_a d^3 k_b d^3 k_c d^3 k_d \delta_D(\mathbf{k}_a + \mathbf{k}_b - \mathbf{k}_3) \delta_D(\mathbf{k}_c + \mathbf{k}_d - \mathbf{k}_4) \\
 & \times F_s^{(2)}[\mathbf{k}_a, \mathbf{k}_b] F_s^{(2)}[\mathbf{k}_c, \mathbf{k}_d] (2\pi)^9 \delta_D(\mathbf{k}_1 + \mathbf{k}_2) \delta_D(\mathbf{k}_a + \mathbf{k}_c) \delta_D(\mathbf{k}_b + \mathbf{k}_d) P_{k_1}^m P_{k_a}^m P_{k_b}^m + 5 \text{ p.} \\
 & = 2(2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2) P_g^s(\mathbf{k}_1) \int d^3 k_a d^3 k_b F_s^{(2)}[\mathbf{k}_a, \mathbf{k}_b] F_s^{(2)}[-\mathbf{k}_a, -\mathbf{k}_b] \\
 & \times \delta_D(\mathbf{k}_a + \mathbf{k}_b - \mathbf{k}_3) \delta_D(\mathbf{k}_3 + \mathbf{k}_4) P_{k_a}^m P_{k_b}^m + 5 \text{ p.} \\
 & = 2(2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2) \delta_D(\mathbf{k}_3 + \mathbf{k}_4) P_g^s(\mathbf{k}_1) \int d^3 k_a F_s^{(2)}[\mathbf{k}_a, \mathbf{k}_3 - \mathbf{k}_a]^2 P_{k_a}^m P_{|\mathbf{k}_3 - \mathbf{k}_a|}^m + 5 \text{ p.} \\
 & = 2(2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4) \delta_D(\mathbf{k}_3 + \mathbf{k}_4) P_g^s(\mathbf{k}_1) \\
 & \times \int d^3 k_a F_s^{(2)}[\mathbf{k}_a, \mathbf{k}_3 - \mathbf{k}_a]^2 P_{k_a}^m P_{|\mathbf{k}_3 - \mathbf{k}_a|}^m + 5 \text{ p.} \quad (3.37)
 \end{aligned}$$

We then look at the first connected tree level term  $T_g^{s(2b)}$

$$\begin{aligned}
 & (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4) T_g^{s(2b)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) = \\
 & = \frac{4}{(2\pi)^6} F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] \int d^3 k_a d^3 k_b d^3 k_c d^3 k_d \delta_D(\mathbf{k}_a + \mathbf{k}_b - \mathbf{k}_3) \delta_D(\mathbf{k}_c + \mathbf{k}_d - \mathbf{k}_4) \\
 & \times F_s^{(2)}[\mathbf{k}_a, \mathbf{k}_b] F_s^{(2)}[\mathbf{k}_c, \mathbf{k}_d] (2\pi)^9 \delta_D(\mathbf{k}_1 + \mathbf{k}_a) \delta_D(\mathbf{k}_2 + \mathbf{k}_c) \delta_D(\mathbf{k}_b + \mathbf{k}_d) P_{k_1}^m P_{k_a}^m P_{k_b}^m + 5 \text{ p.} \\
 & = 4 \times (2\pi)^3 F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] P_{k_1}^m P_{k_2}^m \int d^3 k_b d^3 k_d \delta_D(\mathbf{k}_b - \mathbf{k}_1 - \mathbf{k}_3) \delta_D(\mathbf{k}_d - \mathbf{k}_2 - \mathbf{k}_4) \\
 & \times F_s^{(2)}[-\mathbf{k}_1, \mathbf{k}_1 + \mathbf{k}_3] F_s^{(2)}[-\mathbf{k}_2, \mathbf{k}_2 + \mathbf{k}_4] P_{k_b}^m \delta_D(\mathbf{k}_b + \mathbf{k}_d) + 5 \text{ p.} \\
 & = 4 \times (2\pi)^3 F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] F_s^{(2)}[-\mathbf{k}_1, \mathbf{k}_1 + \mathbf{k}_3] F_s^{(2)}[-\mathbf{k}_2, -\mathbf{k}_1 - \mathbf{k}_3] \\
 & \times \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4) P_{k_1}^m P_{k_2}^m P_{|\mathbf{k}_1 + \mathbf{k}_3|}^m + 5 \text{ p.} \tag{3.38}
 \end{aligned}$$

from which it is possible to directly write the third term which result to be exactly the same

$$\begin{aligned}
 & (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4) T_g^{s(2c)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) = \\
 & = 4 \times (2\pi)^3 F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] F_s^{(2)}[-\mathbf{k}_2, \mathbf{k}_2 + \mathbf{k}_3] F_s^{(2)}[-\mathbf{k}_1, -\mathbf{k}_2 - \mathbf{k}_3] \\
 & \times \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4) P_{k_1}^m P_{k_2}^m P_{|\mathbf{k}_2 + \mathbf{k}_3|}^m + 5 \text{ p.} \tag{3.39}
 \end{aligned}$$

Finally it is possible to consider the last term  $T_g^{s(3)}$

$$\begin{aligned}
 & (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4) T_g^{s(3)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) = \\
 & = \langle F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] F_s^{(1)}[\mathbf{k}_3] \delta_{\mathbf{k}_1} \delta_{\mathbf{k}_2} \delta_{\mathbf{k}_3} \rangle \\
 & \times \frac{1}{(2\pi)^6} \int d^3 k_a d^3 k_b d^3 k_c \delta_D(\mathbf{k}_a + \mathbf{k}_b + \mathbf{k}_c - \mathbf{k}_4) F_s^{(3)}[\mathbf{k}_a, \mathbf{k}_b, \mathbf{k}_c] \delta_{\mathbf{k}_a} \delta_{\mathbf{k}_b} \delta_{\mathbf{k}_c} \rangle + 3 \text{ p.} \\
 & = \frac{1}{(2\pi)^6} F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] F_s^{(1)}[\mathbf{k}_3] \int d^3 k_a d^3 k_b d^3 k_c \delta_D(\mathbf{k}_a + \mathbf{k}_b + \mathbf{k}_c - \mathbf{k}_4) F_s^{(3)}[\mathbf{k}_a, \mathbf{k}_b, \mathbf{k}_c] \\
 & \times \langle \delta_{\mathbf{k}_1} \delta_{\mathbf{k}_2} \delta_{\mathbf{k}_3} \delta_{\mathbf{k}_a} \delta_{\mathbf{k}_b} \delta_{\mathbf{k}_c} \rangle + 3 \text{ p.} \tag{3.40}
 \end{aligned}$$

Like we did before, we apply now the Wick's theorem. In this case there are two different ways to pair the wave-vectors.

$$T_g^{s(3)} \begin{cases} T_g^{s(3a)} & \Leftrightarrow (\langle \delta_{\mathbf{k}_1} \delta_{\mathbf{k}_2} \rangle \langle \delta_{\mathbf{k}_3} \delta_{\mathbf{k}_a} \rangle \langle \delta_{\mathbf{k}_b} \delta_{\mathbf{k}_c} \rangle + 2 \text{ p.}) \times 3 (\text{sym.}) \\ T_g^{s(3b)} & \Leftrightarrow \langle \delta_{\mathbf{k}_1} \delta_{\mathbf{k}_a} \rangle \langle \delta_{\mathbf{k}_2} \delta_{\mathbf{k}_b} \rangle \langle \delta_{\mathbf{k}_3} \delta_{\mathbf{k}_c} \rangle \times 6 (\text{sym.}) \end{cases}$$

The first option represents again a non connected 1-loop correction to the power spectrum covariance matrix.

$$\begin{aligned}
 & (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4) T_g^{s(3a)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) = \\
 & = \frac{3}{(2\pi)^6} F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] F_s^{(1)}[\mathbf{k}_2] \int d^3 k_a d^3 k_b d^3 k_c \delta_D(\mathbf{k}_a + \mathbf{k}_b + \mathbf{k}_c - \mathbf{k}_4) F_s^{(3)}[\mathbf{k}_a, \mathbf{k}_b, \mathbf{k}_c] \\
 & \times (2\pi)^9 \delta_D(\mathbf{k}_1 + \mathbf{k}_2) \delta_D(\mathbf{k}_3 + \mathbf{k}_a) \delta_D(\mathbf{k}_b + \mathbf{k}_c) P_{k_1}^m P_{k_3}^m P_{k_b}^m + 2p. \\
 & = 3 \times (2\pi)^3 P_g^s(\mathbf{k}_1) F_s^{(1)}[\mathbf{k}_3] P_{k_3}^m \delta_D(\mathbf{k}_1 + \mathbf{k}_2) \int d^3 k_a d^3 k_b d^3 k_c \delta_D(\mathbf{k}_a + \mathbf{k}_b + \mathbf{k}_c - \mathbf{k}_4) \\
 & \times F_s^{(3)}[\mathbf{k}_a, \mathbf{k}_b, \mathbf{k}_c] \delta_D(\mathbf{k}_3 + \mathbf{k}_a) \delta_D(\mathbf{k}_b + \mathbf{k}_c) P_{k_b}^m + 2p. \\
 & = 3 \times (2\pi)^3 P_g^s(\mathbf{k}_1) F_s^{(1)}[\mathbf{k}_3] P_{k_3}^m \delta_D(\mathbf{k}_1 + \mathbf{k}_2) \int d^3 k_b d^3 k_c \delta_D(\mathbf{k}_b + \mathbf{k}_c - \mathbf{k}_3 - \mathbf{k}_4) \\
 & \times F_s^{(3)}[-\mathbf{k}_3, \mathbf{k}_b, \mathbf{k}_c] \delta_D(\mathbf{k}_b + \mathbf{k}_c) P_{k_b}^m + 2p. \\
 & = 3 \times (2\pi)^3 P_g^s(\mathbf{k}_1) F_s^{(1)}[\mathbf{k}_3] P_{k_3}^m \delta_D(\mathbf{k}_1 + \mathbf{k}_2) \delta_D(\mathbf{k}_3 + \mathbf{k}_4) \\
 & \times \int d^3 k_b F_s^{(3)}[-\mathbf{k}_3, \mathbf{k}_b, -\mathbf{k}_b] P_{k_b}^m + 2p. \\
 & = 3 \times (2\pi)^3 P_g^s(\mathbf{k}_1) F_s^{(1)}[\mathbf{k}_3] P_{k_3}^m \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4) \delta_D(\mathbf{k}_3 + \mathbf{k}_4) \\
 & \times \int d^3 k_b F_s^{(3)}[-\mathbf{k}_3, \mathbf{k}_b, -\mathbf{k}_b] P_{k_b}^m + 2p.. \tag{3.41}
 \end{aligned}$$

From the last line it is clear that this is also an un-connected term. The second term gives the tree level contribution

$$\begin{aligned}
 & (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4) T_g^{s(3b)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) = \\
 & = \frac{6}{(2\pi)^6} F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] F_s^{(1)}[\mathbf{k}_2] \int d^3 k_a d^3 k_b d^3 k_c \delta_D(\mathbf{k}_a + \mathbf{k}_b + \mathbf{k}_c - \mathbf{k}_4) F_s^{(3)}[\mathbf{k}_a, \mathbf{k}_b, \mathbf{k}_c] \\
 & \times (2\pi)^9 \delta_D(\mathbf{k}_1 + \mathbf{k}_a) \delta_D(\mathbf{k}_2 + \mathbf{k}_b) \delta_D(\mathbf{k}_3 + \mathbf{k}_c) P_{k_1}^m P_{k_2}^m P_{k_3}^m + 3p. \\
 & = 6 \times (2\pi)^3 F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] F_s^{(1)}[\mathbf{k}_3] \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4) \\
 & \times F_s^{(3)}[-\mathbf{k}_1, -\mathbf{k}_2, -\mathbf{k}_3] P_{k_1}^m P_{k_2}^m P_{k_3}^m + 3p.. \tag{3.42}
 \end{aligned}$$

Therefore the full trispectrum at tree level is given by

$$T_g^s = T_g^{s(2b)} + T_g^{s(2c)} + T_g^{s(3b)}. \tag{3.43}$$

### 3.A.4 Tetraspectrum definition

The tetraspectrum is defined as

$$\begin{aligned}
 & \langle \delta_g^s(\mathbf{k}_1) \delta_g^s(\mathbf{k}_2) \delta_g^s(\mathbf{k}_3) \delta_g^s(\mathbf{k}_4) \delta_g^s(\mathbf{k}_5) \rangle_c = \\
 & = (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4 + \mathbf{k}_5) T_{e,g}^s(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4, \mathbf{k}_5), \tag{3.44}
 \end{aligned}$$

where the subscript "c" stands for the connected part of the five points correlation function in Fourier space.

Drawing the diagrams as shown in Fry (1984), it is possible to see that the leading term of the connected part has order  $O(\delta_m^8)$ . There are two possible ways through which  $O(\delta_m^8)$  can be reached, either by having  $F_s^{(2)^3}$  or  $F_s^{(2)}F_s^{(3)}$  terms:

$$\begin{aligned}
 & (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4 + \mathbf{k}_5) T_{e,g}^s(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4, \mathbf{k}_5) = \\
 & = \frac{1}{(2\pi)^9} F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] \int d\mathbf{k}_a^3 d\mathbf{k}_b^3 \delta_D(\mathbf{k}_a + \mathbf{k}_b - \mathbf{k}_3) F_s^{(2)}[\mathbf{k}_a, \mathbf{k}_b] \\
 & \times \int d\mathbf{k}_c^3 d\mathbf{k}_d^3 \delta_D(\mathbf{k}_c + \mathbf{k}_d - \mathbf{k}_4) F_s^{(2)}[\mathbf{k}_c, \mathbf{k}_d] \\
 & \times \int d\mathbf{k}_e^3 d\mathbf{k}_f^3 \delta_D(\mathbf{k}_e + \mathbf{k}_f - \mathbf{k}_5) F_s^{(2)}[\mathbf{k}_e, \mathbf{k}_f] \langle \delta_1 \delta_2 \delta_a \delta_b \delta_c \delta_d \delta_e \delta_f \rangle + 59 \text{ p.} \\
 & + \frac{1}{(2\pi)^9} F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] F_s^{(1)}[\mathbf{k}_3] \int d\mathbf{k}_a^3 d\mathbf{k}_b^3 \delta_D(\mathbf{k}_a + \mathbf{k}_b - \mathbf{k}_4) F_s^{(2)}[\mathbf{k}_a, \mathbf{k}_b] \\
 & \times \int d\mathbf{k}_c^3 d\mathbf{k}_d^3 d\mathbf{k}_e^3 \delta_D(\mathbf{k}_c + \mathbf{k}_d + \mathbf{k}_e - \mathbf{k}_5) F_s^{(3)}[\mathbf{k}_c, \mathbf{k}_d, \mathbf{k}_e] \langle \delta_1 \delta_2 \delta_3 \delta_a \delta_b \delta_c \delta_d \delta_e \rangle + 19 \text{ p..} \\
 & \hspace{15em} (3.45)
 \end{aligned}$$

Starting from the first term that appears in the above expansion:

$$\begin{aligned}
 & (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4 + \mathbf{k}_5) T_{e,g}^{s(a)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4, \mathbf{k}_5) = \\
 & = \frac{1}{(2\pi)^9} F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] \int d\mathbf{k}_a^3 d\mathbf{k}_b^3 d\mathbf{k}_c^3 d\mathbf{k}_d^3 d\mathbf{k}_e^3 d\mathbf{k}_f^3 \delta_D(\mathbf{k}_a + \mathbf{k}_b - \mathbf{k}_3) \delta_D(\mathbf{k}_c + \mathbf{k}_d - \mathbf{k}_4) \\
 & \times \delta_D(\mathbf{k}_e + \mathbf{k}_f - \mathbf{k}_5) F_s^{(2)}[\mathbf{k}_a, \mathbf{k}_b] F_s^{(2)}[\mathbf{k}_c, \mathbf{k}_d] F_s^{(2)}[\mathbf{k}_e, \mathbf{k}_f] \langle \delta_1 \delta_2 \delta_a \delta_b \delta_c \delta_d \delta_e \delta_f \rangle + 59 \text{ p.} \\
 & = 8 (2\pi)^3 F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] \int d\mathbf{k}_a^3 d\mathbf{k}_b^3 d\mathbf{k}_c^3 d\mathbf{k}_d^3 d\mathbf{k}_e^3 d\mathbf{k}_f^3 \delta_D(\mathbf{k}_a + \mathbf{k}_b - \mathbf{k}_3) \\
 & \times \delta_D(\mathbf{k}_c + \mathbf{k}_d - \mathbf{k}_4) \delta_D(\mathbf{k}_e + \mathbf{k}_f - \mathbf{k}_5) \delta_D(\mathbf{k}_1 + \mathbf{k}_f) \delta_D(\mathbf{k}_2 + \mathbf{k}_a) \delta_D(\mathbf{k}_b + \mathbf{k}_c) \delta_D(\mathbf{k}_d + \mathbf{k}_e) \\
 & \times F_s^{(2)}[\mathbf{k}_a, \mathbf{k}_b] F_s^{(2)}[\mathbf{k}_c, \mathbf{k}_d] F_s^{(2)}[\mathbf{k}_e, \mathbf{k}_f] P^m(k_1) P^m(k_2) P^m(k_b) P^m(k_c) + 359 \text{ p.} \\
 & = 8 (2\pi)^3 F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] \int d\mathbf{k}_b^3 d\mathbf{k}_d^3 \delta_D(-\mathbf{k}_2 + \mathbf{k}_b - \mathbf{k}_3) \delta_D(-\mathbf{k}_b + \mathbf{k}_d - \mathbf{k}_4) \\
 & \times \delta_D(-\mathbf{k}_d - \mathbf{k}_1 - \mathbf{k}_5) F_s^{(2)}[-\mathbf{k}_2, \mathbf{k}_b] F_s^{(2)}[-\mathbf{k}_b, \mathbf{k}_d] F_s^{(2)}[-\mathbf{k}_d, -\mathbf{k}_1] \\
 & \times P^m(k_1) P^m(k_2) P^m(k_b) P^m(k_b) + 359 \text{ p.} \\
 & = (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4 + \mathbf{k}_5) 8 F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] F_s^{(2)}[-\mathbf{k}_2, \mathbf{k}_2 + \mathbf{k}_3] \\
 & \times F_s^{(2)}[-\mathbf{k}_2 - \mathbf{k}_3, -\mathbf{k}_1 - \mathbf{k}_5] F_s^{(2)}[\mathbf{k}_1 + \mathbf{k}_5, -\mathbf{k}_1] \\
 & \times P^m(k_1) P^m(k_2) P^m(|\mathbf{k}_2 + \mathbf{k}_3|) P^m(|\mathbf{k}_2 + \mathbf{k}_3|) + 359 \text{ p..} \\
 & \hspace{15em} (3.46)
 \end{aligned}$$

where the factor of 8 comes from permutations which do not influence the relations between the starting five wave vectors. On the contrary for each the original 60 permutations, there are 6 for which the relation between the initial wave vectors varies. Considering now the second term:

$$\begin{aligned}
 & (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4 + \mathbf{k}_5) T_{e,g}^{s(b)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4, \mathbf{k}_5) = \\
 & = \frac{1}{(2\pi)^9} F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] F_s^{(1)}[\mathbf{k}_3] \int d\mathbf{k}_a^3 d\mathbf{k}_b^3 d\mathbf{k}_c^3 d\mathbf{k}_d^3 d\mathbf{k}_e^3 \delta_D(\mathbf{k}_a + \mathbf{k}_b - \mathbf{k}_4) \\
 & \times \delta_D(\mathbf{k}_c + \mathbf{k}_d + \mathbf{k}_e - \mathbf{k}_5) F_s^{(2)}[\mathbf{k}_a, \mathbf{k}_b] F_s^{(3)}[\mathbf{k}_c, \mathbf{k}_d, \mathbf{k}_e] \langle \delta_1 \delta_2 \delta_3 \delta_a \delta_b \delta_c \delta_d \delta_e \rangle + 19 \text{ p.} \\
 & = 6 (2\pi)^3 F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] F_s^{(1)}[\mathbf{k}_3] \int d\mathbf{k}_a^3 d\mathbf{k}_b^3 d\mathbf{k}_c^3 d\mathbf{k}_d^3 d\mathbf{k}_e^3 \delta_D(\mathbf{k}_a + \mathbf{k}_b - \mathbf{k}_4) \\
 & \times \delta_D(\mathbf{k}_c + \mathbf{k}_d + \mathbf{k}_e - \mathbf{k}_5) \delta_D(\mathbf{k}_1 + \mathbf{k}_e) \delta_D(\mathbf{k}_2 + \mathbf{k}_d) \delta_D(\mathbf{k}_3 + \mathbf{k}_b) \delta_D(\mathbf{k}_a + \mathbf{k}_c) \\
 & \times F_s^{(2)}[\mathbf{k}_a, \mathbf{k}_b] F_s^{(3)}[\mathbf{k}_c, \mathbf{k}_d, \mathbf{k}_e] P^m(k_1) P^m(k_2) P^m(k_3) P^m(k_a) + 119 \text{ p.} \\
 & = 6 (2\pi)^3 F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] F_s^{(1)}[\mathbf{k}_3] \int d\mathbf{k}_a^3 \delta_D(\mathbf{k}_a - \mathbf{k}_3 - \mathbf{k}_4) \delta_D(-\mathbf{k}_a - \mathbf{k}_2 - \mathbf{k}_1 - \mathbf{k}_5) \\
 & \times F_s^{(2)}[\mathbf{k}_a, -\mathbf{k}_3] F_s^{(3)}[-\mathbf{k}_a, -\mathbf{k}_2, -\mathbf{k}_1] P^m(k_1) P^m(k_2) P^m(k_3) P^m(k_a) + 119 \text{ p.} \\
 & = (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4 + \mathbf{k}_5) 6 F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] F_s^{(1)}[\mathbf{k}_3] F_s^{(2)}[\mathbf{k}_3 + \mathbf{k}_4, -\mathbf{k}_3] \\
 & \times F_s^{(3)}[-\mathbf{k}_3 - \mathbf{k}_4, -\mathbf{k}_2, -\mathbf{k}_1] P^m(k_1) P^m(k_2) P^m(k_3) P^m(|\mathbf{k}_3 + \mathbf{k}_4|) + 119 \text{ p.} \\
 & \hspace{15em} (3.47)
 \end{aligned}$$

where the factor of 6 comes from permutations which do not influence the relations between the starting five wave vectors. On the contrary for each the original 19 permutations, there are 6 for which the relation between the initial wave vectors varies. From the above expressions of both  $T_{e,g}^{s(a)}$  and  $T_{e,g}^{s(b)}$  it is possible to see that at leading order the tetraspectrum has order  $O(\delta_m^8)$ . In other words it is proportional to the fourth power of the linear matter power spectrum.

### 3.A.5 Unconnected part of the five points correlation function

If one looks at the unconnected part of the five point correlation function, the leading term has order  $O(\delta_m^6)$ . Below are derived the analytical expression for this unconnected part.

$$\begin{aligned}
 & U_{5p,g}^s(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4, \mathbf{k}_5) = \\
 & = \frac{1}{(2\pi)^3} \prod_{i=1}^4 F_s^{(1)}[\mathbf{k}_i] \int d\mathbf{k}_a^3 d\mathbf{k}_b^3 \delta_D(\mathbf{k}_a + \mathbf{k}_b - \mathbf{k}_5) F_s^{(2)}[\mathbf{k}_a, \mathbf{k}_b] \langle \delta_1 \delta_2 \delta_3 \delta_4 \delta_a \delta_b \rangle + 5 \text{ p.} \\
 & \hspace{15em} (3.48)
 \end{aligned}$$

Using Wick's theorem the six-points correlator can be approximated using in to products of two points correlators. For example, for the first permutation one obtains

$$\begin{aligned}
 \langle \delta_1 \delta_2 \delta_3 \delta_4 \delta_a \delta_b \rangle &= 2 \times \langle \delta_1 \delta_a \rangle \langle \delta_2 \delta_b \rangle \langle \delta_3 \delta_4 \rangle \leftrightarrow U_{5p}^1 \\
 &+ 2 \times \langle \delta_1 \delta_a \rangle \langle \delta_3 \delta_b \rangle \langle \delta_2 \delta_4 \rangle \leftrightarrow U_{5p}^2 \\
 &+ 2 \times \langle \delta_1 \delta_a \rangle \langle \delta_4 \delta_b \rangle \langle \delta_2 \delta_3 \rangle \leftrightarrow U_{5p}^3 \\
 &+ 2 \times \langle \delta_2 \delta_a \rangle \langle \delta_3 \delta_b \rangle \langle \delta_1 \delta_4 \rangle \leftrightarrow U_{5p}^4 \\
 &+ 2 \times \langle \delta_2 \delta_a \rangle \langle \delta_4 \delta_b \rangle \langle \delta_1 \delta_3 \rangle \leftrightarrow U_{5p}^5 \\
 &+ 2 \times \langle \delta_3 \delta_a \rangle \langle \delta_4 \delta_b \rangle \langle \delta_1 \delta_2 \rangle \leftrightarrow U_{5p}^6. \tag{3.49}
 \end{aligned}$$

Since all these terms are similar to each other, we compute below only the first one of them as an example on how to derive the others.

$$\begin{aligned}
 (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4 + \mathbf{k}_5) U_{5p}^1 &= \\
 &= \frac{2}{2(\pi)^3} \prod_{i=1}^4 F_s^{(1)}(\mathbf{k}_i) \int d\mathbf{k}_a^3 d\mathbf{k}_b^3 \delta_D(\mathbf{k}_a + \mathbf{k}_b - \mathbf{k}_5) F_s^{(2)}[\mathbf{k}_a, \mathbf{k}_b] (2\pi)^9 \\
 &\times \delta_D(\mathbf{k}_1 + \mathbf{k}_a) \delta_D(\mathbf{k}_2 + \mathbf{k}_b) \delta_D(\mathbf{k}_3 + \mathbf{k}_4) P^m(k_1) P^m(k_2) P^m(k_3) \\
 &= 2 \times (2\pi)^6 F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] F_s^{(1)}[\mathbf{k}_3] F_s^{(1)}[\mathbf{k}_4] F_s^{(2)}[\mathbf{k}_1, \mathbf{k}_2] \delta_D(\mathbf{k}_3 + \mathbf{k}_4) \\
 &\times \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_5) P^m(k_1) P^m(k_2) P^m(k_3) \\
 &= 2 \times (2\pi)^6 F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] F_s^{(1)}[\mathbf{k}_3] F_s^{(1)}[\mathbf{k}_4] F_s^{(2)}[\mathbf{k}_1, \mathbf{k}_2] \delta_{125}^K \delta_D(\mathbf{k}_3 + \mathbf{k}_4) \\
 &\times \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4 + \mathbf{k}_5) P^m(k_1) P^m(k_2) P^m(k_3), \tag{3.50}
 \end{aligned}$$

where it is important to note that, in the last line, when substituting the Dirac's delta for the sum of three wave-vectors with the sum of all five of them, a Kronecker's delta must be added to keep track of the original relation between the three  $\mathbf{k}$ 's. From the last line one can immediately write another one of the six terms

$$\begin{aligned}
 (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4 + \mathbf{k}_5) U_{5p}^5(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4, \mathbf{k}_5) &= \\
 &= 2(2\pi)^6 F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] F_s^{(1)}[\mathbf{k}_3] F_s^{(1)}[\mathbf{k}_4] F_s^{(2)}[\mathbf{k}_2, \mathbf{k}_4] \\
 &\times \delta_{245}^K \delta_D(\mathbf{k}_1 + \mathbf{k}_3) \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4 + \mathbf{k}_5) P^m(k_2) P^m(\mathbf{k}_4) P^m(k_3). \tag{3.51}
 \end{aligned}$$

Therefore, it is possible to write down the complete expression that takes into account all the permutations and terms as follows

$$\begin{aligned}
 U_{5p,g}^s(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4, \mathbf{k}_5) &= \sum_{i=1}^5 U_{5p,g}^s(\delta_i^{(2)}) \text{ where } U_{5p,g}^s(\delta_i^{(2)}) \\
 &= U_{5p}^1 + U_{5p}^2 + U_{5p}^3 + U_{5p}^4 + U_{5p}^5 + U_{5p}^6. \tag{3.52}
 \end{aligned}$$

The argument  $\delta_i^{(2)}$  identifies which wave-vectors corresponds to the second order perturbation expansion for each term.

## 3.B Covariance terms derivation

In this Appendix we present the derivation of the covariance terms, starting from the power spectrum one. Theoretically one could compute all the terms up to any arbitrary order in perturbation theory. However in this work we limit our computations to all the terms proportional to  $O(\delta_m^6)$ . For consistency all the higher order terms (e.g  $O(\delta_m^8)$ ) will be ignored. Of course the more higher order terms are included in the analysis, the better is the approximation used for the analytical covariance matrix.

### 3.B.1 Covariance term: $C^{\text{PP}}$

As stated in the main text, the covariance matrix element can be computed as

$$\begin{aligned} C^{\text{Ps}} &\equiv \text{Cov} \left[ P_g^{\text{s}}(\bar{\mathbf{k}}_1), P_g^{\text{s}}(\bar{\mathbf{k}}_3) \right] = \left\langle \left( \hat{P}_{g,1}^{\text{s}} - \langle \hat{P}_{g,1}^{\text{s}} \rangle \right) \right\rangle \left\langle \left( \hat{P}_{g,3}^{\text{s}} - \langle \hat{P}_{g,3}^{\text{s}} \rangle \right) \right\rangle \\ &= \langle \hat{P}_{g,1}^{\text{s}} \hat{P}_{g,3}^{\text{s}} \rangle - \hat{P}_{g,1}^{\text{s}} \hat{P}_{g,3}^{\text{s}}. \end{aligned} \quad (3.53)$$

The computation of the covariance matrix involves a four points correlator of  $\delta_g^{\text{s}}$  that can be expanded into its connected parts

$$\begin{aligned} \langle \delta_g^{\text{s}}(\mathbf{k}_1) \delta_g^{\text{s}}(\mathbf{k}_2) \delta_g^{\text{s}}(\mathbf{k}_3) \delta_g^{\text{s}}(\mathbf{k}_4) \rangle &= \langle \delta_g^{\text{s}}(\mathbf{k}_1) \delta_g^{\text{s}}(\mathbf{k}_2) \rangle_c \langle \delta_g^{\text{s}}(\mathbf{k}_3) \delta_g^{\text{s}}(\mathbf{k}_4) \rangle_c + 2 \text{ perms.} \\ &\quad + \langle \delta_g^{\text{s}}(\mathbf{k}_1) \delta_g^{\text{s}}(\mathbf{k}_2) \delta_g^{\text{s}}(\mathbf{k}_3) \delta_g^{\text{s}}(\mathbf{k}_4) \rangle_c. \end{aligned} \quad (3.54)$$

In this work, the following short hand for the integrals will be also adopted

$$\int_{V_{\bar{k}_i}^{\text{c}}} \frac{d\phi_i dk_i^{\perp} dk_i^{\parallel} k_i^{\perp}}{2\pi \bar{k}_i^{\perp} \Delta k^2} \equiv \int_i. \quad (3.55)$$

#### 3.B.1.1 $C_{\text{P}}^{\text{PP}}$ term:

Starting from the first term  $C_{\text{P}}^{\text{PP}}$  in the particular case in which the pairs are made of identical wave vectors (in this case, from the power spectrum estimator definition:  $\mathbf{k}_1 = \mathbf{k}_2$  and  $\mathbf{k}_3 = \mathbf{k}_4$ ) we have that

$$\begin{aligned} C_{\text{P}}^{\text{PP}}(\bar{\mathbf{k}}_1; \bar{\mathbf{k}}_3) &= \frac{1}{V_s^2 V_1^{\text{c}} V_3^{\text{c}}} \prod_{i=0}^4 \int dV_i^{\text{c}} \delta_{\text{D}}(\mathbf{k}_1 + \mathbf{k}_2) \delta_{\text{D}}(\mathbf{k}_3 + \mathbf{k}_4) \\ &\quad \times \langle \delta_g^{\text{s}}(\mathbf{k}_1) \delta_g^{\text{s}}(\mathbf{k}_2) \rangle \langle \delta_g^{\text{s}}(\mathbf{k}_3) \delta_g^{\text{s}}(\mathbf{k}_4) \rangle, \end{aligned} \quad (3.56)$$

it is straightforward to see that this terms cancels the second term of Equation 3.53. Considering now the other two possibilities for  $C_{\text{P}}^{\text{PP}}$

$$\begin{aligned}
 C_P^{\text{PP}}(\bar{\mathbf{k}}_1; \bar{\mathbf{k}}_3) &= \\
 &= \frac{1}{V_s^2 V_1^c V_3^c} \prod_{i=0}^4 \int dV_i^c \delta_D(\mathbf{k}_1 + \mathbf{k}_2) \delta_D(\mathbf{k}_3 + \mathbf{k}_4) \langle \delta_g^s(\mathbf{k}_1) \delta_g^s(\mathbf{k}_3) \rangle \langle \delta_g^s(\mathbf{k}_2) \delta_g^s(\mathbf{k}_4) \rangle + 1\text{p.} \\
 &= \frac{(2\pi)^6}{V_s^2 V_1^c V_3^c} \prod_{i=0}^4 \int dV_i^c \delta_D(\mathbf{k}_1 + \mathbf{k}_2) \delta_D(\mathbf{k}_3 + \mathbf{k}_4) \delta_D(\mathbf{k}_1 + \mathbf{k}_3) \\
 &\times \delta_D(\mathbf{k}_2 + \mathbf{k}_4) P_g^s(\mathbf{k}_1) P_g^s(\mathbf{k}_2) + 1\text{p.} \\
 &= \frac{(2\pi)^6}{V_s^2 V_1^c V_2^c} (\delta_{13}^K \delta_{24}^K + \delta_{14}^K \delta_{23}^K) \int dV_1^c \int dV_2^c \delta_D(\mathbf{k}_1 + \mathbf{k}_2)^2 P_g^s(\mathbf{k}_1) P_g^s(\mathbf{k}_2) \\
 &= \frac{(2\pi)^3}{V_s V_1^c V_2^c} D_{1234}^{\text{P}} \int dV_1^c \int dV_2^c \delta_D(\mathbf{k}_1 + \mathbf{k}_2) P_g^s(\mathbf{k}_1) P_g^s(\mathbf{k}_2) \\
 &= \frac{(2\pi)^3}{V_s V_1^c} D_{1234}^{\text{P}} \int_{\bar{k}_1^\perp - \frac{\Delta k}{2}}^{\bar{k}_1^\perp + \frac{\Delta k}{2}} \int_{\bar{k}_1^\parallel - \frac{\Delta k}{2}}^{\bar{k}_1^\parallel + \frac{\Delta k}{2}} \frac{dk_1^\perp dk_1^\parallel k_1^\perp}{\bar{k}_1^\perp \Delta k^2} P_g^s(\mathbf{k}_1)^2 \\
 &\approx \frac{(2\pi)^2}{V_s \bar{k}_1^\perp \Delta k^2} D_{1234}^{\text{P}} P_g^s(\bar{\mathbf{k}}_1)^2 = \frac{2 \times (2\pi)^2}{V_s \bar{k}_1^\perp \Delta k^2} \delta_{13}^K P_g^s(\bar{\mathbf{k}}_1)^2, \tag{3.57}
 \end{aligned}$$

where  $D_{1234}^{\text{P}} = \delta_{13}^K \delta_{24}^K + \delta_{14}^K \delta_{23}^K$  has been defined in order to take in account the additional permutation from the initial Dirac's deltas conditions, which now reduces to  $D_{1234}^{\text{P}} = 2 \times \delta_{13}^K$ . If, when obtaining an expression for the trispectrum, one considers also one-loop corrections, in addition to this Gaussian term on the diagonal of the power spectrum covariance matrix, one should add the unconnected terms encountered in Appendix 3.A,  $T_g^{s(2a)}$  and  $T_g^{s(3a)}$ . This has been recently well described in Mohammed et al. (2017). As stated in Sec. 3.3.2 the error made with this tree level approximation for the diagonal term 3.57 is around  $\sim 1\%$  for  $k_{\text{max}} = 0.2 \text{ Mpc}^{-1}h$  and up to  $\sim 9\%$  for  $k_{\text{max}} = 0.3 \text{ Mpc}^{-1}h$  (Taruya et al., 2008).

### 3.B.1.2 $C_T^{\text{PP}}$ term:

Proceeding in the same way it is possible to compute as well the other term of the cross-correlation matrix, the one containing the trispectrum contribution.

$$\begin{aligned}
 C_T^{\text{PP}}(\bar{\mathbf{k}}_1; \bar{\mathbf{k}}_3) &= \frac{1}{V_s^2 V_1^c V_3^c} \prod_{i=0}^4 \int dV_i^c \delta_D(\mathbf{k}_1 + \mathbf{k}_2) \delta_D(\mathbf{k}_3 + \mathbf{k}_4) \\
 &\times \langle \delta_g^s(\mathbf{k}_1) \delta_g^s(\mathbf{k}_3) \delta_g^s(\mathbf{k}_2) \delta_g^s(\mathbf{k}_4) \rangle \\
 &= \frac{(2\pi)^3}{V_s^2 V_1^c V_3^c} \prod_{i=0}^4 \int dV_i^c \delta_D(\mathbf{k}_1 + \mathbf{k}_2) \delta_D(\mathbf{k}_3 + \mathbf{k}_4) \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4) \\
 &\times T_g^s(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) \\
 &= \frac{(2\pi)^3}{V_s^2 V_1^c V_3^c} \delta_{34}^K \int dV_1^c dV_2^c dV_3^c \delta_D(\mathbf{k}_1 + \mathbf{k}_2)^2 T_g^s(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, -\mathbf{k}_3) \\
 &= \frac{1}{V_s V_1^c V_3^c} \delta_{34}^K \int dV_1^c dV_2^c dV_3^c \delta_D(\mathbf{k}_1 + \mathbf{k}_2) T_g^s(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, -\mathbf{k}_3) \\
 &= \frac{1}{V_s} \delta_{12}^K \delta_{34}^K \int \frac{dV_1^c}{V_1^c} \int \frac{dV_3^c}{V_3^c} T_g^s(\mathbf{k}_1, -\mathbf{k}_1, \mathbf{k}_3, -\mathbf{k}_3) \\
 &\approx \frac{1}{2\pi V_s} \int d\phi_{13} T_g^s(\bar{k}_1^\perp, \bar{k}_1^\parallel, \bar{k}_3^\perp, \bar{k}_3^\parallel, \phi_{13}). \tag{3.58}
 \end{aligned}$$

Following what was done in [Pielorz et al. \(2010\)](#), i.e, substituting in here the previously derived expression for the trispectrum, this expression can be "simplified" at tree level to

$$\begin{aligned}
 C_T^{\text{PP}}(\bar{\mathbf{k}}_1; \bar{\mathbf{k}}_3) &\approx \frac{1}{2\pi V_s} \int d\phi_{13} T_g^s(\bar{k}_1^\perp, \bar{k}_1^\parallel, \bar{k}_3^\perp, \bar{k}_3^\parallel, \phi_{13}) \\
 &= \frac{1}{2\pi V_s} \int d\phi_{13} \left\{ 4 P_g^s(\mathbf{k}_1)^2 \left[ F_s^{(2)}[\mathbf{k}_1, -\mathbf{k}_+]^2 P^m(\mathbf{k}_+) + F_s^{(2)}[\mathbf{k}_1, \mathbf{k}_-]^2 P^m(\mathbf{k}_-) \right] \right. \\
 &+ 4 P_g^s(\mathbf{k}_3)^2 \left[ F_s^{(2)}[\mathbf{k}_3, -\mathbf{k}_+]^2 P^m(\mathbf{k}_+) + F_s^{(2)}[\mathbf{k}_3, -\mathbf{k}_-]^2 P^m(\mathbf{k}_-) \right] \\
 &+ 8 P_g^s(\mathbf{k}_1) P_g^s(\mathbf{k}_3) \\
 &\times \left[ F_s^{(2)}[\mathbf{k}_1, -\mathbf{k}_+] F_s^{(2)}[\mathbf{k}_3, -\mathbf{k}_+] P^m(\mathbf{k}_+) + F_s^{(2)}[\mathbf{k}_1, \mathbf{k}_-] F_s^{(2)}[\mathbf{k}_3, -\mathbf{k}_-] P^m(\mathbf{k}_-) \right] \\
 &\left. + 12 P_g^s(\mathbf{k}_1) P_g^s(\mathbf{k}_3) \left[ P_g^s(\mathbf{k}_1) F_s^{(3)}[\mathbf{k}_1, -\mathbf{k}_1, \mathbf{k}_3] + P_g^s(\mathbf{k}_3) F_s^{(3)}[\mathbf{k}_1, \mathbf{k}_3, -\mathbf{k}_3] \right] \right\}, \tag{3.59}
 \end{aligned}$$

where  $\mathbf{k}_+ = \mathbf{k}_1 + \mathbf{k}_3$  and  $\mathbf{k}_- = \mathbf{k}_3 - \mathbf{k}_1$  and  $P^m$  is the linear matter power spectrum and  $\phi_{13}$  is the difference between the two azimuthal angles of the two wave-vectors. Therefore this is the only term that requires an integration since this angular dependence defines the relation between the two wave-vectors involved.

### 3.B.2 Covariance term: $C^{\text{BB}}$

By definition, the covariance matrix element can be computed as

$$\begin{aligned}
 \mathbf{C}^{\text{B}_g^s} &\equiv \mathbf{Cov} \left[ \mathbf{B}_g^s \left( \bar{k}_1^\perp, \bar{k}_2^\perp, \bar{k}_3^\perp, \bar{k}_1^\parallel, \bar{k}_2^\parallel \right), \mathbf{B}_g^s \left( \bar{k}_4^\perp, \bar{k}_5^\perp, \bar{k}_6^\perp, \bar{k}_4^\parallel, \bar{k}_5^\parallel \right) \right] \\
 &= \left\langle \left( \hat{\mathbf{B}}_{g,123}^s - \langle \hat{\mathbf{B}}_{g,123}^s \rangle \right) \right\rangle \left\langle \left( \hat{\mathbf{B}}_{g,456}^s - \langle \hat{\mathbf{B}}_{g,456}^s \rangle \right) \right\rangle \\
 &= \langle \hat{\mathbf{B}}_{g,123}^s \hat{\mathbf{B}}_{g,456}^s \rangle - \hat{\mathbf{B}}_{g,123}^s \hat{\mathbf{B}}_{g,456}^s.
 \end{aligned} \tag{3.60}$$

The six points correlation function can be decomposed into all the possible combinations of connected parts

$$\begin{aligned}
 &\langle \delta_g^s(\mathbf{k}_1) \delta_g^s(\mathbf{k}_2) \delta_g^s(\mathbf{k}_3) \delta_g^s(\mathbf{k}_4) \delta_g^s(\mathbf{k}_5) \delta_g^s(\mathbf{k}_6) \rangle = \\
 &= \langle \delta_g^s(\mathbf{k}_1) \delta_g^s(\mathbf{k}_2) \rangle_c \langle \delta_g^s(\mathbf{k}_3) \delta_g^s(\mathbf{k}_4) \rangle_c \langle \delta_g^s(\mathbf{k}_5) \delta_g^s(\mathbf{k}_6) \rangle_c + 14 \text{ perms.} \\
 &+ \langle \delta_g^s(\mathbf{k}_1) \delta_g^s(\mathbf{k}_2) \delta_g^s(\mathbf{k}_3) \rangle_c \langle \delta_g^s(\mathbf{k}_4) \delta_g^s(\mathbf{k}_5) \delta_g^s(\mathbf{k}_6) \rangle_c + 9 \text{ perms.} \\
 &+ \langle \delta_g^s(\mathbf{k}_1) \delta_g^s(\mathbf{k}_2) \delta_g^s(\mathbf{k}_3) \delta_g^s(\mathbf{k}_4) \rangle_c \langle \delta_g^s(\mathbf{k}_5) \delta_g^s(\mathbf{k}_6) \rangle_c + 14 \text{ perms.} \\
 &+ \langle \delta_g^s(\mathbf{k}_1) \delta_g^s(\mathbf{k}_2) \delta_g^s(\mathbf{k}_3) \delta_g^s(\mathbf{k}_4) \delta_g^s(\mathbf{k}_5) \delta_g^s(\mathbf{k}_6) \rangle_c.
 \end{aligned} \tag{3.61}$$

Below we present the full derivation of the of the bispectrum Covariance matrix to leading order in Gaussian terms

$$\begin{aligned}
 \mathbf{C}^{\text{BB}} \left( \bar{\mathbf{k}}_1, \bar{\mathbf{k}}_2, \bar{\mathbf{k}}_3; \bar{\mathbf{k}}_4, \bar{\mathbf{k}}_5, \bar{\mathbf{k}}_6 \right) &= \\
 &= \frac{(2\pi)^4 \Delta k_3^\parallel \Delta k_6^\parallel}{V_s^2} \Lambda_{123}^{-1} \Lambda_{456}^{-1} \prod_{i=1}^6 \int_i \frac{dV_i^c}{V_i^c} \mathbf{F}_{s,i}^{(1)} \delta_{\text{D}}(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) \delta_{\text{D}}(\mathbf{k}_4 + \mathbf{k}_5 + \mathbf{k}_6) \\
 &\times \langle \delta_{\text{m}}(\mathbf{k}_1) \delta_{\text{m}}(\mathbf{k}_4) \rangle \langle \delta_{\text{m}}(\mathbf{k}_2) \delta_{\text{m}}(\mathbf{k}_5) \rangle \langle \delta_{\text{m}}(\mathbf{k}_3) \delta_{\text{m}}(\mathbf{k}_6) \rangle + 5 \text{ p.} \\
 &= \frac{(2\pi)^{13} \Delta k_3^{\parallel 2}}{V_s^2 V_1^c V_2^c V_3^c} \Lambda_{123}^{-2} \delta_{14}^{\text{K}} \delta_{25}^{\text{K}} \delta_{36}^{\text{K}} \prod_{i=1}^3 \int_i \mathbf{F}_{s,i}^{(1)2} \mathbf{P}^{\text{m}}(k_i) \delta_{\text{D}}(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3)^2 + 5 \text{ p.} \\
 &= \frac{(2\pi)^{10} \Delta k_3^{\parallel 2}}{V_s V_1^c V_2^c V_3^c} \Lambda_{123}^{-2} \mathbf{D}_{123456} \prod_{i=1}^3 \int_i \mathbf{P}_g^s(\mathbf{k}_i) \delta_{\text{D}}(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) \\
 &= \frac{(2\pi)^{11} \Delta k_3^{\parallel 2}}{V_s V_1^c V_2^c V_3^c} \Lambda_{123}^{-1} \mathbf{D}_{123456} \prod_{i=1}^3 \int_{\bar{k}_i^\perp - \frac{\Delta k}{2}}^{\bar{k}_i^\perp + \frac{\Delta k}{2}} \int_{\bar{k}_i^\parallel - \frac{\Delta k}{2}}^{\bar{k}_i^\parallel + \frac{\Delta k}{2}} \frac{dk_i^\perp dk_i^\parallel k_i^\perp}{V_i^c} \mathbf{P}_g^s(\mathbf{k}_i) \delta_{\text{D}}(k_1^\parallel + k_2^\parallel + k_3^\parallel) \\
 &= \frac{(2\pi)^5 \Delta k_3^\parallel}{V_s \bar{k}_1^\perp \bar{k}_2^\perp \bar{k}_3^\perp \Delta k^6} \Lambda_{123}^{-1} \mathbf{D}_{123456} \prod_{i=1}^3 \int_{\bar{k}_i^\perp - \frac{\Delta k}{2}}^{\bar{k}_i^\perp + \frac{\Delta k}{2}} \frac{dk_i^\perp k_i^\perp}{\bar{k}_i^\perp \Delta k_i^\perp} \prod_{j=1}^2 \int_{\bar{k}_j^\parallel - \frac{\Delta k}{2}}^{\bar{k}_j^\parallel + \frac{\Delta k}{2}} \frac{dk_j^\parallel}{\Delta k_j^\parallel} \mathbf{P}_g^s(\mathbf{k}_i),
 \end{aligned} \tag{3.62}$$

where  $\mathbf{D}_{123456}$  has been used as a shorthand notation for all the possible permutations allowed by the fact that only pairs formed by wave-vectors from different triplets survive, which are in total 6. For all the other steps we have used relations described previously. Finally, defining a bin average by using the thin shell approximation, it is possible to write down the final result

$$C^{\text{BB}} = \frac{(2\pi)^5 \Delta k_3^{\parallel}}{V \bar{k}_1^{\perp} \bar{k}_2^{\perp} \bar{k}_3^{\perp} \Delta k^6} \Lambda_{123}^{-1} D_{123456} \prod_{i=1} F_s^{(1)2} [\bar{k}_i^{\perp}, \bar{k}_i^{\parallel}] P^{\text{m}}(\bar{k}_i^{\perp}, \bar{k}_i^{\parallel}). \quad (3.63)$$

### 3.B.3 Cross - variance term: $C^{\text{BP}}$

The cross-correlation part of the covariance matrix of the joint data-vector  $[P_g^{\text{s}}, B_g^{\text{s}}]$  can be computed as

$$C^{\text{BP}} \equiv \text{Cov} [P_g^{\text{s}}(\bar{\mathbf{k}}_1), B_g^{\text{s}}(\bar{\mathbf{k}}_3, \bar{\mathbf{k}}_4, \bar{\mathbf{k}}_5)] = \langle \hat{P}_{g,1}^{\text{s}} \hat{B}_{g,345}^{\text{s}} \rangle - \hat{P}_{g,1}^{\text{s}} \hat{B}_{g,345}^{\text{s}}. \quad (3.64)$$

The computation of the covariance matrix involves a five points correlator of  $\delta_g^{\text{s}}$  that can be expanded in its connected parts

$$\begin{aligned} \langle \delta_g^{\text{s}}(\mathbf{k}_1) \delta_g^{\text{s}}(\mathbf{k}_2) \delta_g^{\text{s}}(\mathbf{k}_3) \delta_g^{\text{s}}(\mathbf{k}_4) \delta_g^{\text{s}}(\mathbf{k}_5) \rangle &= \langle \delta_g^{\text{s}}(\mathbf{k}_1) \delta_g^{\text{s}}(\mathbf{k}_2) \rangle_c \langle \delta_g^{\text{s}}(\mathbf{k}_3) \delta_g^{\text{s}}(\mathbf{k}_4) \delta_g^{\text{s}}(\mathbf{k}_5) \rangle_c \\ &\quad + \langle \delta_g^{\text{s}}(\mathbf{k}_1) \delta_g^{\text{s}}(\mathbf{k}_3) \rangle_c \langle \delta_g^{\text{s}}(\mathbf{k}_2) \delta_g^{\text{s}}(\mathbf{k}_4) \delta_g^{\text{s}}(\mathbf{k}_5) \rangle_c \quad + \quad 5 \text{ p.} \\ &\quad + \langle \delta_g^{\text{s}}(\mathbf{k}_3) \delta_g^{\text{s}}(\mathbf{k}_4) \rangle_c \langle \delta_g^{\text{s}}(\mathbf{k}_1) \delta_g^{\text{s}}(\mathbf{k}_2) \delta_g^{\text{s}}(\mathbf{k}_5) \rangle_c \quad + \quad 2 \text{ p.} \\ &\quad + \langle \delta_g^{\text{s}}(\mathbf{k}_3) \delta_g^{\text{s}}(\mathbf{k}_4) \delta_g^{\text{s}}(\mathbf{k}_1) \delta_g^{\text{s}}(\mathbf{k}_2) \delta_g^{\text{s}}(\mathbf{k}_5) \rangle_c \\ &\quad + \langle \delta_g^{\text{s}}(\mathbf{k}_1) \delta_g^{\text{s}}(\mathbf{k}_2) \delta_g^{\text{s}}(\mathbf{k}_3) \delta_g^{\text{s}}(\mathbf{k}_4) \delta_g^{\text{s}}(\mathbf{k}_5) \rangle_{\text{unc.}} \\ &= C_{m0}^{\text{BP}} + C_{m1}^{\text{BP}} + C_{m2}^{\text{BP}} + C_{\text{Te}}^{\text{BP}} + C_{U_{5p}}^{\text{BP}}. \end{aligned} \quad (3.65)$$

In the case of the five points correlation function it has to be included also the unconnected part in the expansion. This is because while for an even number  $2n$  the leading term of unconnected part of the  $2n$ -points correlation function is proportional to the product of  $n$  power spectra, for an odd number  $m$  this is no longer the case. As shown in last subsection of Appendix 3.A, the unconnected part of the 5-points correlation function has a leading term proportional to the product of three power spectra. Therefore in this case the unconnected part needs to be considered since it has the same order of the other leading terms of the above expansion. From the last equation, it is possible to immediately that  $C_{m0}^{\text{BP}}$  cancels out with the last term of Equation 3.64 and that  $C_{m2}^{\text{BP}} = 0$  because the Dirac's deltas combination leads to terms having  $\delta_{\text{D}}(\mathbf{k}_5) = 0$ .

#### 3.B.3.1 $C_{m1}^{\text{BP}}$ term:

Starting from the first term written previously, below we report the full derivation

$$\begin{aligned}
 C_{\text{m1}}^{\text{BP}}(\bar{\mathbf{k}}_1; \bar{\mathbf{k}}_3, \bar{\mathbf{k}}_4, \bar{\mathbf{k}}_5) &= \\
 &= \frac{(2\pi)^2 \Delta k_5^{\parallel}}{V_s^2} \Lambda_{345}^{-1} \int dV_1^c \prod_{i=2}^5 \int \frac{dV_i^c}{V_i^c} \delta_{\text{D}}(\mathbf{k}_1 + \mathbf{k}_2) \delta_{\text{D}}(\mathbf{k}_3 + \mathbf{k}_4 + \mathbf{k}_5) \\
 &\times \langle \delta_{\text{g}}^{\text{s}}(\mathbf{k}_1) \delta_{\text{g}}^{\text{s}}(\mathbf{k}_3) \rangle_c \langle \delta_{\text{g}}^{\text{s}}(\mathbf{k}_2) \delta_{\text{g}}^{\text{s}}(\mathbf{k}_4) \delta_{\text{g}}^{\text{s}}(\mathbf{k}_5) \rangle_c + 5 \text{ p.} \\
 &= \frac{(2\pi)^8 \Delta k_5^{\parallel}}{V_s^2} \Lambda_{345}^{-1} \int dV_1^c \prod_{i=2}^5 \int \frac{dV_i^c}{V_i^c} \delta_{\text{D}}(\mathbf{k}_1 + \mathbf{k}_2) \delta_{\text{D}}(\mathbf{k}_3 + \mathbf{k}_4 + \mathbf{k}_5) \delta_{\text{D}}(\mathbf{k}_1 + \mathbf{k}_3) \\
 &\times \delta_{\text{D}}(\mathbf{k}_2 + \mathbf{k}_4 + \mathbf{k}_5) \text{P}_{\text{g}}^{\text{s}}(\mathbf{k}_3) \text{B}_{\text{g}}^{\text{s}}(\mathbf{k}_2, \mathbf{k}_4, \mathbf{k}_5) + 5 \text{ p.} \\
 &= \frac{(2\pi)^5 \Delta k_5^{\parallel}}{V_s} \Lambda_{345}^{-1} \int dV_1^c \prod_{i=2}^5 \int \frac{dV_i^c}{V_i^c} \delta_{\text{D}}(\mathbf{k}_1 + \mathbf{k}_2) \delta_{\text{D}}(\mathbf{k}_1 + \mathbf{k}_3) \delta_{\text{D}}(\mathbf{k}_3 + \mathbf{k}_4 + \mathbf{k}_5) \\
 &\times \text{P}_{\text{g}}^{\text{s}}(\mathbf{k}_3) \text{B}_{\text{g}}^{\text{s}}(\mathbf{k}_2, \mathbf{k}_4, \mathbf{k}_5) + 5 \text{ p.} \\
 &= \frac{(2\pi)^5 \Delta k_5^{\parallel}}{V_s} \Lambda_{345}^{-1} \delta_{12}^{\text{K}} \prod_{i=2}^5 \int \frac{dV_i^c}{V_i^c} \delta_{\text{D}}(-\mathbf{k}_2 + \mathbf{k}_3) \delta_{\text{D}}(\mathbf{k}_3 + \mathbf{k}_4 + \mathbf{k}_5) \\
 &\times \text{P}_{\text{g}}^{\text{s}}(\mathbf{k}_3) \text{B}_{\text{g}}^{\text{s}}(\mathbf{k}_2, \mathbf{k}_4, \mathbf{k}_5) + 5 \text{ p.} \\
 &= \frac{(2\pi)^5 \Delta k_5^{\parallel}}{V_s V_2^c} \Lambda_{345}^{-1} \delta_{12}^{\text{K}} \delta_{-23}^{\text{K}} \prod_{i=3}^5 \int \frac{dV_i^c}{V_i^c} \delta_{\text{D}}(\mathbf{k}_3 + \mathbf{k}_4 + \mathbf{k}_5) \text{P}_{\text{g}}^{\text{s}}(\mathbf{k}_3) \text{B}_{\text{g}}^{\text{s}}(\mathbf{k}_3, \mathbf{k}_4, \mathbf{k}_5) + 5 \text{ p.} \\
 &= \frac{(2\pi)^5 \Delta k_5^{\parallel}}{V_s V_2^c} \Lambda_{345}^{-1} \delta_{12}^{\text{K}} \delta_{-23}^{\text{K}} \prod_{i=3}^5 \int \frac{dk_i^{\parallel} dk_i^{\perp} k_i^{\perp}}{2\pi k_i^{\perp} \Delta k^2} 2\pi \Lambda_{345} \delta_{\text{D}}(k_3^{\parallel} + k_4^{\parallel} + k_5^{\parallel}) \\
 &\times \text{P}_{\text{g}}^{\text{s}}(\mathbf{k}_3) \text{B}_{\text{g}}^{\text{s}}(\mathbf{k}_3, \mathbf{k}_4, \mathbf{k}_5) + 5 \text{ p.} \\
 &\approx \frac{(2\pi)^3}{V_s V_2^c} \delta_{-23}^{\text{K}} \text{P}_{\text{g}}^{\text{s}}(\bar{\mathbf{k}}_3) \text{B}_{\text{g}}^{\text{s}}(\bar{\mathbf{k}}_3, \bar{\mathbf{k}}_4, \bar{\mathbf{k}}_5) + 5 \text{ p.} \tag{3.66}
 \end{aligned}$$

Note that only three of the six permutations are different terms, since the other three are just obtained by switching  $\mathbf{k}_1$  with  $\mathbf{k}_2$  which does not change the final result. Therefore the final expression for this term can be written as

$$C_{\text{m1}}^{\text{BP}} = 2 \times \frac{(2\pi)^3}{V_s V_2^c} (\delta_{-13}^{\text{K}} + \delta_{-14}^{\text{K}} + \delta_{-15}^{\text{K}}) \text{P}_{\text{g}}^{\text{s}}(\bar{\mathbf{k}}_1) \text{B}_{\text{g}}^{\text{s}}(\bar{\mathbf{k}}_3, \bar{\mathbf{k}}_4, \bar{\mathbf{k}}_5). \tag{3.67}$$

Note that the argument of the power spectrum depend on the wave-vector selected by the Kronecker Delta.

### 3.B.3.2 $C_{\text{Te}}^{\text{BP}}$ term:

The other contribution at tree level to the connected five points correlator resulting in the cross covariance term between power spectrum and bispectrum is the one proportional to the tetraspectrum defined before, we get

$$\begin{aligned}
 C_{U_{5p}}^{\text{BP}}(\bar{\mathbf{k}}_2; \bar{\mathbf{k}}_3, \bar{\mathbf{k}}_4, \bar{\mathbf{k}}_5) &= \\
 &= \frac{(2\pi)^2 \Delta k_5^{\parallel}}{V_s^2} \Lambda_{345}^{-1} \int dV_1^c \prod_{i=2}^5 \int \frac{dV_i^c}{V_i^c} \delta_D(\mathbf{k}_1 + \mathbf{k}_3) \delta_D(\mathbf{k}_3 + \mathbf{k}_4 + \mathbf{k}_5) \\
 &\times \langle \delta_g^s(\mathbf{k}_1) \delta_g^s(\mathbf{k}_3) \delta_g^s(\mathbf{k}_2) \delta_g^s(\mathbf{k}_4) \delta_g^s(\mathbf{k}_5) \rangle_c \\
 &= \frac{(2\pi)^5 \Delta k_5^{\parallel}}{V_s^2} \Lambda_{345}^{-1} \int dV_1^c \prod_{i=2}^5 \int \frac{dV_i^c}{V_i^c} \delta_D(\mathbf{k}_1 + \mathbf{k}_2) \delta_D(\mathbf{k}_3 + \mathbf{k}_4 + \mathbf{k}_5) \\
 &\times \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4 + \mathbf{k}_5) T_{e,g}^s(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4, \mathbf{k}_5) \\
 &= \frac{(2\pi)^5 \Delta k_5^{\parallel}}{V_s^2} \Lambda_{345}^{-1} \delta_{12}^{\text{K}} \prod_{i=2}^5 \int \frac{dV_i^c}{V_i^c} \delta_D(\mathbf{k}_3 + \mathbf{k}_4 + \mathbf{k}_5)^2 T_{e,g}^s(\mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4, \mathbf{k}_5) \\
 &= \frac{(2\pi)^2 \Delta k_5^{\parallel}}{V_s} \Lambda_{345}^{-1} \prod_{i=2}^5 \int \frac{dV_i^c}{V_i^c} \delta_D(\mathbf{k}_3 + \mathbf{k}_4 + \mathbf{k}_5) T_{e,g}^s(\mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4, \mathbf{k}_5). \quad (3.68)
 \end{aligned}$$

Since the leading order of the tetraspectrum is higher than the one considered in this work, the above covariance term has not been included in the numerical computation. For what concerns the unconnected part of the five point correlation function described in the end of Appendix 3.A, it represents loop correction to the standard power spectrum times bispectrum cross covariance  $C_{m1}^{\text{BP}}$ . Coherently with what done in the same case for the trispectrum it has also not been included in the numerical computations.

### 3.C Compressed covariance matrix

We run the MCMC not on the original data-vector, e.g.  $B_g^s$  or  $[P_g^s + B_g^s]$  but on their compressed version obtained through MOPED weighting. Given a set of parameters  $\{\theta_1, \dots, \theta_m\}$ , a data-vector  $\mathbf{x} = \{x_1, \dots, x_n\}$  with mean  $\langle \mathbf{x} \rangle$  the set of weights with respect to each parameter  $\theta_i$  is given by

$$\mathbf{b}_i = \mathbf{Cov}^{-1} \langle \mathbf{x} \rangle_{,i}, \quad (3.69)$$

where  $\mathbf{Cov}(\mathbf{x})$  is the covariance matrix relative to the data-vector  $\mathbf{x}$ . Assuming the covariance matrix to be approximately independent of the cosmology, then for each considered parameter the original data-vector is compressed to a single scalar

$$y_{\theta_i} = \mathbf{b}_i \cdot \mathbf{x}. \quad (3.70)$$

Therefore stacking all the weighting vectors as rows of an  $m \times n$  matrix  $\mathbf{B}$  the new  $m$ -dimensional data-vector will be

$$\mathbf{y} = \mathbf{B}\mathbf{x}. \quad (3.71)$$

For what concerns the covariance matrix for the new data-vector, it will be related to the original one by

$$\begin{aligned} \mathbf{Cov}_{\mathbf{y},ij} = \mathbf{Cov}[y_i, y_j] &= \mathbf{Cov}\left[\sum_k^n b_i^k x_k, \sum_l^n b_j^l x_l\right] \\ &= \sum_k^n \sum_l^n b_i^k b_j^l \mathbf{Cov}[x_k, x_l] \\ &= \mathbf{b}_i^\top \cdot \mathbf{Cov}_{\mathbf{x}} \cdot \mathbf{b}_j, \end{aligned} \quad (3.72)$$

Therefore when running the MCMC using  $\mathbf{y}$  as data-vector the natural logarithm of likelihood will be proportional to

$$\log \mathcal{L} \propto -\frac{1}{2} (\mathbf{y} - \mathbf{y}_{\text{fid.}})^\top \mathbf{Cov}_{\mathbf{y}}^{-1} (\mathbf{y} - \mathbf{y}_{\text{fid.}}). \quad (3.73)$$

Where  $\mathbf{y}_{\text{fid.}}$  is the compressed data-vector obtained using the fiducial values of the cosmological parameters. In our case of interest this will be applied to  $\mathbf{B}_g^s$  and  $[\mathbf{P}_g^s, \mathbf{B}_g^s]$ . An additional option consists in leaving  $\mathbf{P}_g^s$  uncompressed, compressing only the bispectrum. In that case the cross-variance term would be given by

$$\mathbf{Cov}_{ij}^{Py_B} = \mathbf{Cov}[\mathbf{P}_g^s(\mathbf{k}_i), y_j] = \mathbf{Cov}[\mathbf{P}_g^s(\mathbf{k}_i), \mathbf{B}_g^s] \cdot \mathbf{b}_j. \quad (3.74)$$

Putting everything together the full covariance matrix is

$$\mathbf{Cov} = \begin{bmatrix} \mathbf{Cov}^{PP} & \mathbf{Cov}^{Py_B} \\ \mathbf{Cov}^{Py_B} & \mathbf{Cov}^{y_B y_B} \end{bmatrix}, \quad (3.75)$$

where  $\mathbf{Cov}^{y_B y_B}$  is the covariance matrix for the compressed data-vector derived in Equation 3.72 in the case of  $\mathbf{x}$  being the Bispectrum.

### 3.D Weights orthogonalisation

As stated in [Zablocki and Dodelson \(2016\)](#) each compressed data set  $y_i$  contains all the information regarding the parameter  $\theta_i$  but at the same time it will have some sensitivity to the other parameters. It is possible to remove this sensitivity by marginalising at linear level over these other parameters. This is done by orthogonalising the weighting vectors for all the individual parameters. If the model is parametrised by  $m$  parameters  $\theta_i$ , compressing with respect to all these

will return  $m$  different linear combinations  $y_i$  (or in other words  $m$  weighting vectors). Therefore it is possible to define another set of new scalars  $y'_i$  given by a linear combination of the original ones, for example:

$$y'_1 = c_1 y_1 + c_2 y_2 + \dots + c_m y_m, \quad (3.76)$$

with the constants  $c_i$  such that  $y'_1$  contains all the information regarding  $\theta_1$  while at the same time removing all the sensitivity with respect to the other parameters at linear level. This means that when taking the derivative with respect to the second parameter  $\theta_2$  we are looking for  $c$  coefficients such that:

$$\begin{aligned} \frac{dy'_1}{d\theta_2} &= c_1 \frac{dy_1}{d\theta_2} + c_2 \frac{dy_2}{d\theta_2} + \dots + c_m \frac{dy_m}{d\theta_2} \\ &= c_1 \sum_j b_1^j \frac{dx_j}{d\theta_2} + c_2 \sum_j b_2^j \frac{dx_j}{d\theta_2} + \dots + c_m \sum_j b_m^j \frac{dx_j}{d\theta_2} \\ &= c_1 F_{12} + c_2 F_{22} + \dots + c_m F_{m2} = 0, \end{aligned} \quad (3.77)$$

where for the Fisher information matrix elements it has been used Equation 3.21 together with the expression for the weights  $\mathbf{b}_i = \mathbf{Cov}^{-1}(\mathbf{x})_{,i}$ . Imposing this for the derivative with respect to all the  $m - 1$  parameters beside  $\theta_1$  it gives the following matrix problem to be solved in order to get the  $m - 1$  coefficients needed to compute  $y'_1$ :

$$\begin{bmatrix} F_{22} & F_{23} & \cdot & \cdot & \cdot & F_{2m} \\ F_{32} & F_{33} & \cdot & \cdot & \cdot & F_{3m} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ F_{m2} & F_{m3} & \cdot & \cdot & \cdot & F_{mm} \end{bmatrix} \begin{bmatrix} c_2 \\ c_3 \\ \cdot \\ \cdot \\ \cdot \\ c_m \end{bmatrix} = \begin{bmatrix} -F_{12} \\ -F_{13} \\ \cdot \\ \cdot \\ \cdot \\ -F_{1m} \end{bmatrix}. \quad (3.78)$$

Solving this equation returns  $m - 1$  unique constants for the  $c_i$  with  $i > 1$ .  $c_1$  can be set equal to one. Once the  $c$ 's are known,  $y'_1$  can be obtained using Equation 3.76, which corresponds to defining the new weighting vector for the parameter  $\theta_1$ :

$$\mathbf{b}'_1 = \mathbf{b}_1 + c_2 \mathbf{b}_2 + \dots + c_m \mathbf{b}_m, \quad (3.79)$$

where  $\mathbf{b}'_1$  is the orthogonalised weighting vector for  $\theta_1$ . The same procedure can be repeated for all the other parameters and relative weighting vectors. In general, for the mode with  $i = \alpha$ , the coefficients are determined by the general equation:

$$\sum_j F'_{\alpha,ij} c_j = -F_{\alpha i}, \quad (3.80)$$

where  $F'_\alpha$  is the Fisher matrix with row and column  $\alpha$  removed.

“May your beer be laid under an enchantment of surpassing excellence for seven years!”

- J.R.R. Tolkien, The Lord of the Rings

## 4 | Enhancing BOSS bispectrum cosmological constraints with maximal compression

### 4.1 Abstract

We apply two compression methods to the galaxy power spectrum monopole, quadrupole and bispectrum monopole measurements from the BOSS DR12 CMASS sample. Both methods reduce the dimension of the original data-vector to the number of cosmological parameters considered, using the MOPED algorithm with an analytic covariance model. In the first case, we infer the posterior through MCMC sampling from the likelihood of the compressed data-vector (MC-MP). The second, faster option, works by first Gaussianising and then orthogonalising the parameter space before the compression; in this option (G-PCA) we only need to run a low-resolution preliminary MCMC sample for the Gaussianization to compute our posterior. Both compression methods accurately reproduce the posterior distributions obtained by standard MCMC sampling on the CMASS dataset for a  $k$ -space range of  $0.03 - 0.12 h/\text{Mpc}$ . The compression enables us to increase the number of bispectrum measurements by a factor of  $\sim 23$  over the standard binning (from 116 to 2734 triangles used), which is otherwise limited by the number of mock catalogues available. This reduces the 68% credible intervals for the parameters  $(b_1, b_2, f, \sigma_8)$  by  $(-24.8\%, -52.8\%, -26.4\%, -21\%)$ , respectively. The best-fit values we obtain are  $(b_1 = 2.31 \pm 0.17, b_2 = 0.77 \pm 0.19, f(z_{\text{CMASS}}) = 0.67 \pm 0.06, \sigma_8(z_{\text{CMASS}}) = 0.51 \pm 0.03)$ . Using these methods for future redshift surveys like DESI, Euclid and PFS will drastically reduce the number of simulations needed to compute accurate covariance matrices and will facilitate tighter constraints on cosmological parameters.

## 4.2 Introduction

Large datasets have recently become available from current cosmological surveys (*Planck*, <sup>1</sup> Ade et al., 2014 ; Sloan Digital Sky Survey <sup>2</sup>, Eisenstein et al., 2011; KiDS de Jong et al., 2013a; DES, Dark Energy Survey Collaboration et al., 2016 <sup>3</sup>) and even larger ones will be provided in future by DESI<sup>4</sup>, Levi et al. (2013); Euclid <sup>5</sup>, Laureijs et al. (2011); PFS <sup>6</sup>, Takada et al. (2014) and the LSST<sup>7</sup>, LSST Science Collaboration et al. (2009). In order to exploit their full potential, it is desirable to go beyond standard two-points statistics (2pt).

Three-points statistics (3pt) are a complementary probe that is possible to investigate both in configuration and Fourier space and have been used extensively in galaxy clustering analyses (Groth and Peebles, 1977, Fry, 1984, Fry and Gaztanaga, 1993, Frieman and Gaztanaga, 1994, Matarrese et al., 1997, Verde et al., 1998, Heavens et al., 1998, Scoccimarro et al., 1998, Scoccimarro, 2000, Sefusatti et al., 2006). Deviations from General Relativity (Borisov and Jain, 2009; Bernardeau and Brax, 2011; Gil-Marín et al., 2011) and primordial non-Gaussianities (Fry and Scherrer, 1994; Gangui et al., 1994; Verde et al., 2000; Liguori et al., 2010; Tellarini et al., 2016) have been investigated using 3pt statistics. Their potential in lifting degeneracies present at 2pt level has been shown by the most recent measurement on the BOSS dataset, for the bispectrum by Gil-Marín (2017) and for the 3pt correlation function by Slepian et al. (2017a). Baryonic acoustic oscillations (BAO) have also been measured using the 3pt correlation function by Slepian et al. (2017b) and detected using the bispectrum by Pearson and Samushia (2017).

Recently, 3pt statistics have been studied in the case of 21cm emission lines by Hoffmann et al. (2018). For what concerns weak lensing, its effect on 3pt galaxy clustering have been studied by Schmidt et al. (2008). Moreover the weak lensing bispectrum has been object of several studies in recent years (Takada and Jain, 2004; Joachimi et al., 2009; Kayo, I. and Takada, M. and Jain, B., 2013; Kayo and Takada, 2013). The skewness of mass aperture statistic was considered by Jarvis et al. (2004) while the 3pt correlation function of cosmic shear was analysed by Schneider et al. (2005); Kilbinger and Schneider (2005). higher-order statistics like the bispectrum via gravitational lensing have been investigated also by Simon (2013); Fu et al. (2014); Simon et al. (2015); Pyne et al. (2017).

<sup>1</sup><http://sci.esa.int/planck/>

<sup>2</sup><http://www.sdss3.org/surveys/boss.php>

<sup>3</sup><https://www.darkenergysurvey.org>

<sup>4</sup><http://desi.lbl.gov>

<sup>5</sup><http://sci.esa.int/euclid/>

<sup>6</sup><http://pfs.ipmu.jp>

<sup>7</sup><https://www.lsst.org/>

Besides being computationally more expensive than 2pt statistics, 3pt statistics present the drawback to be described by very large data-vectors, which in turn require a high number of simulations to accurately estimate their covariance matrix (Hartlap et al., 2007). In Gualdi et al. (2018b), Chapter 3 from now on, we presented two methods to compress the redshift-space galaxy bispectrum, namely MC-MP (Markov chain Monte Carlo sampling + MOPED compression) and PCA + MP (principal component analysis transformation + MOPED compression). MOPED is the method presented in Heavens et al. (2000), which achieves maximal compression of the original data-vector by extending to the multiple parameters case the algorithm introduced in Tegmark et al. (1997). MC-MP consists in sampling via MCMC the compressed data-vector’s likelihood. PCA + MP reconstructs the multidimensional physical posterior distribution from the 1D posterior of orthogonalised parameters obtained by diagonalising the Fisher information matrix. Improvements of the MOPED algorithm were introduced recently by Heavens et al. (2017); Alsing and Wandelt (2018); Alsing et al. (2018) also with the target of data compression.

In this work we apply our compression methods to both the power spectrum monopole/quadrupole and to the bispectrum monopole measurements from the CMASS sample of BOSS DR12. While the MC-MP is more flexible than the PCA + MP method since doesn’t require the multidimensional Gaussian posterior assumption, the PCA + MP is much faster in terms of computational time and requires far fewer computational resources (it can be run on standard laptop). We compare both methods and test their convergence in terms of deriving equivalent posterior distributions.

In order to make the PCA + MP method applicable also to parameter spaces with strong degeneracies, for which the posterior Gaussianity approximation is no longer valid, we introduce a pre-Gaussianisation step based on the algorithm developed by Schuhmann et al. (2016).

We measure the bispectrum monopole using the same code used for the BOSS DR12 analysis done by Gil-Marín (2017). We vary the size of the triangle vectors by changing the bin size  $\Delta k$  for  $k$ , which returns different number of triangular shapes given the minimum and maximum scales. For the same number of triangles the compression returns posterior distributions slightly larger than the MCMC counterparts. However, when compressing a much larger number of triangles (which cannot be done for the MCMC on the full data-vector because of the limited number of mocks available constraint), the posterior distribution becomes more Gaussian and narrow. It eventually returns tighter constraints than the ones obtained by the standard analysis.

In Sec. 4.3 we describe the data set and the galaxy mocks used to estimate the covariance matrix together with the settings of our analysis. In Sec. 4.4 we present the analytical model used for the data-vector considered and the analytical expression of the covariance matrix used to derive the weights for the compression. In Sec. 4.5 we recap the compression methods applied including the Gaussianisation extension for the original PCA + MP method. We report the performance of the compression methods compared to the MCMC sampling for the cases in which it is possible to run it on the full data-vector in Sec. 4.6. We describe the gain in parameter constraints as a function of the number of triangle configurations used in the bispectrum monopole data-vector component in Sec. 4.7. We test the flexibility and accuracy of the compression methods presented in Sec. 4.8. Finally we conclude summarising our results in Sec. 4.9. In Appendix 4.A we report the full derivation of all the analytic expressions used in the analysis. In Appendix 4.B additional validation tests are presented.

## 4.3 Data, mocks and analysis

### 4.3.1 DR12 BOSS data and mocks catalogues

In this paper we use the CMASS galaxy sample ( $0.43 \leq z \leq 0.70$ ) of the Baryon Oscillation Spectroscopic Survey (BOSS Dawson et al., 2013) which is part of the Sloan Digital Sky Survey III (Eisenstein et al., 2011). In the final data release DR12 the CMASS sample contains the spectroscopic redshift of 777202 galaxies (see Gil-Marín 2017 and Alam, 2017 for more details).

In order to accurately numerically estimate the covariance matrix it is necessary to employ a large suite of mock galaxy catalogues. These are different realizations of the same region of the Universe based on methods such as second-order Lagrangian perturbation theory (2LPT Scoccimarro and Sheth 2002; Manera et al. 2013) or augmented Lagrangian perturbation theory (ALPT) as described in Kitaura and Heß (2013). By measuring the data-vector of interest on each one of these catalogues we can numerically estimate the covariance matrix which will be used in the likelihood evaluation. In this work we use subsets of the 2048 realisations of the MultiDark Patchy BOSS DR12 mocks by Kitaura et al. (2016). This set of mocks has been run using the underlying cosmology:  $\Omega_\Lambda = 0.693$ ,  $\Omega_m(z=0) = 0.307$ ,  $\Omega_b(z=0) = 0.048$ ,  $\sigma_8(z=0) = 0.829$ ,  $n_s = 0.96$ ,  $h_0 = 0.678$ .

### 4.3.2 Analysis settings

For the power spectrum monopole and quadrupole the bin size was fixed to  $\Delta k = 0.01h/\text{Mpc}$ . We measured the bispectrum monopole from both data and mocks

using different multiples of the fundamental frequency defined as  $k_f^3 = \frac{(2\pi)^3}{V_s}$  where  $V_s$  is the survey volume which in this case was the cubic box volume  $V_s = L_b^3 = (3500 \text{ Mpc}/h)^3$  used to analyse the galaxy mocks. In particular, the considered bin sizes for the bispectrum are  $\Delta k = (6, 5, 4, 2) \times k_f$  respectively, corresponding to 116, 195, 404 and 2734 triangles used between  $0.03 < k_i [h/\text{Mpc}] < 0.12$ . The largest bin size  $\Delta k = 6 \times k_f$  corresponds to the one used in the BOSS collaboration analysis done by [Gil-Marín \(2017\)](#). For the  $k$ -range considered in the BOSS analysis the  $\Delta k_6$  ( $\Delta k = 6 \times k_f$ ) binning case corresponded to 825 fundamental triangle configurations while  $\Delta k_2$  would have corresponded to more than  $\sim 7000$  triangles.

In all the parameter estimation analyses that we are going to perform, we use the covariance matrix derived from the galaxy catalogues described above (see [Sec. 4.3.1](#)). In particular, we use 1400 mocks to estimate the covariance matrix when running the MCMC sampling on the full data-vector. We use 700 when the analysis is performed using the compressed data-vector.

The largest scales considered in this work are  $k_{\min} = 0.03 h/\text{Mpc}$  for both power spectrum monopole and quadrupole and  $k_{\min} = 0.02 h/\text{Mpc}$  for the bispectrum monopole. The smallest scales considered are  $k_{\max} = 0.09 h/\text{Mpc}$  and  $k_{\max} = 0.12 h/\text{Mpc}$  for power spectrum (monopole and quadrupole) and bispectrum monopole respectively. The lower  $k_{\max}$  used for the power spectrum is due to the fact that we did not include 1-loop corrections for it, hence we consider only scales belonging to the quasi-linear regime. We chose a higher  $k_{\max}$  for the bispectrum since we implemented the effective model developed by [Gil-Marín et al. \(2014\)](#) which works up to non-linear scales.

The fiducial cosmology chosen for the analysis corresponds to a flat- $\Lambda$ CDM model close to the one reported in [Planck Collaboration et al. \(2016a\)](#). In particular, we set  $\Omega_m(z=0) = 0.31$ ,  $\Omega_b(z=0) = 0.049$ ,  $A_s = 2.21 \times 10^{-9}$ ,  $n_s = 0.9624$ ,  $h_0 = 0.6711$ . In order to compute the covariance terms and the derivatives of the model necessary for the compression, we fix the fiducial value of the galaxy bias model parameters, the growth rate and the amplitude of dark matter fluctuations to the ones obtained by running a preliminary low-resolution MCMC ( $b_1 = 2.5478$ ,  $b_2 = 1.2127$ ,  $f = 0.7202$ ,  $\sigma_8 = 0.4722$ ). The Finger-of-God parameters for both power spectrum and bispectrum  $\sigma_{\text{FoG}}^{\text{B}}$  and  $\sigma_{\text{FoG}}^{\text{P}}$  have been set to zero after checking that for the range of scales considered (quasi-linear regime) they were compatible with zero. In [Section 4.8](#) we check that the choice of fiducial parameters used to compute the derivatives of the mean of the data-vector and the analytical covariance matrix does not significantly influence the results of the compression.

## 4.4 Data-vector and covariance matrix

In order to measure the power spectrum and bispectrum from the data and the mocks catalogues we use the estimators described in [Gil-Marín et al. \(2016a,b\)](#). These are based on the weighted field of density fluctuations ([Feldman et al., 1994](#)):

$$F_\lambda(\mathbf{r}) = \frac{w_{\text{FKP}}(\mathbf{r})}{I_\lambda^{1/2}} [w_c(\mathbf{r})n(\mathbf{r}) - \alpha n_{\text{syn}}(\mathbf{r})], \quad (4.1)$$

where  $w_c$  is the weight taking into account all the measurement systematics (redshift failure, fiber collision, target density variations),  $w_{\text{FKP}}$  (Feldman, Kaiser and Peacock) ensures the condition of minimum variance,  $n$  is the observed number density of galaxies,  $n_{\text{syn}}$  is the number density of objects in a synthetic catalogue and  $I_\lambda$  is the normalisation of the amplitude of the observed power ( $\lambda = 2, 3$  for power spectrum and bispectrum, respectively).  $\alpha$  is the ratio between weighted number of observed galaxies over the weighted number of objects in the synthetic catalogues.

### 4.4.1 Power spectrum monopole and quadrupole

The redshift-space galaxy power spectrum model adopted in this work is a linear one including redshift-space distortions (RSD) plus a damping function taking into account the Finger-of-God (FoG) effect:

$$P_g^s(k, \mu) = D_{\text{FoG}}^P(k, \mu, \sigma_{\text{FoG}}^P[z]) Z_1^s(\mathbf{k})^2 P_m^{\text{lin.}}(k), \quad (4.2)$$

where  $k$  is the module of the wave vector  $\mathbf{k}$  and  $\mu$  is the cosine of the angle between the wave vector and the line of sight. The standard redshift-space distortion kernels  $Z_i^s$  are reported in the Appendix of [Gil-Marín et al. \(2014\)](#) together with the FoG damping function expression.  $\sigma_{\text{FoG}}^P[z]$  is the FoG free parameter for the power spectrum. For the range of scales considered in this work the linear RSD model has been proved to be a good approximation ([Taruya et al., 2010](#)). The redshift-space galaxy power spectrum can be expanded in terms of Legendre polynomials using its dependence on  $\mu$ :

$$P_g^s(k, \mu) = \sum_{\ell=0}^{\infty} P_g^{(\ell)}(k) L_\ell(\mu), \quad (4.3)$$

where  $L_\ell(\mu)$  is the  $\ell$ -order Legendre polynomial. Almost all the signal is contained in the first two even multipoles, the monopole and the quadrupole ( $\ell = 0, 2$ ). These can be found by inverting the above expression:

$$P_g^{(\ell)}(k) = \frac{2\ell + 1}{2} \int_{-1}^{+1} d\mu P_g^s(k, \mu) L_\ell(\mu) . \quad (4.4)$$

#### 4.4.2 Analytical expression for $P_g^{(0,2)}$ covariance matrices

Defining an estimator as in Appendix 4.A.1, it is possible to derive the expression for the Gaussian term of the power spectrum monopole and quadrupole covariance matrices (Appendix 4.A.2):

$$C_G^{P_g^{(\ell)} P_g^{(\ell)}}(k_1; k_2) = \left( \frac{2\ell + 1}{2} \right)^2 \frac{2\delta_{12}^K}{N_p(k_1)} P_g^{(\ell)}(k_1)^2 , \quad (4.5)$$

where  $\delta_{12}^K$  is the Kronecker delta between  $k_1$  and  $k_2$ , while  $N_p(k_1)$  is the number of pairs of grid points inside the estimator integration volume in Fourier space  $V_k = 4\pi k^2 \Delta k$  (Scoccimarro et al., 1998) and it is proportional to an effective survey volume  $V_e$ . The  $V_e$  normalisation is used to obtain a closer match between the analytic and mocks covariance matrices. Please refer to Equations 4.18 and 4.27 for more details. We set the cross covariance between power spectrum monopole and quadrupole to zero.

#### 4.4.3 Bispectrum monopole

For the redshift-space galaxy bispectrum we adopt the effective model presented in Gil-Marín et al. (2014), which modifies the redshift-space distortion kernels derived from perturbations theory in order to better fit the data at non-linear scales (see the Appendix of the paper above for the full expressions). The tree level has also been corrected to take into account the Finger-of-God damping effect:

$$B_g^s(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = D_{\text{FoG}}^B(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \sigma_{\text{FoG}}^B[z]) \times \left[ Z_1^s(\mathbf{k}_1) Z_1^s(\mathbf{k}_2) Z_{2,\text{eff}}^s[\mathbf{k}_1, \mathbf{k}_2] P_m^{\text{lin.}}(k_1) P_m^{\text{lin.}}(k_2) + \text{cyc.} \right] , \quad (4.6)$$

where  $\sigma_{\text{FoG}}^B[z]$  is the FoG free parameter for the bispectrum. The monopole of the bispectrum corresponds to the average of all the possible orientations of a determinate triangle with respect to the line of sight. It can therefore be obtained by integrating over two angular coordinates:

$$B_g^{(0)}(k_1, k_2, k_3) = \frac{1}{4} \int_{-1}^1 d\mu_1 \int_{-1}^1 d\mu_2 B_g^s(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = \frac{1}{4\pi} \int_{-1}^1 d\mu_1 \int_0^{2\pi} d\phi B_g^s(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) , \quad (4.7)$$

where  $\mu_i$  is the cosine of the angle between the  $\mathbf{k}_i$  vector and the line of sight. The angle  $\phi$  is defined as  $\mu_2 \equiv \mu_1 x_{12} - \sqrt{1 - \mu_1^2} \sqrt{1 - x_{12}^2} \cos \phi$  and where  $x_{12}$  is the cosine of the angle between  $\mathbf{k}_1$  and  $\mathbf{k}_2$ . More details are given in Appendix 4.A.

#### 4.4.4 Analytical expression for $\mathbf{B}_g^{(0)}$ covariance matrix

In order to apply the compression methods presented in Chapter 3 we need an analytical expression for the bispectrum monopole covariance matrix. This allows us to compress a data-vector with an arbitrarily large number of triangle configurations, which on the contrary wouldn't be possible using a covariance matrix estimated from the galaxy mock catalogues. That is because in order to obtain an accurate numerical estimate of the covariance matrix, the number of simulations used must be much greater than the data-vector's dimension (Hartlap et al., 2007; Percival et al., 2014). As it has been shown in Chapter 3, compressing the power spectrum together with the bispectrum, or leaving it uncompressed, does not make any substantial difference in terms of recovered parameter constraints. However, it makes a huge difference in terms of complexity of the covariance matrix that one has to model analytically in order to compress the data-vector. Compressing the power spectrum as well (monopole and quadrupole) also requires modelling their covariance matrices together with the cross-covariance with the bispectrum monopole. Leaving them uncompressed just requires to model the bispectrum monopole covariance matrix. The expression for the Gaussian term of  $\mathbf{C}_G^{\mathbf{B}_g^{(0)}\mathbf{B}_g^{(0)}}$  is derived in Appendix 4.A.3 and reads:

$$\mathbf{C}_G^{\mathbf{B}_g^{(0)}\mathbf{B}_g^{(0)}}(k_1, k_2, k_3; k_4, k_5, k_6) = \frac{D_{123456}}{16\pi^2} \frac{V_e}{N_t(k_1, k_2, k_3)} \mathbf{P}_g^{(0)}(k_1) \mathbf{P}_g^{(0)}(k_2) \mathbf{P}_g^{(0)}(k_3), \quad (4.8)$$

where  $D_{123456}$  stands for all the possible permutations for which each side of the first triangle is equal to a side of the second one; it has the values (6, 2, 1) respectively for equilateral, isosceles and scalene triangles.  $N_t(k_1, k_2, k_3)$  is the number of independent triplets of grid points in the integration volume in Fourier space  $V_{k_{123}} \simeq 8\pi^2 k_1 k_2 k_3 \Delta k_1 \Delta k_2 \Delta k_3$ . For the values of the effective survey volume and the average galaxy density number used in computing the analytical covariance matrix, we adopt the values  $V_e = 2.43 \times 10^9 \text{ Mpc}^3$  and  $\bar{n}_g = 1.14 \times 10^{-4} \text{ Mpc}^{-3}$  used by Slepian et al. (2017a) for both power spectrum monopole/quadrupole and bispectrum monopole analytical covariance matrices. In practice we use the analytic expression of the covariance matrix only to determine the weights for the compression. Since all the terms considered scale as  $V_e^{-1}$  the effective volume acts only as a scaling factor not affecting the compression performance.

In order to describe the correlation between different triangles in our analytical model of the covariance matrix, we include also a non-Gaussian term of the bispectrum monopole covariance matrix. In the expansion of the bispectrum covariance matrix presented in the Appendix of Chapter 3, for the bispectrum monopole this corresponds to a term proportional to the product of two bispectra monopoles as shown in Appendix 4.A.4:

$$C_{\text{NG}}^{\text{B}_g^0 \text{B}_g^0}(k_1, k_2, k_3; k_4, k_5, k_6) = \frac{\delta_{34}^{\text{K}}}{16\pi^2} \frac{k_f^3}{4\pi k_3^2 \Delta k_3} \text{B}_g^{(0)}(k_1, k_2, k_3) \text{B}_g^{(0)}(k_3, k_5, k_6) + 8 \text{ perm.} \quad (4.9)$$

It is important to include a term modelling the correlation between different triangles since the number of possible configurations increases very quickly as the bin size decreases. We do not include a corresponding non-Gaussian term into the power spectrum monopole and quadrupole covariances, since the number of data points considered is relatively low, thus the separation between the  $k$  modules values is more than sufficient to assume that the correlation between two different modes  $k_i$  and  $k_j$  is negligible with respect to their variance (approximated by the Gaussian term on the diagonal of the covariance matrix).

#### 4.4.5 Analytical expression for $[\mathbf{P}_g^{(0,2)}, \mathbf{B}_g^{(0)}]$ cross-covariance matrix

Finally we also model the cross-covariance between power spectrum multipoles and bispectrum monopole as described in Appendix 4.A.5:

$$C^{\text{P}_g^{(\ell)} \text{B}_g^0}(k_1; k_2, k_3, k_4) = \frac{1}{2\pi} \left( \frac{2\ell + 1}{2} \right) \frac{\delta_{12}^{\text{K}}}{N_p(k_2)} \text{P}_g^{(\ell)}(k_2) \text{B}_g^{(0)}(k_2, k_3, k_4) + 2 \text{ perm..} \quad (4.10)$$

As done in Chapter 3, we made the assumption that the shot noise is well approximated by a Gaussian distribution (which is reasonable if the galaxy number density is fairly high). Therefore, we just modify the galaxy power spectrum expressions by adding a  $\bar{n}_g^{-1}$  term. We did not take into account the effect of the survey geometry in the theoretical covariance matrix expression, which would affect the large scales inducing an extra correlation among the modes. We leave the inclusion of this correction for future work. Please refer to [Howlett and Percival \(2017\)](#) for a more detailed study on how to include this effect in the covariance matrix.

## 4.5 Compression methods

In Chapter 3 we presented two compression methods and applied them to the galaxy bispectrum and power spectrum: MC-MP and PCA + MP. Both methods rely on the MOPED method presented in [Heavens et al. \(2000\)](#), which achieves maximal compression of the original data-vector by extending to the multiple parameters case the algorithm introduced in [Tegmark et al. \(1997\)](#). Using this MP compression it is possible to shrink an arbitrarily large data-vector  $\mathbf{x}$  to a compressed one  $\mathbf{y}$  having dimension equal to the number of model parameters considered preserving Fisher information. This is obtained by deriving a set of weights for the full data-vector for each model parameter. Taking the scalar product between the weighting vectors and the original full data-vector  $\mathbf{x}$  gives the elements  $y_i$  of the compressed data-vector. Here we report only the main equations, please refer to Chapter 3 for more details. The weighting vector for each parameter  $\theta_i$  is given by:

$$\mathbf{b} = \mathbf{Cov}_x^{-1} \langle \mathbf{x} \rangle_{,i}, \quad (4.11)$$

where  $\mathbf{Cov}^{-1}$  is the inverse of the original full data-vector covariance matrix and  $\langle \mathbf{x} \rangle_{,i}$  is the derivative with respect to the model parameter  $\theta_i$  of the mean of the modelled data-vector  $\mathbf{x}$ , computed at a fiducial parameter vector  $\boldsymbol{\theta}_{\text{fid}}$ . In our case the fiducial values are reported in Section 4.3.2. Therefore, the elements of the compressed data-vector  $\mathbf{y}$  are given by:

$$y_i = \langle \mathbf{x} \rangle_{,i}^\top \mathbf{Cov}_x^{-1} \mathbf{x} \equiv \mathbf{b}^\top \mathbf{x}. \quad (4.12)$$

In the MC-MP method a MCMC sampling algorithm using  $\mathbf{y}$  as data-vector is ran after compression. An estimate of the compressed covariance matrix from the mock catalogues can be obtained as shown in the Appendix of Chapter 3:

$$\mathbf{Cov}_{\mathbf{y},ij} = \mathbf{Cov}[y_i, y_j] = \mathbf{b}_i^\top \cdot \mathbf{Cov}_x \cdot \mathbf{b}_j, \quad (4.13)$$

where  $\mathbf{Cov}_x$  is the original covariance matrix.

### 4.5.1 PCA + MP

As described in Chapter 3, instead of orthogonalising the weights as in [Zablocki and Dodelson \(2016\)](#), we perform a principal component analysis (PCA) transformation of our parameter space before applying the MP compression. This is

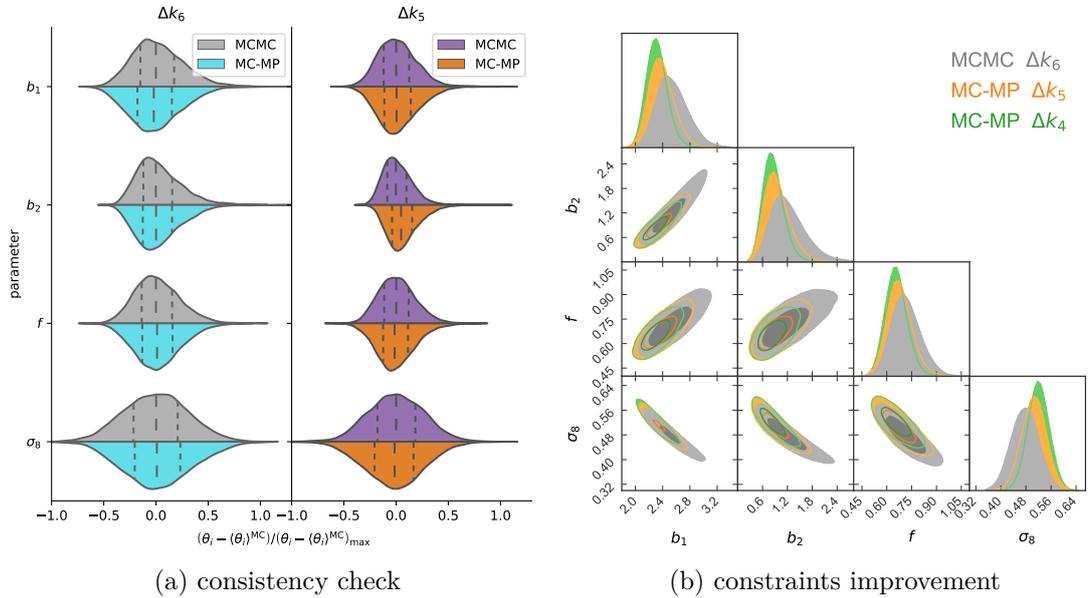


Figure 4.1: Joint data-vector  $[P_g^{(0)}, P_g^{(2)}, B_g^{(0)}]$  posteriors: MC-MP four-parameter case.

**a)** the violin plots show for two test cases ( $\Delta k_6$  and  $\Delta k_5$  binning) the comparison between the 1D posterior densities obtained via MCMC and MC-MP for all parameters. The vertical lines represent the 25%, 50% and 75% quartiles. All distributions have been centered by subtracting the mean value obtained from the MCMC analysis and they have been normalised by dividing by the maximum difference between the parameter value of each sample and the mean of the distribution. Even if the 1D distributions are not Gaussian, the agreement between MCMC and MC-MP results is qualitatively good. For a quantitative comparison see Table 4.1 and additionally Figure 4.B.1 and 4.B.2 in Appendix 4.B.

**b)** the 2D 68% and 95% credible regions are shown in order to highlight the improved constraints obtained by employing a higher number of triangles in the data-vector thanks to the compression with respect to the standard MCMC for the full data vector. In particular, the grey contours correspond to the standard binning  $\Delta k_6$  used to run the MCMC for the full data-vector. The orange and green contours correspond to the distributions for the compressed data-vector for the binnings  $\Delta k_5$  and  $\Delta k_4$  (which corresponds to  $N_{\text{triangles}} = 195, 404$ ). See also Table 4.2.

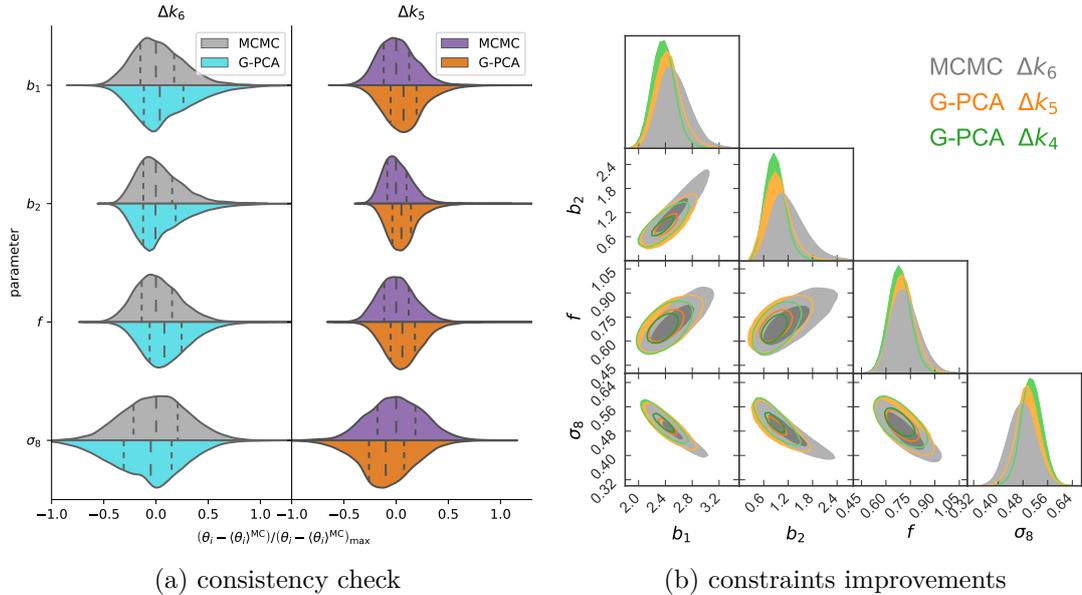


Figure 4.2: Joint data-vector  $[P_g^{(0)}, P_g^{(2)}, B_g^{(0)}]$  posteriors: G-PCA four-parameters case. Same as Figure 4.1 but for the G-PCA method.

done by diagonalising the Fisher information matrix using the eigenvalue decomposition

$$F_{\theta_{\text{phys.}}} = \mathbf{P} F_{\theta_{\text{PCA}}} \mathbf{P}^T \quad \text{where} \quad \theta_{\text{PCA}} = \mathbf{P}^T \theta_{\text{phys.}}, \quad (4.14)$$

and  $\mathbf{P}$  is the linear transformation matrix. After having diagonalised the Fisher matrix we compress the data-vector with respect to this new set of parameters  $\theta_{\text{PCA}}$ . The effect of a PCA decomposition is to rotate the parameter space to the axes corresponding to the degeneracies between the original set of parameters. Therefore, taking the outer product of the 1D posteriors of the parameters  $\theta_{\text{PCA}}$  in order to get the multidimensional posterior distribution should return a good approximation to the one sampled by the MCMC code.

Since the  $\theta_{\text{PCA}}$  are uncorrelated, one can randomly sample the 1D posteriors and then rotate the resulting parameter vector using  $\mathbf{P}$  back into the physical space. Doing this avoids the use of the MCMC sampling altogether.

As shown in Chapter 3, this works only for those parameter sets which have a sufficiently low degree of degeneracy such that the approximation of Gaussianity for the multidimensional posterior can be assumed to be valid (no or very weak "banana-shaped" contours). Since this is not always the case, as for our choice of parameters, an additional Gaussianisation pre-step is required.

### 4.5.2 Gaussianisation pre-step

In Chapter 3 the PCA + MP method assumed that it was possible to rotate through a linear transformation the physical parameter space into a new one where the new parameters are orthogonal/uncorrelated between each other. In order to be able to deal with distributions containing non-linear degeneracies (e.g. "banana-shaped" contours), we add a pre-Gaussianisation transformation of the parameter space using the procedure described in [Schuhmann et al. \(2016\)](#). In their work they introduced an extension of the Box-Cox transformations, which are functions of two parameters  $(a, \lambda)$ :

$$\tilde{\theta}^i = BC_{(a,\lambda)}(\theta^i) = \begin{cases} \lambda^{-1}[(\theta^i + a)^\lambda - 1] & (\lambda \neq 0) \\ \log(\theta^i + a) & (\lambda = 0) \end{cases} \quad (4.15)$$

where  $\tilde{\theta}^i$  is the transformed  $i$ -th model parameter while  $\theta^i$  is the original  $i$ -th model parameter. Their method was labelled Arcsinh-Box-Cox transformation (ABC). For each of the model parameters, a set of three ABC transformation parameters  $(a, \lambda, t)$  are computed by the algorithm which are then used in the following way:

$$\theta_{\text{Gauss.}}^i = ABC(\theta_{\text{phys.}}^i) = \begin{cases} t^{-1} \sinh[t BC_{(a,\lambda)}(\theta_{\text{phys.}}^i)] & (t > 0) \\ BC_{(a,\lambda)}(\theta_{\text{phys.}}^i) & (t = 0) \\ t^{-1} \text{arcsinh}[t BC_{(a,\lambda)}(\theta_{\text{phys.}}^i)] & (t < 0) \end{cases} \quad (4.16)$$

where  $\theta_{\text{Gauss.}}^i$  is the Gaussianised  $i$ -th model parameter while  $\theta_{\text{phys.}}^i$  is the original  $i$ -th physical model parameter. We then relabel this compression as G-PCA. In order to obtain the transformation parameters of the Gaussianising transformations it is necessary to run a preliminary MCMC sampling using the full data-vector. What we want to prove is that once the transformation parameters have been obtained for the standard number of triangles corresponding to the  $\Delta k_6$  binning case, these are valid also for a higher number of triangle configurations included in the bispectrum.

### 4.5.3 Analytical covariance matrix: usage

In the following analysis, we are going to use two different options for the analytical covariance matrices. For the MC-MP method we compress only the bispectrum monopole part of the data-vector. To derive the weights in Equation 4.11 we use the analytical covariance matrix of the bispectrum monopole given by the sum of the Gaussian term in Equation 4.8 and the non-Gaussian one given in Equation 4.9. For the G-PCA method the full data-vector needs to be

compressed since the computation of the 1D posteriors of the  $\theta_{PCA}$  parameters requires each data vector element to be sensitive to the variation of just one  $\theta_{PCA}$  parameter, as explained in Chapter 3. Therefore, for the power spectrum monopole/quadrupole we use Equation 4.5 as our analytical covariance matrix; similarly for the bispectrum monopole we use Equation 4.8 for the covariance matrix (the same as the one we used for the MC-MP case), and finally, we use Equation 4.10 for our cross-covariance matrix.

## 4.6 Recover MCMC-derived posterior distribution

For MCMC sampling we use EMCEE<sup>8</sup> (Foreman-Mackey et al., 2013). All the likelihoods have been corrected as suggested by Sellentin and Heavens (2016) in order to take into account the bias induced by estimating the inverse of the real covariance matrix from a limited number of mocks. In order to check whether our analytical estimate of the covariance matrix is good enough to be used for deriving the weights as explained in Sec. 4.5, we compare to the full MCMC 1D posterior distributions in the left panels of Figures 4.1 and 4.2 with results from the MCMC+ MC-MP and G-PCA methods, respectively.

The violin plots include the standard binning case  $\Delta k_6$  (116 triangles) and the  $\Delta k_5$  case (195 triangles). For these two cases we compare the MCMC (grey and purple) with the compression results (cyan and orange). From each point we subtract the mean of the model parameters obtained using the MCMC. This makes it easier to check that the shift in the mean of the compression results with respect to the MCMC ones is small when compared to the size of the inner quartiles of the distribution. This concept is also quantified in the bottom half of Table 4.1, which shows the shifts in the mean values is relative to the 1D 68% credible intervals. In the top half of Table 4.1 we report the precise values of both the means and the 68% credible intervals for all model parameters. Additionally, Figure 4.B.1 in Appendix 4.B shows the comparison between the 2D MCMC posterior distributions and the MC-MP and G-PCA ones for both  $\Delta k_6$  and  $\Delta k_5$  cases. We conclude that even if a small part of the constraining power is lost (see the  $\Delta k_6$  columns in Table 4.2 for details), both compression methods return posterior distributions which well agree with the MCMC distribution for all model parameters under consideration.

---

<sup>8</sup> We use 192 walkers, 1100 burn-in steps and 1700 steps. For the low-resolution MCMC we use half of the previous quantities.

Table 4.1: Four parameter-case, check consistency.

**Upper half:** Mean values of the posterior distributions and 68% credible intervals for the MCMC and the MC-MP / G-PCA compression methods. We report the values for a range of  $k$ -binnings. From the largest bin  $\Delta k_6$ , the size used in the BOSS analysis, corresponding to the lowest number of triangles (116), to the thinnest binning  $\Delta k_2$  corresponding to the highest number of triangles (2734). The observed shift in the mean values as a function of the number of triangles considered is due to the strong the degeneracy present between the model parameters. As can be seen in Figure 4.4, the shift does not have any effect on the goodness of fit.

**Lower half:** In the compression columns we report the relative difference between the posterior modes obtained via MCMC and the ones obtained via compression (MC-MP or G-PCA). In the MCMC columns the relative size of the 68% credible intervals obtained via MCMC sampling is shown. Comparing the MCMC columns to the compression ones, it is that the difference between the mean parameter values obtained via MCMC and the ones obtained via compression (MC-MP or G-PCA) are evidently within the 68% credible intervals given by the MCMC on the full data-vector.

	$\Delta k_6$			$\Delta k_5$			$\Delta k_4$			$\Delta k_2$		
	MCMC	MC-MP	G-PCA									
$b_1$	$2.41 \pm 0.22$	$2.41 \pm 0.23$	$2.49 \pm 0.27$	$2.34 \pm 0.17$	$2.38 \pm 0.18$	$2.42 \pm 0.17$	$2.27 \pm 0.14$	$2.27 \pm 0.14$	$2.38 \pm 0.16$	$2.33 \pm 0.14$	$2.33 \pm 0.14$	$2.31 \pm 0.17$
$b_2$	$1.00 \pm 0.40$	$1.04 \pm 0.42$	$1.08 \pm 0.47$	$0.82 \pm 0.26$	$0.83 \pm 0.29$	$0.85 \pm 0.26$	$0.79 \pm 0.23$	$0.79 \pm 0.23$	$0.81 \pm 0.22$	$0.72 \pm 0.22$	$0.72 \pm 0.22$	$0.77 \pm 0.19$
$f$	$0.69 \pm 0.08$	$0.72 \pm 0.09$	$0.72 \pm 0.09$	$0.67 \pm 0.07$	$0.67 \pm 0.07$	$0.70 \pm 0.07$	$0.65 \pm 0.06$	$0.65 \pm 0.06$	$0.68 \pm 0.06$	$0.63 \pm 0.06$	$0.63 \pm 0.06$	$0.67 \pm 0.06$
$\sigma_8$	$0.50 \pm 0.04$	$0.48 \pm 0.05$	$0.48 \pm 0.05$	$0.51 \pm 0.04$	$0.50 \pm 0.04$	$0.49 \pm 0.03$	$0.53 \pm 0.03$	$0.53 \pm 0.03$	$0.51 \pm 0.03$	$0.53 \pm 0.03$	$0.53 \pm 0.03$	$0.51 \pm 0.03$
	$\frac{\Delta\theta_{\Delta k_6}^{\text{MCMC}}}{\theta_{\Delta k_6}^{\text{MCMC}}} [\%]$	$\frac{\theta^{\text{comp.}} - \theta_{\Delta k_6}^{\text{MCMC}}}{\theta_{\Delta k_6}^{\text{MCMC}}} [\%]$	$\frac{\Delta\theta_{\Delta k_6}^{\text{G-PCA}}}{\theta_{\Delta k_6}^{\text{G-PCA}}} [\%]$	$\frac{\Delta\theta_{\Delta k_5}^{\text{MCMC}}}{\theta_{\Delta k_5}^{\text{MCMC}}} [\%]$	$\frac{\theta^{\text{comp.}} - \theta_{\Delta k_5}^{\text{MCMC}}}{\theta_{\Delta k_5}^{\text{MCMC}}} [\%]$	$\frac{\Delta\theta_{\Delta k_5}^{\text{G-PCA}}}{\theta_{\Delta k_5}^{\text{G-PCA}}} [\%]$	$\frac{\Delta\theta_{\Delta k_4}^{\text{MCMC}}}{\theta_{\Delta k_4}^{\text{MCMC}}} [\%]$	$\frac{\theta^{\text{comp.}} - \theta_{\Delta k_4}^{\text{MCMC}}}{\theta_{\Delta k_4}^{\text{MCMC}}} [\%]$	$\frac{\Delta\theta_{\Delta k_4}^{\text{G-PCA}}}{\theta_{\Delta k_4}^{\text{G-PCA}}} [\%]$	$\frac{\Delta\theta_{\Delta k_2}^{\text{MCMC}}}{\theta_{\Delta k_2}^{\text{MCMC}}} [\%]$	$\frac{\theta^{\text{comp.}} - \theta_{\Delta k_2}^{\text{MCMC}}}{\theta_{\Delta k_2}^{\text{MCMC}}} [\%]$	$\frac{\Delta\theta_{\Delta k_2}^{\text{G-PCA}}}{\theta_{\Delta k_2}^{\text{G-PCA}}} [\%]$
$b_1$	9.2	-0.3	3.3	7.3	1.9	3.5	-2.7	-2.7	1.9	-2.7	-2.7	-1.1
$b_2$	40.3	3.5	7.5	32.2	1.9	4.4	-3.6	-3.6	-1.2	-16.5	-16.5	-5.7
$f$	12.1	4.4	4.4	10.1	-1.3	3.8	-3.3	-3.3	0.2	0.5	0.5	-1.1
$\sigma_8$	8.5	-5.1	-5.5	7.3	-1.1	-3.6	4	4	-0.3	2.2	2.2	-1.2

Table 4.2: Four-parameter case, constraints improvement. Below are shown the relative variations in percentage of the size of the 68% credible intervals as a function of the  $k$ -binning considered (number of triangle configurations used for the bispectrum monopole). In orange and green are highlighted respectively the improvements achieved via compression for the  $\Delta k_5$  and at the saturation level (404 triangles -  $\Delta k_4$ ) of the bispectrum monopole constraining power case for the considered set of parameters (e.g. left panel of Figure 4.3). Finally in blue and red are highlighted the improvements obtained via compression for the highest number of triangles considered (2734 triangles -  $\Delta k_2$  binning) for MC-MP and G-PCA respectively.

	$\Delta k_6$		$\Delta k_5$		$\Delta k_4$		$\Delta k_2$	
	MCMC	$\frac{\Delta\theta_{\Delta k_6}^{\text{comp.}} - \Delta\theta_{\Delta k_6}^{\text{mc}}}{\Delta\theta_{\Delta k_6}^{\text{mc}}} [\%]$	MCMC	$\frac{\Delta\theta_{\Delta k_5}^{\text{comp.}} - \Delta\theta_{\Delta k_5}^{\text{mc}}}{\Delta\theta_{\Delta k_5}^{\text{mc}}} [\%]$	MC-MP	$\frac{\Delta\theta_{\Delta k_4}^{\text{comp.}} - \Delta\theta_{\Delta k_4}^{\text{mc}}}{\Delta\theta_{\Delta k_4}^{\text{mc}}} [\%]$	MC-MP	$\frac{\Delta\theta_{\Delta k_2}^{\text{comp.}} - \Delta\theta_{\Delta k_2}^{\text{mc}}}{\Delta\theta_{\Delta k_2}^{\text{mc}}} [\%]$
$\Delta b_1$	0.22	4.4	0.17	-21.3	-35.3	-30.0	-37.1	-24.8
$\Delta b_2$	0.40	2.9	0.26	-28.9	-42.6	-46.0	-46.1	-52.8
$\Delta f$	0.08	3.7	0.07	-16.5	-24.7	-25.1	-27.8	-26.4
$\Delta\sigma_8$	0.04	6.5	0.04	-11.3	-22.3	-24.5	-22.8	-21.0

## 4.7 Information content and number of triangles

The right panels of Figures 4.1 and 4.2 show how using a larger number of triangles tightens the posterior contours of the four model parameters considered. At the same time, the maxima of the 2D posterior distributions converge to the same values for each compression method as the number of triangles is increased.

Note that the shift in the posterior distribution between binning cases is not an artifact of the compression: it is also present when we fit using the standard MCMC method. This can be seen when comparing the location and shape of the 2D contour regions in Figures 4.B.1 and 4.B.2 in Appendix 4.B for the  $\Delta k_6$  and  $\Delta k_5$  binning cases. Quantitatively it can be observed by comparing means and standard deviations in Table 4.1. Thus, both compression algorithms reproduce posterior distributions very similar to the ones derived via MCMC sampling for the relevant binning cases  $\Delta k_6$  and  $\Delta k_5$ . The observed shift between binning cases is due to the strong degeneracy between the model parameters. In particular the shift happens along the degeneration direction of  $b_1, b_2$  and  $f$  with  $\sigma_8$ . It may have a statistical origin. Further checks on this effect may be performed using the galaxy mocks, for example by fitting several different realizations for both the  $\Delta k_6$  and  $\Delta k_5$  binning cases using the G-PCA method (which would be much faster than doing parameter estimation via MCMC or MC-MP). We reserve to do these tests in future work. The main result of this Chapter 3s that the variance of the parameters is reduced when the number of triangles used increases.

For future surveys the compression can be then used for the main analysis and also to find out the minimum number of triangle configurations for a given  $k$ -range needed to fully capture the non-Gaussian information contained in 3pt statistics like the bispectrum. The later will indicate how many mock catalogues/simulations are required in order to accurately estimate the covariance matrix. In our analysis the saturation seems to be reached already for the  $\Delta k_4$  binning case (404 triangles).

For what concerns  $\Delta k_2$ , the smallest  $k$ -bin size considered (2734 triangles), Tables 4.1 and 4.2 show that the  $\Delta k_2$  posterior distribution is very similar to the  $\Delta k_4$  case.

The trend in the information content in terms of the 1D 68% credible intervals as a function of the triangle number used is shown in the left panel of Figure 4.3, and the improvement quantified in Table 4.2. From Figure 4.3 it appears that the parameters constraints improvement as a function of the number of triangles reaches the saturation already for the  $\Delta k_4$  case. In terms

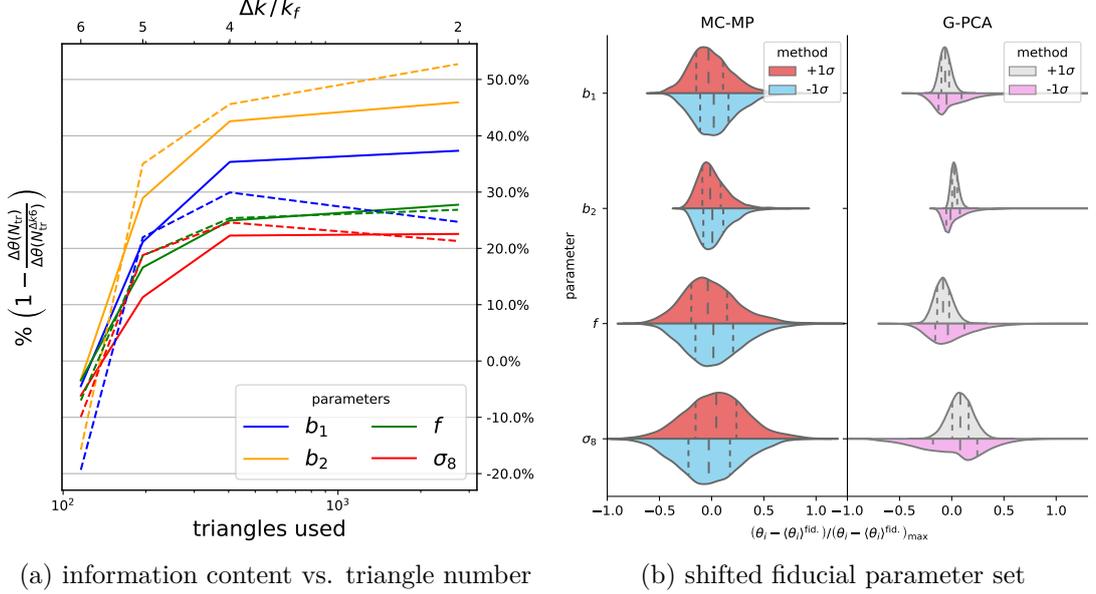


Figure 4.3: **a)** the 1D 68% credible intervals as a function of the number of triangles used in the bispectrum monopole data-vector. Continuous lines represent the MC-MP results while the dashed ones are given by the G-PCA compression method.

**b)** the compression results for the MC-MP and G-PCA cases when the fiducial parameter set used to compute the analytical covariance matrix and the derivatives of the mean are shifted by  $\pm 1\sigma$  credible intervals. The violin plots show, for the test case of the  $\Delta k_6$ -binning, the comparison between the 1D posterior distributions for all parameters, using shifts by  $+1\sigma$  (red/grey) and  $-1\sigma$  (blue/pink) for the MC-MP / G-PCA methods. The vertical lines represent the 25%, 50% and 75% quartiles. All distributions are mean-subtracted using the fiducial parameter set for the compression, and they have been normalised by the maximum difference between the parameter value of each sample and the mean of the distribution. Even if the 1D distributions are not Gaussian, the effect of compressing with respect a shifted cosmology is qualitatively negligible for the MC-MP method while it affects the G-PCA performance more. Nevertheless, the modifications to the fiducial parameter sets are substantial ( $\sim 10 - 40\%$  variations) given the broad posteriors due to the strong degeneracy in the parameter set.

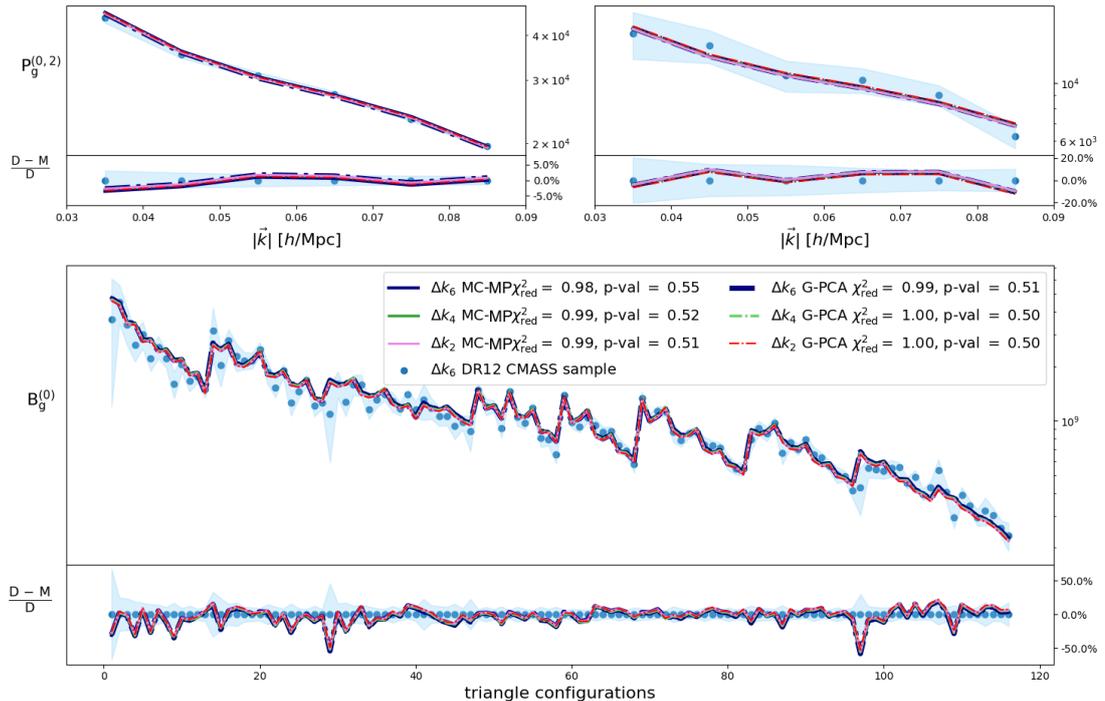


Figure 4.4: Reduced  $\chi^2$  and  $p$ -values for the best-fit models obtained using the MCMC, MC-MP and G-PCA compression methods. The  $k$ -binnings shown are respectively the standard  $\Delta k_6$  (navy), an intermediate size  $\Delta k_4$  (green) and the smallest one  $\Delta k_2$  (pink for MC-MP and red for G-PCA) corresponding to the highest number of triangle used in the bispectrum monopole. The two upper panels are for the power spectrum monopole (left) and quadrupole (right) while the bottom panel refers to the bispectrum monopole. The lower part of each panel shows the relative difference between the data measurements and the different models. Even if for example  $b_1$  and  $\sigma_8$  values are shifted between the cases of  $\Delta k_6$  and  $\Delta k_2$ , the strong degeneracy has the result of making the two models practically identical.

of percentages of the original 1D 68% credible intervals obtained running an MCMC on the full data-vector for the parameters  $(b_1, b_2, f, \sigma_8)$  in the BOSS  $\Delta k_6$  case, the  $\Delta k_2$  MC-MP and G-PCA analyses obtain tighter constraints by  $(-35\%, -45.3\%, -22.6\%, -22.6\%)$  and  $(-24.8\%, -52.8\%, -26.4\%, -21\%)$ , respectively. These optimal constraints as obtained by the compression methods are also shown in summary Figure 4.5.

## 4.8 Consistency check

In order to test the validity of our analysis, we compute the reduced  $\chi^2$  and corresponding  $p$ -value for each set of parameters obtained using either the MCMC sampling or the compression methods. For all parameter vectors (compressed and uncompressed) this has been done using the data-vector corresponding to the standard  $\Delta k_6$  binning. The results can be seen in Figure 4.4. This test proves

that the shift observed in the parameters as the number of triangle configurations is increased is simply due to the strong degeneracy present between  $b_1$ ,  $b_2$ ,  $f$  and  $\sigma_8$ . Indeed both the reduced  $\chi^2$  and  $p$ -values show that all these models fit the data very well. In Figure 4.4 we did not show the lines and statistics for the  $\Delta k_5$  cases just for the sake of clarity and because the results are equivalent to those of the other binnings. From the same figure it can also be noticed that the tightest errorbars are those from the power spectrum case.

To demonstrate the flexibility of the compression methods we check their performance when the fiducial parameter set is shifted by  $\pm 1\sigma$  credible intervals in the  $\Delta k_6$  case. The effect of this is shown in the right panel of Figure 4.3. For this plot, we centre each 1D distribution by subtracting the mean obtained by running the compression pipelines using the fiducial parameters values. In this way it is possible to observe by how much the posterior distributions derived via MC-MP or G-PCA shift as a function of the chosen fiducial parameter set. In Appendix 4.B the precise numbers are reported in Table 4.B.1.

MC-MP appears to be more stable than the G-PCA when the fiducial parameter set is shifted. The explanation of this could be the fact that G-PCA involves several transformations of the parameter space, including a diagonalisation of the Fisher information matrix which is computed from the analytical model of the covariance matrix.

Nevertheless, it should be noted that we are testing the performances of the compression in a regime of strong degeneracy of the parameter space and therefore shifting the fiducial parameter set by  $\pm 1\sigma$  credible intervals actually means increasing/reducing the individual values by  $\sim 10 - 40\%$  (second panel Table 4.1). Therefore, running a preliminary low-resolution MCMC sampling on the full data-vector (which can be shorter than the one that will be later compressed, as we have done in our analysis) is an efficient solution to determining a reasonable fiducial model for deriving the compression.

### 4.8.1 Comparison with BOSS DR12 bias constraints

BOSS galaxy sample results from the bispectrum are reported by Gil-Marín (2017) [in Table 3 at p. 18] from the same CMASS sample data set, at the same redshift, for the following parameter combinations:  $b_1\sigma_8 = 1.2479 \pm 0.0072$ ,  $b_2\sigma_8 = 0.641 \pm 0.066$  and  $f\sigma_8 = 0.432 \pm 0.018$ <sup>9</sup>. If we recast our results obtained using the MCMC for the  $\Delta k_6$  case in terms of the same parameter combinations these are:  $b_1\sigma_8 = 1.203 \pm 0.008$ ,  $b_2\sigma_8 = 0.557 \pm 0.140$  and  $f\sigma_8 = 0.339 \pm 0.019$ .

<sup>9</sup>we compare our results with the BOSS analysis standard deviation values obtained considering only the statistical contributions and not the systematics ones.

In the BOSS analysis a larger range of scales has been considered. In particular, BOSS analysis goes up to  $k \sim 0.2h/\text{Mpc}$  for both power spectrum monopole/quadrupole and bispectrum monopole while we stop at  $k \sim 0.09h/\text{Mpc}$  and  $k \sim 0.12h/\text{Mpc}$ , respectively. This could explain the larger value we obtained for  $b_2\sigma_8$ . A more complex model for the power spectrum was used in the BOSS analysis, including loop corrections beyond the tree level approximation. Moreover the BOSS analysis also modelled the effect of the survey window function for both power spectrum and bispectrum.

As we saw from Figure 4.4, the power spectrum monopole is the most constraining part of the full data-vector, having errorbars of less than 5%. Therefore the tree-level expression for the power spectrum severely limits the possibility of using the power spectrum measurements to reduce the degeneracy between the model parameters since it is inadequate to model quasi-non-linear and non-linear scales. This could explain the high values obtained for the bias parameters  $b_1$ ,  $b_2$  and the growth rate  $f$  together with the low value for  $\sigma_8$  with respect to the BOSS analysis results. The additional one-loop corrections for the power spectrum would have indeed supplied additional amplitude which instead, by including only the tree-level term, needs to be achieved by having higher values of  $b_1$  and  $f$ .

Moreover, in the BOSS analysis the FoG parameters  $\sigma_{\text{FoG}}^{\text{B}}$  and  $\sigma_{\text{FoG}}^{\text{P}}$  were left free to vary in order to better model the non-linear regime and were detected with high significance ( $\sigma_{\text{FoG}}^{\text{B}} = 7.54 \pm 0.70$  and  $\sigma_{\text{FoG}}^{\text{P}} = 3.50 \pm 0.14$ ). The BOSS model also included a noise-amplitude parameter  $A_{\text{noise}}$  which modelled divergence from Poissonian shot noise. In our model we had included  $A_{\text{noise}}$  initially, however we set it to zero after having checked that, if let free to vary, its posterior distribution was compatible with zero. These differences in the modelling and scales considered could explain the discrepancy in the best-fitting parameters. For more details see Appendix 4.B and in particular Figure 4.B.3.

### 4.8.2 Difference in time and computer resources needed

There is no significant difference between MCMC and MC-MP in terms of time taken for the pipeline to run or computing resources needed. For the parameter set  $(b_1, b_2, f, \sigma_8)$  the running time varied between 20 minutes for 116 triangles to  $\sim 10$  hours for 2734 triangles on 14 2.2 GHz Intel i7 cores. G-PCA proved to be faster when many triangle configurations are used. Considering  $\sim 30$  minutes for the preliminary MCMC with 116 triangles and  $\sim 2$  hours for the Gaussianisation part, it took between  $\sim 5$  minutes (116 triangles) and  $\sim 30$  minutes (2734

triangles) using only one 2.2 GHz Intel i7 core for the compression plus posterior evaluation to run. Therefore, by running once the preliminary MCMC and Gaussianisation algorithm, we were able to run the PCA part for all the binning cases considered in less than in total  $\sim 3$  hours wall-clock time.

We used CAMB (Lewis et al., 2000) to compute the linear matter power spectrum. The time difference between MCMC/MC-MP and G-PCA would have been much more significant in the case of a parameter set for which the linear matter power spectrum needs to be recomputed for every model realisation.

## 4.9 Conclusions

In this paper we have shown the results of applying both compression methods for the galaxy redshift-space bispectrum, presented in Chapter 3, to the measurements from the SDSS-III BOSS DR12 CMASS sample (Gil-Marín, 2017). We considered as original data-vector the combination of the power spectrum monopole and quadrupole with the bispectrum monopole, which are obtained by averaging over the angles describing the orientation with respect to the line of sight. The first method called MC-MP consists of running an MCMC sampling on the compressed data-vector obtained by taking the scalar product between the original data-vector and a set of weights derived as first shown by Tegmark et al. (1997). The second method, which we denoted as G-PCA, is the modification of the PCA + MP method presented in Chapter 3 obtained by adding a Gaussianisation transformation of the parameter set (Schuhmann et al., 2016) before rotating it using a principal component analysis transformation (PCA) followed by the MP compression. By transforming the physical parameter space into an orthogonal one it is possible to just randomly sample 1D posterior distributions, avoid altogether the need of running a MCMC routine.

In order to derive the posterior distributions for the set of parameter considered, the galaxy bias parameters  $b_1$  and  $b_2$ , the growth rate  $f$  and the normalisation of the dark matter perturbations amplitude  $\sigma_8$ , we numerically estimated the covariance matrix using 1400 and 700 galaxy mocks catalogues for the full data-vector and compressed data-vector cases, respectively.

The following points represent the main conclusions of our analysis:

- In order to obtain the weights for the compression methods we derived an analytic approximation of the leading terms of the covariance matrix relative to the considered data-vector. The final expressions of these computations are reported in Sec. 4.4 while the full derivations are shown in Appendix 4.A.

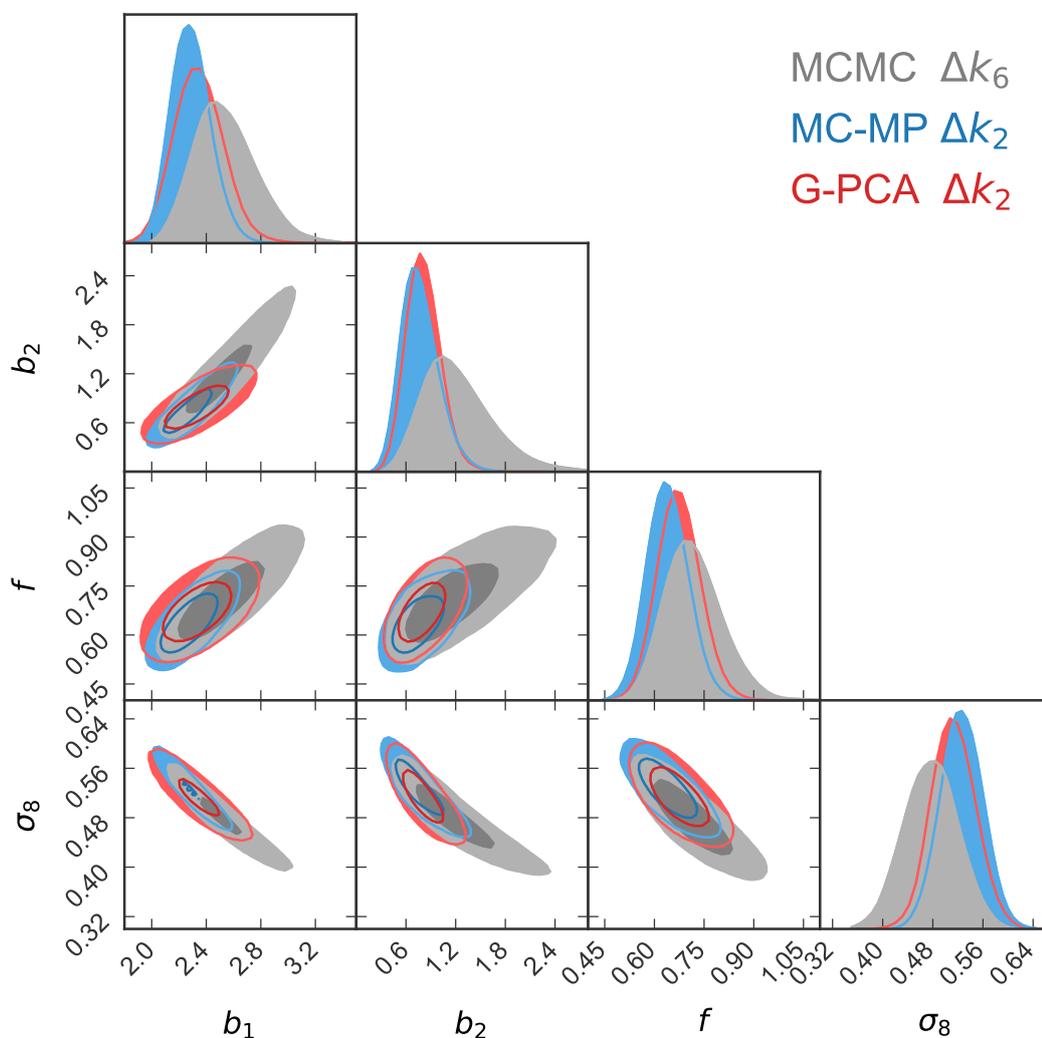


Figure 4.5: MCMC vs. MC-MP vs. G-PCA. 2-D 68% and 95% credible contours are shown respectively for the  $\Delta k_6$  MCMC (grey),  $\Delta k_2$  MC-MP (blue) and  $\Delta k_2$  G-PCA (red) cases. It is possible to observe the substantial improvement in parameter constraints through applying either compression method to a data-vector containing approximately  $\sim 23$  times more triangles than the one used for the MCMC sampling case. The agreement between the MC-MP and G-PCA posterior distributions is remarkable. Using more triangles helps with lifting the strong degeneracy between the model parameters, as can be seen from the shrinkage of the 2-D contours along the degeneracy directions.

- In Sec. 4.6 we have shown that both compression methods recover the posterior distributions obtained via MCMC using the full data-vector with little loss of information ( $\sim 4\%$  and  $\sim 13\%$  larger 68% credible intervals than the MCMC ones in average for MC-MP and G-PCA, respectively). More importantly, even if slightly broader, the posterior distributions recovered through compression have the same shape and modes as the MCMC counterparts.
- Adding a pre-Gaussianisation step removes the PCA + MP limitation linked to a strongly degenerate parameter space described in Chapter 3. It is however necessary to run a preliminary MCMC in order to derive the Gaussianisation transformation parameters. Nevertheless, once these parameters have been derived for a number of triangles case for which it is possible to run an MCMC on the full data-vector, they can then be used to compress a data-vector with an arbitrary number of triangles.
- In Sec. 4.7 we show the main result of this work, namely the substantial improvement in parameter constraints obtained by compressing a much larger number of triangles with respect to standard MCMC data-vector. For the uncompressed data-vector the number of triangles is limited by the number of mock catalogues available to estimate the covariance matrix. For both compression methods and for any number of triangle configuration considered, the dimension of the compressed data-vector is always equal to the number of model parameters constrained.

For the highest number of triangles considered, this leads to an improvement in terms of the 68% 1D credible intervals by  $(-37\%, -46\%, -27\%, -23\%)$  and  $(-25\%, -53\%, -26\%, -21\%)$  for the MC-MP and G-PCA methods, respectively.

- By way of summary, in Figure 4.5 we show the results for both MC-MP and G-PCA methods using 2734 triangles and for the MCMC on the uncompressed data-vector containing 116 triangles. The two compression methods agree well and produce substantially tighter and less degenerate constraints. Furthermore the G-PCA approach allowed for a computational speed up, requiring only approximately a third of the time taken by the MCMC and MC-MP methods, including also the low-resolution MCMC necessary for the Gaussianisation transformation. Considering only the PCA part, the speed up factor rises to  $\sim 20 - 100$  times depending on the parameter set considered.

- Finally we would like to point out that the compressing methods used in this work represents a straightforward approach to include higher-order statistics like the trispectrum or the tetraspectrum in the analysis of current and future data sets. This is due to the fact that the number of elements of the data-vector, after the maximal compression, corresponds exactly to the number of model parameters. Both MC-MP and G-PCA have the potential to fully exploit the constraining power of higher-order statistics applied to data-sets from future surveys like DESI, EUCLID and PFS.

# Appendix

## 4.A Estimators and covariance terms

### 4.A.1 Power spectrum monopole/quadrupole and bispectrum monopole estimators

The analytical model for the redshift-space galaxy power spectrum monopole and quadrupole is given by equation 4.4.

It is therefore natural to define the estimator as:

$$\hat{\mathbb{P}}_g^{(\ell)}(k) = \left(\frac{2\ell+1}{2}\right) \frac{1}{(2\pi)^3 N_p(k)} \int_{V_p} \int_{V_q} d^3\mathbf{p} d^3\mathbf{q} L_\ell(\mu) \delta_D(\mathbf{q} + \mathbf{p}) \delta_g^s(\mathbf{q}) \delta_g^s(\mathbf{p}), \quad (4.17)$$

where  $V_{\mathbf{p},\mathbf{q}}$  are the spherical shell volumes characterised by  $k - \Delta k/2 \leq q, p \leq k + \Delta k/2$ .  $\mu$  is the cosine of the angle with respect to the line of sight of the  $\mathbf{q}$  wave vector and  $L_\ell(\mu)$  is the Legendre polynomial of order  $\ell$ .  $\delta_D$  is the 3-D Dirac delta.  $N_p$  is the number of grid point pairs in the integration volume in Fourier space and can be computed as:

$$N_p(k) = \frac{V_k}{k_f^3} = k_f^{-3} \int_{V_p} \int_{V_q} d^3\mathbf{p} d^3\mathbf{q} \delta_D(\mathbf{q} + \mathbf{p}) \simeq \frac{4\pi k^2 \Delta k}{k_f^3}, \quad (4.18)$$

where  $V_k \simeq 4\pi k^2 \Delta k$  is the spherical integration shell defined by  $k - \Delta k/2 \leq q, p \leq k + \Delta k/2$  as defined in [Scoccimarro et al. \(1998\)](#).  $k_f$  is the fundamental frequency defined in terms of the survey volume  $V_e$  as  $k_f^3 = \frac{(2\pi)^3}{V_e}$ . We check that the estimator defined in Equation 4.17 is unbiased:

$$\begin{aligned}
 \langle \hat{P}_g^{(\ell)}(k) \rangle &= \left( \frac{2\ell+1}{2} \right) \frac{1}{(2\pi)^3 N_p(k)} \int_{V_p} \int_{V_q} d^3 \mathbf{p} d^3 \mathbf{q} L_\ell(\mu) \delta_D(\mathbf{q} + \mathbf{p}) \langle \delta_g^s(\mathbf{q}) \delta_g^s(\mathbf{p}) \rangle \\
 &= \left( \frac{2\ell+1}{2} \right) \frac{1}{(2\pi)^3 N_p(k)} \int_{V_p} \int_{V_q} d^3 \mathbf{p} d^3 \mathbf{q} L_\ell(\mu) \delta_D(\mathbf{q} + \mathbf{p})^2 (2\pi)^3 P_g^s(\mathbf{p}) \\
 &= \left( \frac{2\ell+1}{2} \right) \frac{1}{(2\pi)^3 N_p(k)} \int_{V_p} \int_{V_q} d^3 \mathbf{p} d^3 \mathbf{q} L_\ell(\mu) \delta_D(\mathbf{q} + \mathbf{p}) V_e P_g^s(\mathbf{p}) \\
 &= \left( \frac{2\ell+1}{2} \right) \frac{1}{V_e V_k} \int_{V_p} \int_{V_q} d^3 \mathbf{p} d^3 \mathbf{q} L_\ell(\mu) \delta_D(\mathbf{q} + \mathbf{p}) V_e P_g^s(\mathbf{p}) \\
 &= \left( \frac{2\ell+1}{2} \right) \frac{1}{V_k} \int_{V_p} \int_{V_q} d^3 \mathbf{p} d^3 \mathbf{q} L_\ell(\mu) \delta_D(\mathbf{q} + \mathbf{p}) P_g^s(\mathbf{p}) \\
 &\approx \left( \frac{2\ell+1}{2} \right) \int_{-1}^{+1} d\mu P_g^s(k, \mu) L_\ell(\mu), \tag{4.19}
 \end{aligned}$$

where we used the approximation made in [Joachimi et al. \(2009\)](#) that  $\delta_D^2 \approx \frac{V_e}{(2\pi)^3} \delta_D = k_f^{-3} \delta_D$ . In the last step it has been made the common approximation that  $\mathbf{p}$  and  $\mathbf{q}$  are very close to  $k$  in module for thin enough shells (small  $\Delta k$ ). The standard definition of the redshift galaxy power spectrum has also been used:

$$\langle \delta_g^s(\mathbf{q}) \delta_g^s(\mathbf{p}) \rangle = (2\pi)^3 \delta_D(\mathbf{q} + \mathbf{p}) P_g^s(\mathbf{p}). \tag{4.20}$$

The redshift space galaxy bispectrum is defined as:

$$\langle \delta_g^s(\mathbf{q}_1) \delta_g^s(\mathbf{q}_2) \delta_g^s(\mathbf{q}_3) \rangle = (2\pi)^3 \delta_D(\mathbf{q}_1 + \mathbf{q}_2 + \mathbf{q}_3) B_g^s(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3). \tag{4.21}$$

The analytical expression for the bispectrum monopole model was given in Equation 4.7.

Analogously to the power spectrum multipoles, the estimator for the bispectrum monopole can be defined as:

$$\hat{B}_g^{(0)}(k_1, k_2, k_3) = \frac{1}{4\pi} \frac{V_e (2\pi)^{-6}}{N_t(k_1, k_2, k_3)} \prod_{i=1}^3 \int_{V_{q_i}} d^3 \mathbf{q}_i \delta_D(\mathbf{q}_1 + \mathbf{q}_2 + \mathbf{q}_3) \delta_g^s(\mathbf{q}_1) \delta_g^s(\mathbf{q}_2) \delta_g^s(\mathbf{q}_3). \tag{4.22}$$

where  $N_t(k_1, k_2, k_3)$  is the number of independent grid points triplets inside the integration volume in Fourier space. As shown in the weak lensing 2D case by [Kayo, I. and Takada, M. and Jain, B. \(2013\)](#), this is computed as:

$$\begin{aligned}
 N_t(k_1, k_2, k_3) &= \frac{V_{k_{123}}}{k_f^6} = k_f^{-6} \int_{V_{q_1}} \int_{V_{q_2}} \int_{V_{q_3}} d^3 \mathbf{q}_1 d^3 \mathbf{q}_2 d^3 \mathbf{q}_3 \delta_D(\mathbf{q}_1 + \mathbf{q}_2 + \mathbf{q}_3) \\
 &\simeq \frac{8\pi^2 k_1 k_2 k_3 \Delta k_1 \Delta k_2 \Delta k_3}{k_f^6}. \tag{4.23}
 \end{aligned}$$

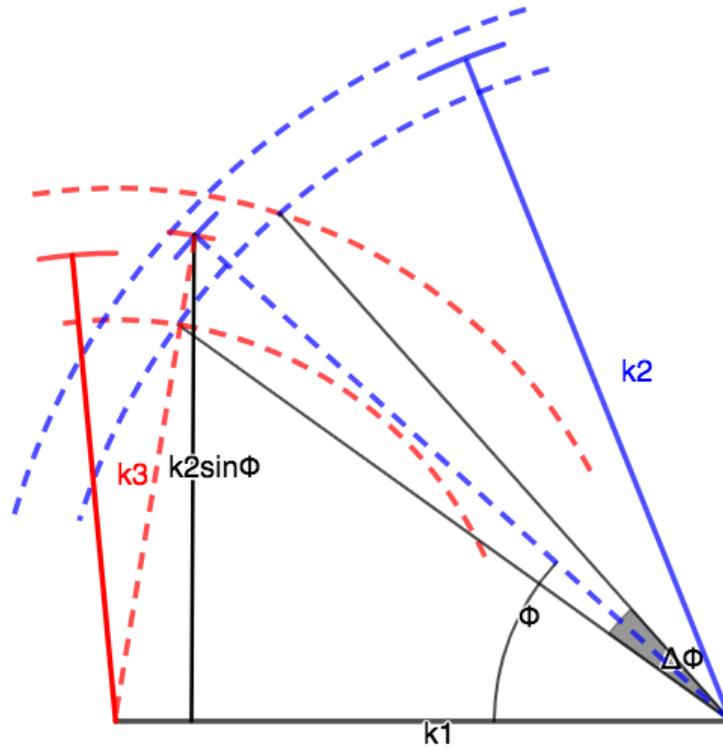


Figure 4.A.1: Computation of the integration volume in Fourier space in the case of the bispectrum monopole. Once the side  $k_1$  of the triangle is fixed, the other two sides are free to vary in the intersection given by two spheres of radius  $k_2 - \Delta k_2/2 \leq r_2 \leq k_2 + \Delta k_2/2$  and  $k_3 - \Delta k_3/2 \leq r_3 \leq k_3 + \Delta k_3/2$  respectively. In the Figure above the 2D projection of the annuli of thickness  $\Delta k_2$  (blue) and  $\Delta k_3$  (red) are shown. The angle  $\phi$  correspond to the angle  $\phi_{12}$  in the text.

It is important to notice that the result of the above integral must be symmetric in the  $k$ -vectors arguments. Therefore, the best way to derive the integral results is through geometrical considerations. Starting from  $q_1$ , this can be chosen in a spherical shell with volume  $V_{k_1} \simeq 4\pi k_1^2 \Delta k$ . Once  $q_1$  is fixed, considering the plane in which both  $q_2$  and  $q_3$  lie, they must connect to each other inside the 2D intersection formed by the two annuli defined by  $k_2 - \Delta k_2/2 \leq q_2 \leq k_2 + \Delta k_2/2$  and  $k_3 - \Delta k_3/2 \leq q_3 \leq k_3 + \Delta k_3/2$ . This has approximately an area equal to  $A_{k_{23}} \simeq k_2 \Delta \phi_{12} \Delta k_2$ . From Figure 4.A.1 it is possible to see that  $\Delta \phi_{12}$  is defined by varying  $k_3$  by  $\Delta k_3$ .  $\phi_{12}$  can be obtained from:

$$\cos \phi_{12} = \frac{k_1^2 + k_2^2 - k_3^2}{2k_1 k_2}, \quad (4.24)$$

and therefore  $\Delta \phi_{12}$  can be found differentiating with respect to  $k_3$ :

$$\frac{d \cos \phi_{12}}{dk_3} = -\frac{d\phi_{12}}{dk_3} \sin \phi_{12} = -\frac{k_3}{k_1 k_2} \implies \Delta \phi_{12} = \frac{\Delta k_3 k_3}{k_1 k_2} (\sin \phi_{12})^{-1} \quad (4.25)$$

Finally the volume of the intersection between  $k_2$  and  $k_3$  is obtained by rotating the area just found around the axis defined by  $k_1$ :

$$V_{k_{23}} = 2\pi A_{k_{23}} (k_2 \sin \phi_{12}), \quad (4.26)$$

which allows to compute  $V_{k_{123}} = V_{k_1} V_{k_{23}}$  in Equation 4.23.

## 4.A.2 Power spectrum monopole and quadrupole covariance matrix: Gaussian term

Following the Appendix of Gualdi et al. (2018b) we can check that also the bispectrum monopole estimator defined in Equation 4.22 is unbiased. Moreover it is possible to compute the Gaussian term of the covariance for the power spectrum monopole and quadrupole as follows:

$$\begin{aligned}
 C_G^{\text{P}_g^{(\ell)} \text{P}_g^{(\ell)}}(k_1; k_2) &= \\
 &= \left(\frac{2\ell+1}{2}\right)^2 \frac{(2\pi)^{-6}}{N_p(k_1) N_p(k_2)} \int_{V_{\mathbf{q}_1}} \int_{V_{\mathbf{q}_2}} \int_{V_{\mathbf{p}_1}} \int_{V_{\mathbf{p}_2}} d^3\mathbf{q}_1 d^3\mathbf{q}_2 d^3\mathbf{p}_1 d^3\mathbf{p}_2 L_\ell(\mu_1) L_\ell(\mu_2) \\
 &\times \delta_D(\mathbf{q}_1 + \mathbf{p}_1) \delta_D(\mathbf{q}_2 + \mathbf{p}_2) 2(2\pi)^6 \delta_D(\mathbf{q}_1 + \mathbf{q}_2) \delta_D(\mathbf{p}_1 + \mathbf{p}_2) \text{P}_g^s(\mathbf{q}_1) \text{P}_g^s(\mathbf{p}_2) \\
 &= \left(\frac{2\ell+1}{2}\right)^2 \frac{2}{N_p(k_1) N_p(k_2)} \int_{V_{\mathbf{q}_1}} \int_{V_{\mathbf{q}_2}} d^3\mathbf{q}_1 d^3\mathbf{q}_2 L_\ell(\mu_1) L_\ell(\mu_2) \\
 &\times \delta_D(\mathbf{q}_1 + \mathbf{q}_2)^2 \text{P}_g^s(\mathbf{q}_1) \text{P}_g^s(\mathbf{q}_2) \\
 &= \left(\frac{2\ell+1}{2}\right)^2 \frac{2k_f^{-3}}{N_p(k_1) N_p(k_2)} \int_{V_{\mathbf{q}_1}} \int_{V_{\mathbf{q}_2}} d^3\mathbf{q}_1 d^3\mathbf{q}_2 L_\ell(\mu_1) L_\ell(\mu_2) \\
 &\times \delta_D(\mathbf{q}_1 + \mathbf{q}_2) \text{P}_g^s(\mathbf{q}_1) \text{P}_g^s(\mathbf{q}_2) \\
 &\approx \left(\frac{2\ell+1}{2}\right)^2 \frac{2k_f^{-3}}{N_p(k_1) N_p(k_2)} \text{P}_g^{(\ell)}(k_1) \text{P}_g^{(\ell)}(k_2) \int_{V_{\mathbf{q}_1}} \int_{V_{\mathbf{q}_2}} d^3\mathbf{q}_1 d^3\mathbf{q}_2 \delta_D(\mathbf{q}_1 + \mathbf{q}_2) \\
 &= \left(\frac{2\ell+1}{2}\right)^2 \frac{2\delta_{12}^K}{N_p(k_1)} \text{P}_g^{(\ell)}(k_1)^2, \tag{4.27}
 \end{aligned}$$

where again we used the approximation made in [Joachimi et al. \(2009\)](#) that  $\delta_D^2 \approx \frac{V_e}{(2\pi)^3} \delta_D = k_f^{-3} \delta_D$ .  $\delta_{12}^K$  is the Kronecker delta indicating that the vector  $\mathbf{q}_1$  and  $\mathbf{q}_2$  are identical (in the second step trivial  $\delta_K$  have been omitted in order to avoid making the notation heavier by adding also the wave-vector letter). In the last steps we made the approximation that the power spectrum monopole and quadrupoles do not vary significantly when integrated over the the bin in Fourier space.

### 4.A.3 Bispectrum monopole covariance matrix: Gaussian term

Analogously to the above we now compute the diagonal term of the bispectrum monopole covariance matrix:

$$\begin{aligned}
 C_G^{\text{B}_g^0 \text{B}_g^0}(k_1, k_2, k_3; k_4, k_5, k_6) &= \\
 &= \frac{1}{16\pi^2} \frac{(2\pi k_f)^{-6}}{N_t(k_1, k_2, k_3) N_t(k_4, k_5, k_6)} \prod_{i=1}^6 \int_{V_{\mathbf{q}_i}} d^3 \mathbf{q}_i \delta_D(\mathbf{q}_1 + \mathbf{q}_2 + \mathbf{q}_3) \delta_D(\mathbf{q}_4 + \mathbf{q}_5 + \mathbf{q}_6) \\
 &\times (2\pi)^9 \delta_D(\mathbf{q}_1 + \mathbf{q}_4) \delta_D(\mathbf{q}_2 + \mathbf{q}_5) \delta_D(\mathbf{q}_3 + \mathbf{q}_6) P_g^s(\mathbf{q}_1) P_g^s(\mathbf{p}_2) P_g^s(\mathbf{q}_3) + 5 \text{ perm.} \\
 &= \frac{D_{123456}}{16\pi^2} \frac{(2\pi)^3 k_f^{-6}}{N_t(k_1, k_2, k_3)^2} \prod_{i=1}^3 \int_{V_{\mathbf{q}_i}} d^3 \mathbf{q}_i \delta_D(\mathbf{q}_1 + \mathbf{q}_2 + \mathbf{q}_3)^2 P_g^s(\mathbf{q}_1) P_g^s(\mathbf{p}_2) P_g^s(\mathbf{q}_3) \\
 &= \frac{D_{123456}}{16\pi^2} \frac{V_e k_f^{-6}}{N_t(k_1, k_2, k_3)^2} \prod_{i=1}^3 \int_{V_{\mathbf{q}_i}} d^3 \mathbf{q}_i \delta_D(\mathbf{q}_1 + \mathbf{q}_2 + \mathbf{q}_3) P_g^s(\mathbf{q}_1) P_g^s(\mathbf{p}_2) P_g^s(\mathbf{q}_3) \\
 &\approx \frac{D_{123456}}{16\pi^2} \frac{V_e k_f^{-6}}{N_t(k_1, k_2, k_3)^2} P_g^{(0)}(k_1) P_g^{(0)}(k_2) P_g^{(0)}(k_3) \prod_{i=1}^3 \int_{V_{\mathbf{q}_i}} d^3 \mathbf{q}_i \delta_D(\mathbf{q}_1 + \mathbf{q}_2 + \mathbf{q}_3) \\
 &= \frac{D_{123456}}{16\pi^2} \frac{V_e}{N_t(k_1, k_2, k_3)} P_g^{(0)}(k_1) P_g^{(0)}(k_2) P_g^{(0)}(k_3), \tag{4.28}
 \end{aligned}$$

where  $D_{123456}$  stands for all the possible permutations and has values 6, 2, 1 respectively for equilateral, isosceles and scalene triangles. Again it has been assumed that the power spectrum monopole does not vary significantly inside the integration volume.

#### 4.A.4 Bispectrum monopole covariance matrix: non-Gaussian term

In this work we use only one of the non-Gaussian terms of the bispectrum monopole covariance matrix. This is because we just need to model the covariance matrix analytically in order to derive the weights for the compression. This additional term allows to better capture the correlation between different triangles. We leave to future work the analytic computation of the remaining terms (one proportional to the product between trispectrum and power spectrum and the other given by the pentaspectrum).

$$\begin{aligned}
 C_{\text{NG}}^{\text{B}_g^0 \text{B}_g^0}(k_1, k_2, k_3; k_4, k_5, k_6) &= \\
 &= \frac{1}{16\pi^2} \frac{(2\pi k_f)^{-6}}{N_t(k_1, k_2, k_3) N_t(k_4, k_5, k_6)} \prod_{i=1}^6 \int_{V_{\mathbf{q}_i}} d^3 \mathbf{q}_i \delta_D(\mathbf{q}_1 + \mathbf{q}_2 + \mathbf{q}_3) \delta_D(\mathbf{q}_4 + \mathbf{q}_5 + \mathbf{q}_6) \\
 &\times (2\pi)^6 \delta_D(\mathbf{q}_1 + \mathbf{q}_2 + \mathbf{q}_4) \delta_D(\mathbf{q}_3 + \mathbf{q}_5 + \mathbf{q}_6) \text{B}_g^s(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_4) \text{B}_g^s(\mathbf{q}_3, \mathbf{q}_5, \mathbf{q}_6) + 8 \text{ perm.} \\
 &= \frac{1}{16\pi^2} \frac{k_f^{-6} \delta_{34}^{\text{K}}}{N_t(k_1, k_2, k_3) N_t(k_3, k_5, k_6)} \int_{V_{\mathbf{q}_1}} \int_{V_{\mathbf{q}_2}} \int_{V_{\mathbf{q}_3}} \int_{V_{\mathbf{q}_5}} \int_{V_{\mathbf{q}_6}} d^3 \mathbf{q}_1 d^3 \mathbf{q}_2 d^3 \mathbf{q}_3 d^3 \mathbf{q}_5 d^3 \mathbf{q}_6 \\
 &\times \delta_D(\mathbf{q}_1 + \mathbf{q}_2 + \mathbf{q}_3) \delta_D(\mathbf{q}_3 + \mathbf{q}_5 + \mathbf{q}_6)^2 \text{B}_g^s(\mathbf{q}_1, \mathbf{q}_2, -\mathbf{q}_3) \text{B}_g^s(\mathbf{q}_3, \mathbf{q}_5, \mathbf{q}_6) + 8 \text{ perm.} \\
 &= \frac{1}{16\pi^2} \frac{k_f^{-9} \delta_{34}^{\text{K}}}{N_t(k_1, k_2, k_3) N_t(k_3, k_5, k_6)} \int_{V_{\mathbf{q}_1}} \int_{V_{\mathbf{q}_2}} \int_{V_{\mathbf{q}_3}} \int_{V_{\mathbf{q}_5}} \int_{V_{\mathbf{q}_6}} d^3 \mathbf{q}_1 d^3 \mathbf{q}_2 d^3 \mathbf{q}_3 d^3 \mathbf{q}_5 d^3 \mathbf{q}_6 \\
 &\times \delta_D(\mathbf{q}_1 + \mathbf{q}_2 + \mathbf{q}_3) \delta_D(\mathbf{q}_3 + \mathbf{q}_5 + \mathbf{q}_6) \text{B}_g^s(\mathbf{q}_1, \mathbf{q}_2, -\mathbf{q}_3) \text{B}_g^s(\mathbf{q}_3, \mathbf{q}_5, \mathbf{q}_6) + 8 \text{ perm.} \\
 &\approx \frac{1}{16\pi^2} \frac{k_f^{-3} \delta_{34}^{\text{K}}}{N_t(k_3, k_5, k_6)} \text{B}_g^{(0)}(k_1, k_2, k_3) \text{B}_g^{(0)}(k_3, k_5, k_6) \\
 &\times \int_{V_{\mathbf{q}_i}} d^3 \mathbf{q}_5 d^3 \mathbf{q}_6 \delta_D(\mathbf{q}_3 + \mathbf{q}_5 + \mathbf{q}_6) + 8 \text{ perm.} \\
 &= \frac{\delta_{34}^{\text{K}}}{16\pi^2} \frac{k_f^3}{4\pi k_3^2 \Delta k_3} \text{B}_g^{(0)}(k_1, k_2, k_3) \text{B}_g^{(0)}(k_3, k_5, k_6) + 8 \text{ perm.}, \tag{4.29}
 \end{aligned}$$

where the usual approximations have been used together with Equation 4.26 which in the last step has been used to simplify the integration over the volume in Fourier space once one of the  $\mathbf{k}$ -vectors is fixed.

#### 4.A.5 Cross-covariance term

For what concerns the cross-covariance term between power spectrum (monopole and quadrupole) and bispectrum monopole we use only the first leading term in our model:

$$\begin{aligned}
 C^{\text{P}_g^{(\ell)} \text{B}_g^0}(k_1; k_2, k_3, k_4) &= \\
 &= \frac{1}{4\pi} \left( \frac{2\ell + 1}{2} \right) \frac{(2\pi)^{-6} k_f^{-3}}{N_p(k_1) N_t(k_2, k_3, k_4)} \int_{V_{\mathbf{q}_1}} \int_{V_{\mathbf{p}_1}} d^3 \mathbf{q}_1 d^3 \mathbf{p}_1 \prod_{i=2}^4 \int_{V_{\mathbf{q}_i}} d^3 \mathbf{q}_i \delta_D(\mathbf{q}_1 + \mathbf{p}_1) \\
 &\times \delta_D(\mathbf{q}_2 + \mathbf{q}_3 + \mathbf{q}_4) L_\ell(\mu_1) 2(2\pi)^6 \delta_D(\mathbf{q}_1 + \mathbf{q}_2) \delta_D(\mathbf{p}_1 + \mathbf{q}_3 + \mathbf{q}_4) \\
 &\times \text{P}_g^s(\mathbf{q}_2) \text{B}_g^s(\mathbf{q}_2, \mathbf{q}_3, \mathbf{q}_4) + 2 \text{ perm.} \\
 &= \frac{1}{2\pi} \left( \frac{2\ell + 1}{2} \right) \frac{k_f^{-3}}{N_p(k_1) N_t(k_2, k_3, k_4)} \prod_{i=1}^4 \int_{V_{\mathbf{q}_i}} d^3 \mathbf{q}_i L_\ell(\mu_1) \\
 &\times \delta_D(\mathbf{q}_1 + \mathbf{q}_2) \delta_D(\mathbf{q}_2 + \mathbf{q}_3 + \mathbf{q}_4)^2 \text{P}_g^s(\mathbf{q}_2) \text{B}_g^s(\mathbf{q}_2, \mathbf{q}_3, \mathbf{q}_4) + 2 \text{ perm.} \\
 &= \frac{1}{2\pi} \left( \frac{2\ell + 1}{2} \right) \frac{k_f^{-6} \delta_{12}^{\text{K}}}{N_p(k_2) N_t(k_2, k_3, k_4)} \prod_{i=2}^4 \int_{V_{\mathbf{q}_i}} d^3 \mathbf{q}_i L_\ell(\mu_2) \delta_D(\mathbf{q}_2 + \mathbf{q}_3 + \mathbf{q}_4) \\
 &\times \text{P}_g^s(\mathbf{q}_2) \text{B}_g^s(\mathbf{q}_2, \mathbf{q}_3, \mathbf{q}_4) + 2 \text{ perm.} \\
 &\approx \frac{1}{2\pi} \left( \frac{2\ell + 1}{2} \right) \frac{\delta_{12}^{\text{K}}}{N_p(k_2)} \text{P}_g^{(\ell)}(k_2) \text{B}_g^{(0)}(k_2, k_3, k_4) + 2 \text{ perm.}, \tag{4.30}
 \end{aligned}$$

where once more we have used the same approximation of the power spectrum multipoles and bispectrum monopole not varying significantly inside the integration volume.

## 4.B Validation tests

In Table 4.B.1 we report the results obtained compressing the bispectrum with respect to the shifted fiducial parameter sets. This is to test whether the performance of the compression is affected by the choice of fiducial set of parameter values. In particular, we consider two cases by varying the fiducial cosmology by adding/subtracting  $1\sigma$  1D credible intervals (derived from the MCMC) to all the parameters. The table quantifies that the shifts in the means of the 1D posterior distributions produced by considering a non-optimal fiducial cosmology are small compared to the  $1\sigma$  1D credible intervals of the MCMC results.

In Figures 4.B.1 and 4.B.2 the 1 and 2-D posterior distributions obtained via MCMC/MC-MP/G-PCA for the test cases relative to the  $\Delta k_6$  and  $\Delta k_5$  binning cases are shown. MC-MP recovers with very good approximation the 1 and 2-D posterior distributions derived by the MCMC. G-PCA shows a slightly greater loss of information for the  $\Delta k_6$  case. However this is noticeably closer to the MCMC/MC-MP result when the number of triangles used is increased ( $\Delta k_5$  case).

In Figure 4.B.3 we compare the best-fit model obtained by varying four parameters ( $b_1, b_2, f, \sigma_8$ ) with the best-fit model corresponding to a fit done via standard MCMC sampling with only three parameters varied, ( $b_1, b_2, f$ ), with  $\sigma_8 = \sigma_8^{\text{fid}}$ . For the three parameter case we find running the MCMC:  $b_1 = 1.98 \pm 0.01$ ,  $b_2 = 0.39 \pm 0.06$ ,  $f(z_{\text{CMASS}}) = 0.52 \pm 0.03$  with  $\sigma_8^{\text{fid}}(z_{\text{CMASS}}) = 0.61$ .

Thereby we show that the discrepancy between the results of this paper and the ones presented in the BOSS collaboration analysis Gil-Marín (2017) is only due to the different range of scales considered. Indeed, by limiting our analysis to a smaller range of scales in  $k$ -space, the degeneracy between the amplitude-like parameters  $b_1$  and  $\sigma_8$  is much stronger. That is visible in Figure 4.B.3, where the models given by sets of parameters with very different  $b_1$ ,  $b_2$  and  $\sigma_8$  parameters produce very similar predictions of the signals all with good  $\chi_{\text{red.}}^2$  and  $p$ -values.

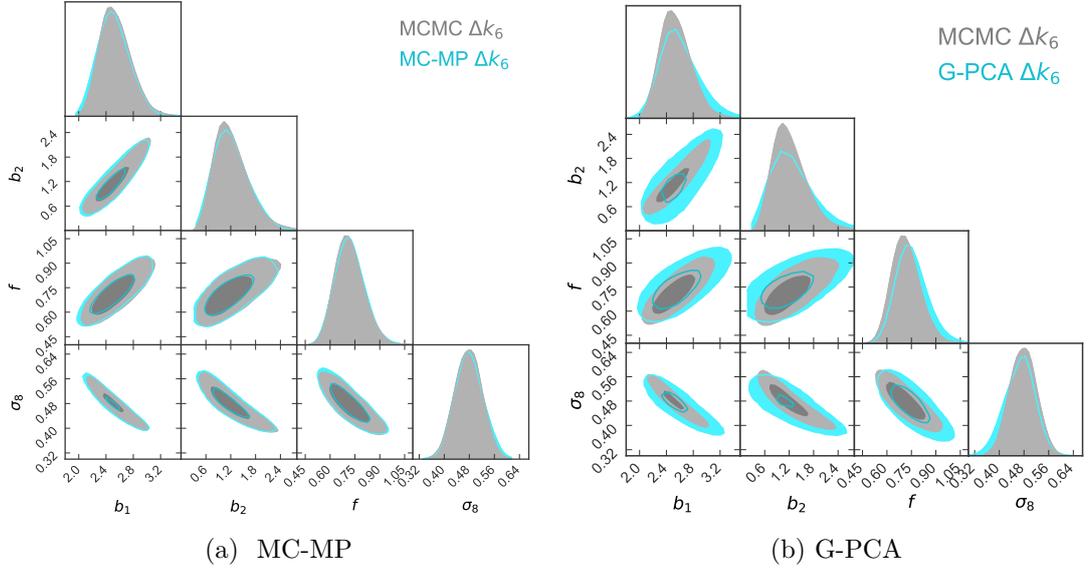


Figure 4.B.1: Joint data-vector  $[P_g^{(0)}, P_g^{(2)}, B_g^{(0)}]$  posteriors: MC-MP and G-PCA four-parameter  $\Delta k_6$  case.

**a)** 2-D 68% and 95% credible regions are shown in order to compare the MC-MP (cyan) performance to the one of the standard MCMC (grey) for the full data vector. The difference between MC-MP and MCMC contours is quantified in Table 4.1.

**b)** The same as a) but for the G-PCA method.

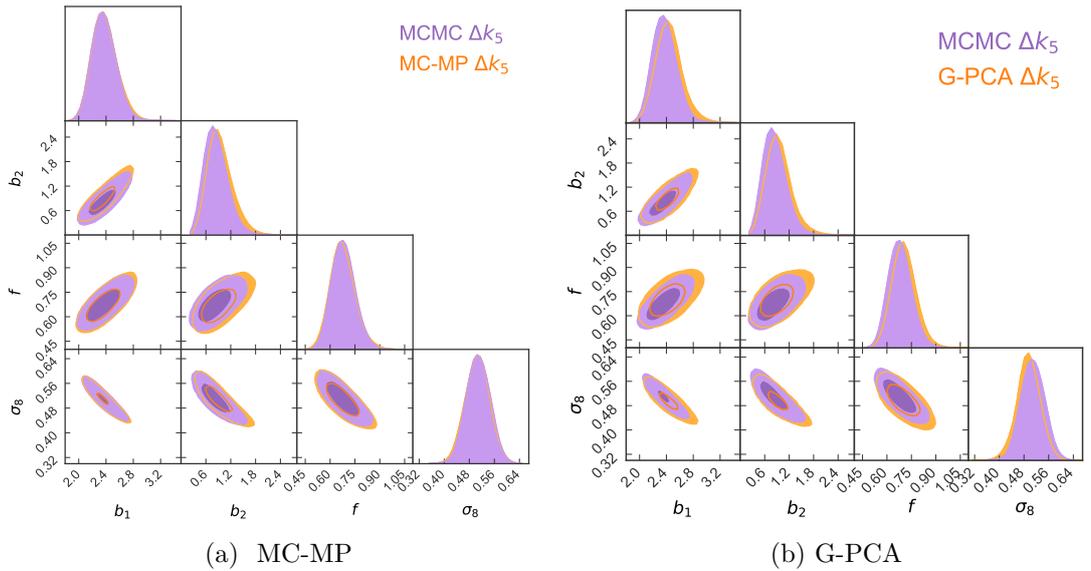


Figure 4.B.2: Joint data-vector  $[P_g^{(0)}, P_g^{(2)}, B_g^{(0)}]$  posteriors: MC-MP and G-PCA four-parameter  $\Delta k_5$  case.

Both **a)** and **b)** are the same as for Figure 4.B.1 for the  $\Delta k_5$  case.

Table 4.B.1: Four parameter-case, checking consistency for shifted fiducial cosmology.

**Upper half:** Mean values of the posterior distributions and 68% credible intervals for the MCMC and the MC-MP / G-PCA compression methods. We report the values for the  $\Delta k_6$  binning case for both compression methods in three cases consisting in using for the compression: the fiducial cosmology, the fiducial cosmology shifted by  $+1\sigma$  and the fiducial cosmology shifted by  $-1\sigma$ .

**Lower half:** In the compression columns we report the relative difference between the posterior modes obtained via MCMC and the ones obtained via compression (MC-MP or G-PCA). In the MCMC columns the relative size of the 68% credible intervals obtained via MCMC sampling is shown. By comparing the MCMC columns to the compression ones, it is clear that the difference between the mean parameter values obtained via MCMC and the ones obtained via compression (MC-MP or G-PCA) are evidently within the 68% credible intervals given by the MCMC on the full data-vector.

	$\Delta k_6$			$\Delta k_6 + 1\sigma$		$\Delta k_6 - 1\sigma$	
	MCMC	MC-MP	G-PCA	MC-MP	G-PCA	MC-MP	G-PCA
$b_1$	$2.41 \pm 0.22$	$2.41 \pm 0.23$	$2.49 \pm 0.27$	$2.47 \pm 0.23$	$2.41 \pm 0.12$	$2.54 \pm 0.24$	$2.34 \pm 0.37$
$b_2$	$1.00 \pm 0.40$	$1.04 \pm 0.42$	$1.08 \pm 0.47$	$1.04 \pm 0.40$	$1.29 \pm 0.25$	$1.03 \pm 0.44$	$0.93 \pm 0.67$
$f$	$0.69 \pm 0.08$	$0.72 \pm 0.09$	$0.72 \pm 0.09$	$0.70 \pm 0.08$	$0.69 \pm 0.05$	$0.72 \pm 0.09$	$0.68 \pm 0.12$
$\sigma_8$	$0.50 \pm 0.04$	$0.48 \pm 0.05$	$0.48 \pm 0.05$	$0.49 \pm 0.04$	$0.49 \pm 0.03$	$0.46 \pm 0.05$	$0.50 \pm 0.07$
	$\frac{\Delta\theta^{\text{mc}}}{\theta^{\text{mc}}}$ [%]	$\frac{\theta^{\text{comp.}} - \theta^{\text{mc}}}{\theta^{\text{mc}}}$ [%]		$\frac{\theta^{\text{comp.}} - \theta^{\text{mc}}}{\theta^{\text{mc}}}$ [%]		$\frac{\theta^{\text{comp.}} - \theta^{\text{mc}}}{\theta^{\text{mc}}}$ [%]	
$b_1$	9.2	-0.3	3.3	2.15	-0.26	8.57	0.31
$b_2$	40.3	3.5	7.5	3.47	28.68	25.29	13.26
$f$	12.1	4.4	4.4	0.84	0.51	6.96	0.26
$\sigma_8$	8.5	-5.1	-5.5	-3.25	-2.91	-8.94	-1.39

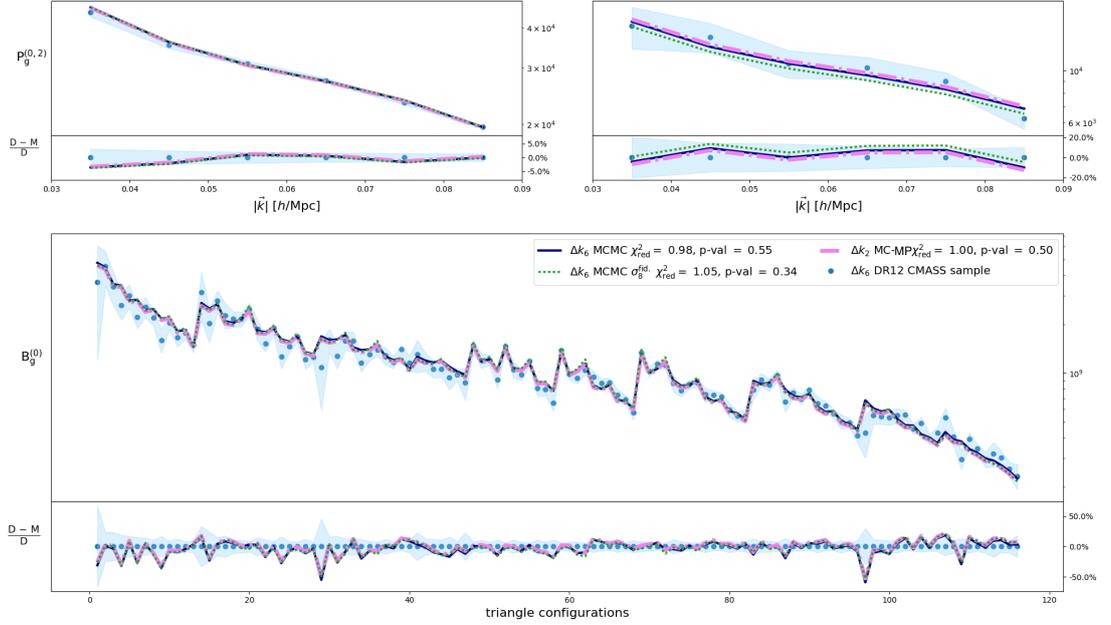


Figure 4.B.3: Reduced  $\chi^2$  and  $p$ -values for the best-fit parameters obtained using the MCMC/MC-MP methods with varying  $\sigma_8$  and for the MCMC leaving  $\sigma_8 = \sigma_8^{\text{fid}}$  fixed. The  $k$ -binnings shown for the four parameter case ( $b_1, b_2, f, \sigma_8$ ) are respectively the standard  $\Delta k_6$  (navy) for the MCMC and the  $\Delta k_2$  (pink) for the MC+MP. The line corresponding to the fit obtained by letting free to vary only the parameters ( $b_1, b_2, f$ ) is shown in green. The two upper panels are for the power spectrum monopole (left) and quadrupole (right) while the bottom panel refers to the bispectrum monopole. The lower part of each panel shows the relative difference between the data measurements and the different models. Even if for example  $b_1$  and  $\sigma_8$  values are shifted in the cases of  $\Delta k_6$  and  $\Delta k_2$ , this is due to the strong degeneracy between them and both models are practically identical to the one given by the three parameters fit ( $b_2, b_2, f$ ) with  $\sigma_8 = \sigma_8^{\text{fid}}$ . The only way to converge to the results obtained by the BOSS collaboration is to consider a larger range of scales (as they have done) for both power spectrum and bispectrum which however involves a more complex modelling of the data-vector.

*“Don’t adventures ever have an end? I suppose not.  
Someone else always has to carry on the story.”*

- J.R.R. Tolkien, *The Fellowship of the Ring*

## 5 | Conclusions and future work

In this thesis we have studied different techniques to compress 3pt statistics in order to efficiently extract cosmological information once applied to data sets from current and future surveys. In particular, we considered two 3pt statistics: the 3pt auto-correlation function (and its ratio with the square of the variance, the skewness), and the redshift space galaxy bispectrum. Below are summarised the main results of each chapter included in this thesis, together with the possible extensions that could be investigated in the future. Finally, an alternative method to compress 3pt statistics is introduced.

### 5.1 Chapter 2: skewness and 3pt auto-correlation function

In the first part of Chapter 2, we have shown the analytic derivation of the 3pt auto-correlation function and of the skewness quantity. Primordial non-Gaussianity contributions were also taken into consideration by computing the relative additional terms up to linear order in  $f_{NL}$ . The main original results of this part are given by the  $f_{NL}$  terms in Equation 2.68 and Equations 2.76, 2.77. The magnitude of their contribution to the standard Gaussian terms is shown in Figure 2.1.

In the second part, we described the measurement of the 2pt and 3pt auto-correlation functions on DR11 SDSS BOSS CMASS NGC sample. The motivating idea was to find out whether it was possible to detect the BAO feature in the skewness. Unfortunately, compared to the cosmic variance contribution, this statistic signal was too low to observe the BAO feature which is visible in the theoretical prediction. Therefore, since the skewness or, more simply, the combination of 2pt and 3pt auto-correlation functions allows to lift the degeneracy between the linear galaxy bias and the "amplitude-like" cosmological parameters (for example  $\sigma_8$  and  $f_{NL}$ ), we focused on constraining the parameter set  $(b_1, \sigma_8, f_{NL})$ . The 1 and 2D posterior distributions derived via MCMC sampling are shown in Figure 2.4.

Auto-correlation functions represent an alternative to the standard correlation functions used in Cosmology. The advantage of these integrated quantities can be seen in the 3pt case. Where the 3pt correlation function requires three different coordinates for each data vector element (e.g. the three sides of a triangle), the skewness needs only one (the radius of the window function). This allows one to compress the data-vector dimension, which can then be used for parameter inference together with far fewer simulations (needed to compute the covariance matrix) than the ones needed for a much longer data-vector (as in the 3pt correlation function case).

However, integration of the statistics also implies the loss of at least part of the information and hence reduces the performance in term of parameter constraints.

### 5.1.1 Chapter 2: future work

In the near future, DESI will produce a huge data set including the redshift and angular position of more than 30 million objects between galaxies and quasars. DESI will map their positions in a volume more than 10 times bigger than that covered by BOSS. It will then be interesting to use the 2pt and 3pt auto-correlation functions as complementary probes to the standard correlation functions. Auto-correlation function can also to provide consistency checks for the standard correlation functions results. With a much larger volume and far more objects, the limit imposed by cosmic variance should be pushed towards larger scales, while the signal to noise ratio will also increase at small scales. The measurement of the 2pt and, in particular, the 3pt auto-correlation function on DESI data could for example produce interesting results on primordial non-Gaussianity, constraining the  $f_{NL}$  parameter at late times, perhaps, at the level of CMB measurements.

For that purpose, the current pipeline for the measurement of 2pt and 3pt auto-correlation functions will need to be improved in order to take care of the predicted systematic errors such as finite survey volume.

## 5.2 Chapter 3: maximal compression of the redshift space galaxy power spectrum and bispectrum

In Chapter 3 we presented the work published in [Gualdi et al. \(2018b\)](#) regarding the compression of the redshift space galaxy power spectrum and bispectrum. The two methods presented are based on the MOPED algorithm. The first

method consists of running a MCMC sampler with the compressed data vector (MCMC-MP). In the second method, before the compression, there is an orthogonalisation of the parameter space step, using a PCA transformation (PCA + MP). In this way it is possible to randomly sample the 1D posterior distributions of the orthogonalised parameters in order to reconstruct the multidimensional posterior distribution of the physical parameters.

In order to apply both methods we derived an expression for the covariance matrix of the joint data-vector  $[P_g^s, B_g^s]$ . The expressions describing all the different terms of the covariance matrix are given in Equations 3.8, 3.10 and 3.12.

We proved that both compression methods recover with minimal loss of information the 1 and 2D posterior distributions obtained via MCMC sampling using the uncompressed data-vector (Figures 3.3 and 3.6). Moreover we forecast the added value, in terms of tighter constraints and lifted degeneracies between parameters, given by including the bispectrum in the analysis with respect to just using the power spectrum.

### 5.2.1 Chapter 3: future work

In order to fully exploit the information contained in the spectroscopic data sets that surveys like DESI and PFS will produce in the near future, the use of higher-order statistics will be indispensable. The extremely large dimension that the data-vectors of higher-order statistics can easily reach, is the main obstacle limiting their employment. Therefore, the compression methods presented here have the potential to solve the problem. Future work would then consist of extending our modelling of the covariance matrix in order to be able to better describe the data-vector self-correlation properties also at non-linear scales. Indeed, at those scales, loop corrections and higher-order terms become non-negligible in the perturbation theory derivation of the covariance matrix expressions.

An immediate extension would also be the development of the formalism necessary to apply the compression to 3pt correlation function in real space.

Galaxy clustering is just one of the several probes in which correlation functions and their Fourier transforms are used to constrain cosmological parameters. In the present and even more in the near future, large data sets will also be available for weak lensing studies (DES, LSST, Euclid, HSC). Therefore both compression methods could be applied to 3pt statics for the convergence and shear fields if the necessary analytical formalism is developed.

The bispectrum of 21 cm emission lines or the Ly- $\alpha$  forest also represent promising applications of the compression methods.

## 5.3 Chapter 4: enhancing BOSS bispectrum cosmological constraints with maximal compression

Chapter 4 is based on the application of the compression methods presented in Chapter 3 (MCMC + MP relabelled as MC-MP) on the power spectrum monopole, quadrupole and the bispectrum monopole measurements from the DR12 BOSS CMASS NGC and SGC data samples (Gil-Marín, 2017).

The idea behind this project was to test the effect of using, thanks to the compression, many more triangle configurations in the bispectrum part of the data-vector. We showed that more triangle configurations improve the overall parameter constraints (tighter 1 and 2D posterior distributions, precise numbers in Table 4.2). From Figure 4.5 one can see that tighter posterior distributions also imply their shape being much closer to a multivariate Gaussian function's one.

In order to apply the compression methods, it was necessary to derive analytically new expressions for the covariance matrix of the  $[P_g^{(0)}, P_g^{(2)}, B_g^{(0)}]$  joint data-vector. These expressions are given in Eqs. 4.5, 4.8, 4.9 and 4.10.

As described in Chapter 3, the PCA + MP method suffered limitations linked to strongly degenerate parameter spaces. For this reason it was added a pre-Gaussianisation step (Schuhmann et al., 2016) before the PCA transformation. The method was hence relabelled G-PCA.

Another important aspect of both MC-MP and G-PCA compression methods is that an extension for higher-order statistics such as the trispectrum and tetraspectrum, is straightforward. It is only necessary to derive the corresponding analytic expressions for the relative covariance matrices. This would enable the use of these higher-order statistics, which up to now has been considered to be prohibitive due to the very large dimension of the data-vectors describing them.

### 5.3.1 Chapter 4: future work

In the analysis presented in Chapter 4, we used a  $k$ -range more conservative than the one adopted in the BOSS analysis (Gil-Marín, 2017). In order to extend the analysis to non-linear scales it is necessary to improve the model of the data-vector, in particular by adding loop-corrections to the power spectrum monopole and quadrupole.

By evaluating the likelihood using a covariance matrix estimate from the galaxy mocks, both compression methods have been shown to work and have

improved constraints obtain through the standard MCMC. We then wish to apply this analysis pipeline to future data-sets (e.g. DESI).

As discussed in Section 5.2.1, an immediate extension is the application of the MC-MP and G-PCA method to higher-order statistics measurements for different cosmological observables (e.g. weak-lensing, 21 cm emission lines, Ly- $\alpha$  forest).

## 5.4 Geometrical compression

The alternative compression method for 3pt statistics, labelled *Geometrical Compression* (MC-GC), is based on exploiting the sensitivity of the geometrical properties of 3pt statistics data-vector elements (functions of triangle characteristics). In other words, the data-vector element coordinates (for example the modulus of the three sides of a triangle) are converted into new parameters characterising the triangle configuration in question (for example the area of the triangle).

As in the MC-MP case, we run an MCMC sampling on the compressed data-vector to derive the 1 and 2D posterior distributions of the model parameters taken into consideration.

We consider the particular case of the bispectrum in order to compare the MC-GC results to the methods presented in Chapter 3 and 4.

### 5.4.1 New triangles parametrisation

We want to regroup the bispectrum data-vector elements in bins defined by different parameters describing the triangle configurations. The idea underlying this procedure is that similar triangular shapes will result in similar bispectrum values. This is because the perturbation kernels depend in particular on the cosine of the angles between the sides of the triangle. The hypothesis is also that the new bins will be less correlated between each other, since similar triangular configurations will contribute to the same new data-vector element.

Given the three triangle sides  $(k_1, k_2, k_3)$  normally characterising an element of the redshift space galaxy bispectrum monopole data-vector, we define three new variables. The first is the square root of the area of the triangle, which we label  $\aleph$  ("aleph"). It can be computed using Heron's formula:

$$A = \sqrt{s(s - k_1)(s - k_2)(s - k_3)} \implies \aleph = \sqrt{A}, \quad (5.1)$$

where  $s = \frac{1}{2}(k_1 + k_2 + k_3)$  is the semi-perimeter of the triangle. The  $\aleph$  parameter keeps track of the scales probed by the triangle configuration.

The second variable which we use to characterise a triangle is the cosine of the largest angle,  $\daleth = \cos \theta_{\max}$  (pronounced "daleth"). This choice allows one to describe whether the triangle is acute or obtuse. If  $\cos(\pi/3) = 1/2 < \daleth < 1$  the triangle is acute. In this case either the three sides are all approximately the same or two of them are larger than a third one. If  $-1 < \daleth < 0$  the triangle is obtuse. The triangle could then have either a side much larger than the other two (the one opposite to  $\theta_{\max}$ ) or two sides of similar length with a third smaller one. In order to distinguish between the pairs of possibilities above described, as a third variable we consider the ratio between the cosines of the intermediate and smallest angles,  $\beth = \cos \theta_{\text{int}} / \cos \theta_{\min}$  (pronounced "gimel"). All the cosines can be computed using the cosine rule for a triangle

$$k_l^2 = k_m^2 + k_n^2 - 2k_mk_n \cos \theta_{mn}. \quad (5.2)$$

Therefore each triangle configuration can be described as a function of the three variables  $(\aleph, \daleth, \beth)$  and the same is true for each bispectrum monopole data-vector element

$$B_g^{(0)}(k_1, k_2, k_3) \implies B_g^{(0)}(\aleph, \daleth, \beth). \quad (5.3)$$

Once the coordinate conversion has been done for all the triangle configurations, the binning for the new coordinates can be defined by finding the minimum and maximum values for the new parameters  $(\aleph, \daleth, \beth)$ . Given a choice for the number of bins for each new coordinate  $(n_\aleph, n_\daleth, n_\beth)$ , the potential dimension of the new data vector is  $n_\aleph \times n_\daleth \times n_\beth$ . However, as is the case when using the three sides  $(k_1, k_2, k_3)$  to describe the triangle, several combinations of  $(\aleph, \daleth, \beth)$  actually do not satisfy the triangle inequalities, and therefore no triplet  $(k_1, k_2, k_3)$  will contribute to that particular bin. Moreover even if a particular combination of  $(\aleph, \daleth, \beth)_i$  does represent a triangle configuration, it is not certain that the triangles bin defined by  $(\aleph, \daleth, \beth)_i$  will contain modes since the original number of triangles in  $(k_1, k_2, k_3)$  coordinates was finite. The new data vector  $\mathbf{g}$  is obtained by averaging over all the bispectra in the non-empty triangle sets defined by different combinations of the coordinates  $(\aleph, \daleth, \beth)$ :

$$g_i(\aleph_i, \daleth_i, \beth_i) = \frac{1}{N_i^{\text{tr.}}} \sum_{j: (k_1^j, k_2^j, k_3^j) \in (\aleph_i, \daleth_i, \beth_i)}^{N_i^{\text{tr.}}} B_g^{(0)}(k_1^j, k_2^j, k_3^j) \quad (5.4)$$

where each new data vector element has been normalised by dividing by the number of triangles belonging to the same set defined by a particular combination

of  $(\aleph, \beth, \beth)$ .  $N_i^{\text{tr.}}$  is the number of triangle configurations belonging to the original data-vector that fall into the new bin defined by the  $i$ -th combination of the new parameters  $(\aleph, \beth, \beth)$ .

### 5.4.2 Number of bins: optimal choice

For the construction of the new data-vector it is necessary to define how many bins will be used to divide the range of each parameter. In order to optimise the choice of these three numbers,  $(n_{\aleph}, n_{\beth}, n_{\beth})$  we suggest the following procedure. The idea is to "sample" the sensitivity of the new data-vector to the considered model parameters for the different choices of  $(n_{\aleph}, n_{\beth}, n_{\beth})$ . The most straightforward way to do so is to consider the derivatives of the data-vector model with respect to the parameters. These can be computed assuming a fiducial cosmology which in our case was described in Sec 4.3.2.

In order to transform the derivatives of the standard bispectrum monopole data-vector into the derivatives of the new one, it is sufficient to apply the same algorithm used to convert the bispectrum into  $\mathbf{g}$ , because the transformation is linear. At this point we have a list of  $\mathbf{g}_{,i} = \partial\mathbf{g}/\partial\theta_i$  for all the elements of the model parameter vector  $\boldsymbol{\theta}$ . The target is to combine these vectors into a unique number expressing the sensitivity of the new data-vector  $\mathbf{g}$  for a determinate choice of  $(n_{\aleph}, n_{\beth}, n_{\beth})$ . We call  $N_{\mathbf{g}}$  the dimension of the new data-vector  $\mathbf{g}$  and  $N_{\boldsymbol{\theta}}$  the number of model parameters.  $N_{\mathbf{g}}$  is of course a function of the number of bins of the new coordinates,  $N_{\mathbf{g}}(n_{\aleph}, n_{\beth}, n_{\beth})$ . For each of the model parameter  $\theta_i$  and for a particular choice of the number of bins  $(n_{\aleph}, n_{\beth}, n_{\beth})_j$  we derive a single number defined as

$$S_{ij} = \sum_{k=0}^{N_{\mathbf{g}}(n_{\aleph}, n_{\beth}, n_{\beth})_j} \frac{1}{N_k^{\text{tr.}}} \left| \frac{\partial g_k}{\partial \theta_i} \right|. \quad (5.5)$$

$S_{ij}$  is a proxy for the sensitivity of the new data-vector  $\mathbf{g}$  defined for a particular choice of number of bins  $(n_{\aleph}, n_{\beth}, n_{\beth})_j$  with respect to variations of the model parameter  $\theta_i$ . Notice that each term of the sum, before being added, is normalised by the number of triangles regrouped in the new bin defined by a set of coordinates  $(\aleph, \beth, \beth)_k$ .

The next step consists of combining these proxies for all the model parameters. This in order to obtain a single number describing the overall sensitivity of  $\mathbf{g}$  for a determinate choice of  $(n_{\aleph}, n_{\beth}, n_{\beth})_j$ . We then normalise each  $i$ -th  $S_{ij}$  dividing by the maximum value of  $S_{ij}$  for all the possible  $(n_{\aleph}, n_{\beth}, n_{\beth})_j$  combinations

$$s_{ij} = \frac{S_{ij}}{\max [S_{ij}]_{\forall j}}, \quad (5.6)$$

so that for all  $\theta_i$  then  $0 < s_{ij} \leq 1$ . Finally all the  $N_{\theta}$   $s_{ij}$  for each  $(n_{\mathbb{N}}, n_{\mathbb{T}}, n_{\mathbb{J}})_j$  combination can be merged into a unique number by doing

$$\bar{s}_j = \sum_{i=0}^{N_{\theta}} s_{ij}. \quad (5.7)$$

We consider  $\bar{s}_j$  as the proxy encoding the overall sensitivity, with respect to the model parameters variation of the new data-vector  $\mathbf{g}$ , defined by a particular choice of the triplet  $(n_{\mathbb{N}}, n_{\mathbb{T}}, n_{\mathbb{J}})_j$ . Since we may want to limit the dimension of  $\mathbf{g}$  in the algorithm, we included a condition setting  $\bar{s}_j = 0$  when  $N_{\mathbf{g}}(n_{\mathbb{N}}, n_{\mathbb{T}}, n_{\mathbb{J}})_j \geq N_{\mathbf{g}}^{\max}$ . The standard BOSS analysis bispectrum data-vector, limited to the range of scales that we consider, has 116 triangles ( $\Delta k_6$  binning case defined in Section 4.3.2). We use the measurements done for the  $\Delta k_2$  binning case corresponding to 2734 triangles for the bispectrum monopole.

We set  $N_{\mathbf{g}}^{\max} = 240$ ,  $\bar{s}_j$  has been computed for all the  $(n_{\mathbb{N}}, n_{\mathbb{T}}, n_{\mathbb{J}})_j$  combinations with  $10 \leq n_{\mathbb{N}}, n_{\mathbb{T}}, n_{\mathbb{J}} \leq 35$ . With these settings we obtained the highest value for  $\bar{s}_j$  in the case of  $(n_{\mathbb{N}} = 24, n_{\mathbb{T}} = 11, n_{\mathbb{J}} = 29)$  corresponding to a dimension  $N_{\mathbf{g}}(24, 11, 29) = 239$ .

### 5.4.3 MC-GC vs. MCMC vs. MC-MP vs. G-PCA

We can compare the results obtained via MC-GC ( $\Delta k_2$  case) in terms of 1 and 2D the posterior distributions obtained via the standard MCMC sampling ( $\Delta k_6$  case), MC-MP ( $\Delta k_2$  case) and G-PCA ( $\Delta k_2$  case). The comparison is shown in Figure 5.1. MC-GC produces a posterior distribution very close to the ones given by MC-MP and G-PCA methods. The agreement is remarkable, especially considering that these compression methods are fairly independent between each other (they have in common only the use of the data-vector derivatives). The precise values of the 1D 68% confidence intervals and of the means of the distribution are reported in Tables 5.1 and 5.2.

The advantage of the MC-GC technique is that it does not require an analytical modelling of the covariance matrix, even if the compression is not as strong as in the MC-MP or G-PCA cases.

Nevertheless, it is not straightforward the application to higher-order statistics like 4pt or 5pt correlation functions. In terms of resources and computing time required, these are approximately the same as for the MC-MP method (Section 4.8.2).

Table 5.1: Four-parameter case, best-fit parameters. Mean values of the posterior distributions and 68% credible intervals for the MCMC and the MC-MP / G-PCA / MC-GC compression methods. The largest  $k$ -binning,  $\Delta k_6$ , the size used in the BOSS analysis, corresponds to the lowest number of triangles (116) and we show the best-fit parameters obtained via MCMC sampling. For the thinnest binning  $\Delta k_2$ , corresponding to the highest number of triangles (2734), we compare the three compression methods. The observed shift in the mean values as a function of the number of considered triangles is due to the strong degeneracy present between the model parameters.

	$\Delta k_6$		$\Delta k_2$		
	MCMC	MC-MP	G-PCA	MC-GC	
$b_1$	$2.41 \pm 0.22$	$2.33 \pm 0.14$	$2.31 \pm 0.17$	$2.25 \pm 0.14$	
$b_2$	$1.00 \pm 0.40$	$0.72 \pm 0.22$	$0.77 \pm 0.19$	$0.68 \pm 0.22$	
$f$	$0.69 \pm 0.08$	$0.63 \pm 0.06$	$0.67 \pm 0.06$	$0.61 \pm 0.06$	
$\sigma_8$	$0.50 \pm 0.04$	$0.53 \pm 0.03$	$0.51 \pm 0.03$	$0.53 \pm 0.03$	

Table 5.2: Four-parameter case, constraints improvement. Below are shown the relative variations in percentage of the size of the 68% credible intervals for the  $\Delta k_2$   $k$ -binning case. MC-GC obtains very similar improvements, in terms of tighter parameter constraints, to the ones obtained via MC-MP and G-PCA.

	$\Delta k_6$	$\Delta k_2$		
	$\Delta \theta_{\Delta k_6}^{\text{mc}}$	$\frac{\Delta \theta^{\text{comp.}} - \Delta \theta_{\Delta k_6}^{\text{mc}}}{\Delta \theta_{\Delta k_6}^{\text{mc}}} [\%]$		
	MCMC	MC-MP	G-PCA	MC-GC
$\Delta b_1$	0.22	-37.1	-24.8	-36.8
$\Delta b_2$	0.40	-46.1	-52.8	-45.0
$\Delta f$	0.08	-27.8	-26.4	-28.7
$\Delta \sigma_8$	0.04	-22.8	-21.0	-22.8
	$\left\langle \frac{\Delta \theta^{\text{comp.}} - \Delta \theta_{\Delta k_6}^{\text{mc}}}{\Delta \theta_{\Delta k_6}^{\text{mc}}} [\%] \right\rangle$	-33.5	-31.3	-33.3

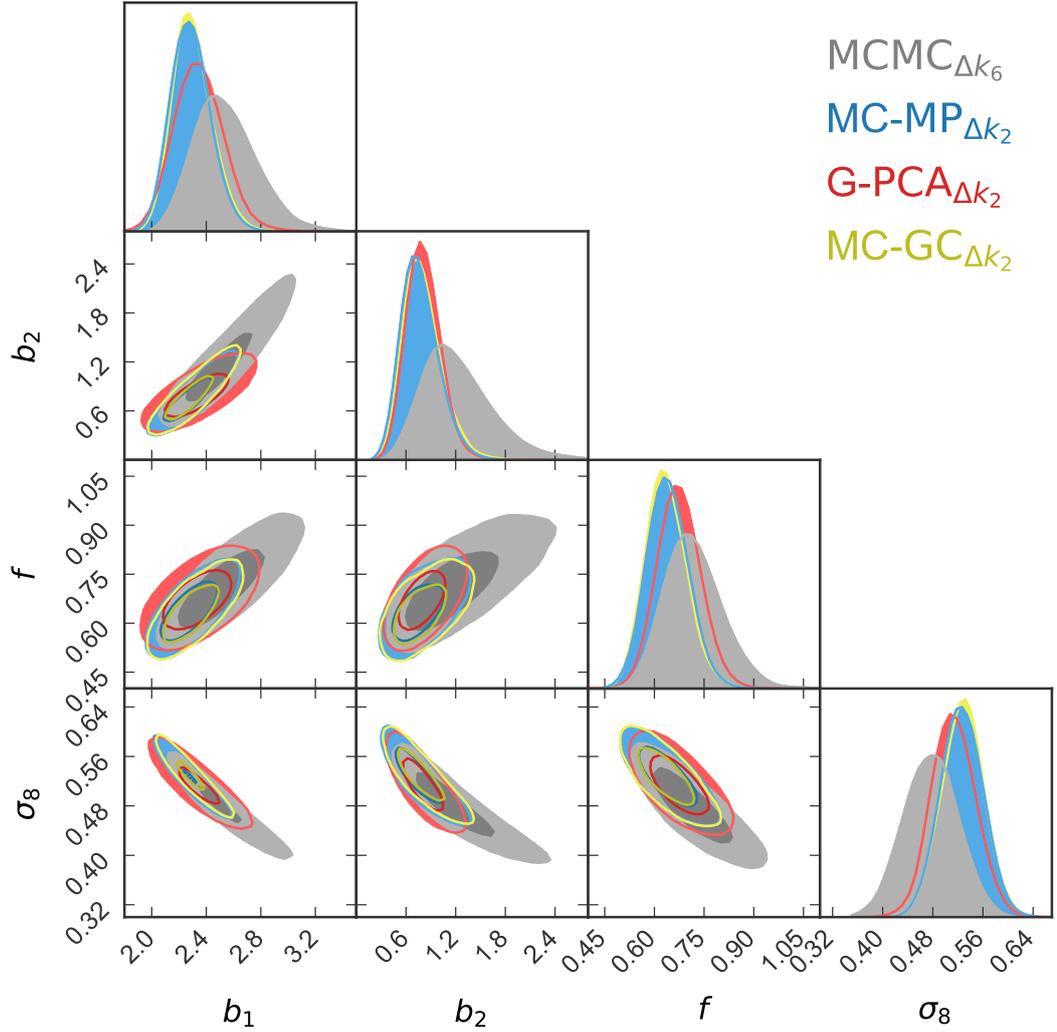


Figure 5.1: MCMC vs. MC-MP vs. G-PCA vs. MC-GC. 2-D 68% and 95% credible contours are shown respectively for the  $\Delta k_6$  MCMC (grey),  $\Delta k_2$  MC-MP (blue),  $\Delta k_2$  G-PCA (red) and  $\Delta k_2$  MC-GC (yellow) cases. The agreement between the MC-MP, G-PCA and MC-GC posterior distributions is remarkable. Without the need of an analytical modelling of the covariance matrix, MC-GC recovers the same posterior distributions derived using MC-MP and G-PCA.

We hope that MC-GC will become, for the future data-sets, the standard procedure to study the bispectra and 3pt functions of the cosmological fields of interest. In the future it would be interesting to find out whether it is possible to reduce the number of parameters characterising each data-vector element using MC-GC, without losing constraining power. This would be very important for example in the case of the redshift space galaxy bispectrum, where each configuration is described by five parameters, as seen in Chapter 3.

“Well, here at last, dear friends, on the shores of the Sea comes the end of our fellowship in Middle-earth. Go in peace! I will not say: do not weep; for not all tears are an evil.”

- J.R.R. Tolkien, *The Return of the King*

# Bibliography

- Milton Abramowitz. *Handbook of Mathematical Functions, With Formulas, Graphs, and Mathematical Tables*. Dover Publications, Incorporated, 1974. ISBN 0486612724.
- P. A. R. Ade et al. Planck 2013 results. XVI. Cosmological parameters. *Astron. Astrophys.*, 571:A16, 2014. doi: 10.1051/0004-6361/201321591.
- P. A. R. Ade et al. Planck 2015 results. XIII. Cosmological parameters. *Astron. Astrophys.*, 594:A13, 2016. doi: 10.1051/0004-6361/201525830.
- I. Affleck and M. Dine. A new mechanism for baryogenesis. *Nuclear Physics B*, 249:361–380, 1985. doi: 10.1016/0550-3213(85)90021-5.
- S.and others Alam. The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample. *MNRAS*, 470:2617–2652, September 2017. doi: 10.1093/mnras/stx721.
- A. Albrecht and P. J. Steinhardt. Cosmology for grand unified theories with radiatively induced symmetry breaking. *Physical Review Letters*, 48:1220–1223, April 1982. doi: 10.1103/PhysRevLett.48.1220.
- C. Alcock and B. Paczynski. An evolution free test for non-zero cosmological constant. *Nature*, 281:358, October 1979. doi: 10.1038/281358a0.
- J. Alsing and B. Wandelt. Generalized massive optimal data compression. *MNRAS*, 476:L60–L64, May 2018. doi: 10.1093/mnrasl/sly029.
- J. Alsing, A. Heavens, and A. H. Jaffe. Cosmological parameters, shear maps and power spectra from CFHTLenS using Bayesian hierarchical inference. *MNRAS*, 466:3272–3292, April 2017. doi: 10.1093/mnras/stw3161.
- J. Alsing et al. Massive optimal data compression and density estimation for scalable, likelihood-free inference in cosmology. *MNRAS*, 477:2514–2525, July 2018. doi: 10.1093/mnras/sty819.
- L. Amendola and S. Tsujikawa. *Dark Energy: Theory and Observations*. 2010.
- L.and others Anderson. The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Data Release 9 spectroscopic galaxy sample. *MNRAS*, 427:3435–3467, December 2012. doi: 10.1111/j.1365-2966.2012.22066.x.

- L. and others Anderson. The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Data Releases 10 and 11 Galaxy samples. *MNRAS*, 441:24–62, June 2014. doi: 10.1093/mnras/stu523.
- P. Astier, J. Guy, N. Regnault, R. Pain, E. Aubourg, D. Balam, S. Basa, R. G. Carlberg, and et al. The Supernova Legacy Survey: measurement of  $\Omega_M$ ,  $\Omega_\Lambda$  and  $w$  from the first year data set. *A&A*, 447:31–48, February 2006. doi: 10.1051/0004-6361:20054185.
- Tobias Baldauf et al. Evidence for Quadratic Tidal Tensor Bias from the Halo Bispectrum. *Phys. Rev.*, D86:083540, 2012. doi: 10.1103/PhysRevD.86.083540.
- K. Bamba, S. Capozziello, S. Nojiri, and S. D. Odintsov. Dark energy cosmology: the equivalent description via different theoretical models and cosmography tests. *Astrophysics and Space Science*, 342:155–228, November 2012. doi: 10.1007/s10509-012-1181-8.
- J. M. Bardeen, P. J. Steinhardt, and M. S. Turner. Spontaneous creation of almost scale-free density perturbations in an inflationary universe. *Phys. Rev. D*, 28:679–693, August 1983. doi: 10.1103/PhysRevD.28.679.
- J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay. The statistics of peaks of Gaussian random fields. *ApJ*, 304:15–61, May 1986a. doi: 10.1086/164143.
- J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay. The statistics of peaks of Gaussian random fields. *ApJ*, 304:15–61, May 1986b. doi: 10.1086/164143.
- M. Bartelmann and P. Schneider. Weak gravitational lensing. *Physics Reports*, 340:291–472, January 2001. doi: 10.1016/S0370-1573(00)00082-X.
- C. M. Baugh and G. Efstathiou. The Three-Dimensional Power Spectrum Measured from the APM Galaxy Survey - Part One - Use of the Angular Correlation Function. *MNRAS*, 265:145, November 1993. doi: 10.1093/mnras/265.1.145.
- J. Baur et al. Lyman-alpha forests cool warm dark matter. *JCAP*, 8:012, August 2016. doi: 10.1088/1475-7516/2016/08/012.
- J. D. Bekenstein. Relativistic gravitation theory for the modified Newtonian dynamics paradigm. *Phys. Rev. D*, 70(8):083509, October 2004. doi: 10.1103/PhysRevD.70.083509.
- Julien Bel, Kai Hoffmann, and Enrique Gaztañaga. Non-local bias contribution to third-order galaxy correlations. *Mon. Not. Roy. Astron. Soc.*, 453(1):259–276, 2015. doi: 10.1093/mnras/stv1600.
- J. Benjamin, C. Heymans, E. Semboloni, L. van Waerbeke, H. Hoekstra, T. Erben, M. D. Gladders, M. Hetterscheidt, Y. Mellier, and H. K. C. Yee. Cos-

- mological constraints from the 100-deg<sup>2</sup> weak-lensing survey. *MNRAS*, 381: 702–712, October 2007. doi: 10.1111/j.1365-2966.2007.12202.x.
- C. L. Bennett, M. Bay, M. Halpern, G. Hinshaw, C. Jackson, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, L. Page, D. N. Spergel, G. S. Tucker, D. T. Wilkinson, E. Wollack, and E. L. Wright. The Microwave Anisotropy Probe Mission. *ApJ*, 583:1–23, January 2003a. doi: 10.1086/345346.
- C. L. Bennett et al. First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Preliminary Maps and Basic Results. *ApJS*, 148:1–27, September 2003b. doi: 10.1086/377253.
- L. Bernard and L. Blanchet. Phenomenology of dark matter via a bimetric extension of general relativity. *Phys. Rev. D*, 91(10):103536, May 2015. doi: 10.1103/PhysRevD.91.103536.
- F. Bernardeau. Skewness and kurtosis in large-scale cosmic fields. *ApJ*, 433:1–18, September 1994. doi: 10.1086/174620.
- F. Bernardeau, S. Colombi, E. Gaztañaga, and R. Scoccimarro. Large-scale structure of the Universe and cosmological perturbation theory. *Physics Reports*, 367:1–248, September 2002. doi: 10.1016/S0370-1573(02)00135-7.
- Francis Bernardeau and Philippe Brax. Cosmological Large-scale Structures beyond Linear Theory in Modified Gravity. *JCAP*, 1106:019, 2011. doi: 10.1088/1475-7516/2011/06/019.
- D. Bertacca et al. Relativistic wide-angle galaxy bispectrum on the light-cone. *ArXiv e-prints*, May 2017.
- C. Blake et al. The WiggleZ Dark Energy Survey: joint measurements of the expansion and growth history at  $z \leq 1$ . *MNRAS*, 425:405–414, September 2012. doi: 10.1111/j.1365-2966.2012.21473.x.
- Alexander Borisov and Bhuvnesh Jain. Three-point correlations in  $f(r)$  models of gravity. *Phys. Rev. D*, 79:103506, May 2009. doi: 10.1103/PhysRevD.79.103506. URL <https://link.aps.org/doi/10.1103/PhysRevD.79.103506>.
- A. Burkert and J. Silk. Dark Baryons and Rotation Curves. *ApJ*, 488:L55–L58, October 1997. doi: 10.1086/310935.
- J. Byun et al. Towards optimal cosmological parameter recovery from compressed bispectrum statistics. *MNRAS*, 471:1581–1618, October 2017. doi: 10.1093/mnras/stx1681.
- Kwan Chuen Chan, Roman Scoccimarro, and Ravi K. Sheth. Gravity and Large-Scale Non-local Bias. *Phys. Rev.*, D85:083509, 2012. doi: 10.1103/PhysRevD.85.083509.
- L. Clerkin, D. Kirk, O. Lahav, F. B. Abdalla, and E. Gaztañaga. A prescription for galaxy biasing evolution as a nuisance parameter. *MNRAS*, 448:1389–1401, April 2015. doi: 10.1093/mnras/stu2754.

- D. Clowe, M. Bradač, A. H. Gonzalez, M. Markevitch, S. W. Randall, C. Jones, and D. Zaritsky. A Direct Empirical Proof of the Existence of Dark Matter. *ApJ*, 648:L109–L113, September 2006. doi: 10.1086/508162.
- S. Cole et al. The 2dF Galaxy Redshift Survey: power-spectrum analysis of the final data set and cosmological implications. *MNRAS*, 362:505–534, September 2005. doi: 10.1111/j.1365-2966.2005.09318.x.
- E. J. Copeland, M. Sami, and S. Tsujikawa. Dynamics of Dark Energy. *International Journal of Modern Physics D*, 15:1753–1935, 2006. doi: 10.1142/S021827180600942X.
- M. Crocce et al. The MICE Grand Challenge lightcone simulation - II. Halo and galaxy catalogues. *MNRAS*, 453:1513–1530, October 2015. doi: 10.1093/mnras/stv1708.
- Dark Energy Survey Collaboration et al. The Dark Energy Survey: more than dark energy - an overview. *MNRAS*, 460:1270–1299, August 2016. doi: 10.1093/mnras/stw641.
- K. S. Dawson et al. The Baryon Oscillation Spectroscopic Survey of SDSS-III. *AJ*, 145:10, January 2013. doi: 10.1088/0004-6256/145/1/10.
- J. T. A. de Jong et al. The Kilo-Degree Survey. *Experimental Astronomy*, 35: 25–44, January 2013a. doi: 10.1007/s10686-012-9306-1.
- J. T. A. de Jong et al. The Kilo-Degree Survey. *The Messenger*, 154:44–46, December 2013b.
- A. Dekel and O. Lahav. Stochastic Nonlinear Galaxy Biasing. *ApJ*, 520:24–34, July 1999a. doi: 10.1086/307428.
- A. Dekel and O. Lahav. Stochastic Nonlinear Galaxy Biasing. *ApJ*, 520:24–34, July 1999b. doi: 10.1086/307428.
- DESI Collaboration and others. The DESI Experiment Part I: Science, Targeting, and Survey Design. *ArXiv e-prints*, October 2016.
- S. Dodelson. *Modern Cosmology*. Academic Press. Academic Press, 2003. ISBN: 9780122191411.
- George Efstathiou. H0 revisited. *Monthly Notices of the Royal Astronomical Society*, 440(2):1138–1152, 2014. doi: 10.1093/mnras/stu278. URL <http://dx.doi.org/10.1093/mnras/stu278>.
- D. J. Eisenstein and W. Hu. Power Spectra for Cold Dark Matter and Its Variants. *ApJ*, 511:5–15, January 1999. doi: 10.1086/306640.
- D. J. Eisenstein et al. Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies. *ApJ*, 633:560–574, November 2005. doi: 10.1086/466512.

- D. J. Eisenstein et al. SDSS-III: Massive Spectroscopic Surveys of the Distant Universe, the Milky Way, and Extra-Solar Planetary Systems. *AJ*, 142:72, September 2011. doi: 10.1088/0004-6256/142/3/72.
- E. E. Falco, C. S. Kochanek, and J. A. Muñoz. Limits on Cosmological Models from Radio-selected Gravitational Lenses. *ApJ*, 494:47–59, February 1998. doi: 10.1086/305207.
- H. A. Feldman et al. Power-spectrum analysis of three-dimensional redshift surveys. *ApJ*, 426:23–37, May 1994. doi: 10.1086/174036.
- J. R. Fergusson, D. M. Regan, and E. P. S. Shellard. Rapid Separable Analysis of Higher Order Correlators in Large Scale Structure. *Phys. Rev.*, D86:063511, 2012. doi: 10.1103/PhysRevD.86.063511.
- D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman. emcee: The MCMC Hammer. *PASP*, 125:306, March 2013. doi: 10.1086/670067.
- W. L. Freedman et al. Final Results from the Hubble Space Telescope Key Project to Measure the Hubble Constant. *ApJ*, 553:47–72, May 2001. doi: 10.1086/320638.
- J. A. Frieman and E. Gaztanaga. The three-point function as a probe of models for large-scale structure. *ApJ*, 425:392–402, April 1994. doi: 10.1086/173995.
- J. A. Frieman, B. Bassett, A. Becker, C. Choi, D. Cinabro, F. DeJongh, D. L. Depoy, B. Dilday, M. Doi, P. M. Garnavich, C. J. Hogan, Holtzman, and et al. The Sloan Digital Sky Survey-II Supernova Survey: Technical Summary. *AJ*, 135:338–347, January 2008a. doi: 10.1088/0004-6256/135/1/338.
- J. A. Frieman, M. S. Turner, and D. Huterer. Dark Energy and the Accelerating Universe. *ARA&A*, 46:385–432, September 2008b. doi: 10.1146/annurev.astro.46.060407.145243.
- J. N. Fry. The Galaxy correlation hierarchy in perturbation theory. *ApJ*, 279: 499–510, April 1984. doi: 10.1086/161913.
- J. N. Fry and E. Gaztanaga. Biasing and hierarchical statistics in large-scale structure. *ApJ*, 413:447–452, August 1993. doi: 10.1086/173015.
- J. N. Fry and R. J. Scherrer. Skewness in large-scale structure and non-Gaussian initial conditions. *ApJ*, 429:36–42, July 1994. doi: 10.1086/174300.
- J. N. Fry and M. Seldner. Transform analysis of the high-resolution Shane-Wirtanen Catalog - The power spectrum and the bispectrum. *ApJ*, 259:474–481, August 1982. doi: 10.1086/160184.
- James N. Fry and Enrique Gaztanaga. Biasing and hierarchical statistics in large scale structure. *Astrophys. J.*, 413:447–452, 1993. doi: 10.1086/173015.
- L. Fu et al. CFHTLenS: cosmological constraints from a combination of cosmic shear two-point and three-point correlations. *MNRAS*, 441:2725–2743, July 2014. doi: 10.1093/mnras/stu754.

- M. Fukugita, C. J. Hogan, and P. J. E. Peebles. The Cosmic Baryon Budget. *ApJ*, 503:518–530, August 1998. doi: 10.1086/306025.
- M. Fukugita et al. The Sloan Digital Sky Survey Photometric System. *AJ*, 111: 1748, April 1996. doi: 10.1086/117915.
- Edgar Gabriel, , et al. Open MPI: Goals, concept, and design of a next generation MPI implementation. In *Proceedings, 11th European PVM/MPI Users' Group Meeting*, pages 97–104, Budapest, Hungary, September 2004.
- Praful Gagrani and Lado Samushia. Information Content of the Angular Multipoles of Redshift-Space Galaxy Bispectrum. *Mon. Not. Roy. Astron. Soc.*, 467 (1):928–935, 2017. doi: 10.1093/mnras/stx135.
- A. Gangui et al. The three-point correlation function of the cosmic microwave background in inflationary models. *ApJ*, 430:447–457, August 1994. doi: 10.1086/174421.
- E. Gaztanaga. High-Order Galaxy Correlation Functions in the APM Galaxy Survey. *MNRAS*, 268:913, June 1994. doi: 10.1093/mnras/268.4.913.
- E. Gaztanaga and J. A. Frieman. Bias and high-order galaxy correlation functions in the APM galaxy survey. *ApJ*, 437:L13–L16, December 1994. doi: 10.1086/187671.
- S. Genel et al. Introducing the Illustris project: the evolution of galaxy populations across cosmic time. *MNRAS*, 445:175–200, November 2014. doi: 10.1093/mnras/stu1654.
- Gil-Marín et al. The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: RSD measurement from the LOS-dependent power spectrum of DR12 BOSS galaxies. *MNRAS*, 460:4188–4209, August 2016a. doi: 10.1093/mnras/stw1096.
- Gil-Marín et al. The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: BAO measurement from the LOS-dependent power spectrum of DR12 BOSS galaxies. *MNRAS*, 460:4210–4219, August 2016b. doi: 10.1093/mnras/stw1264.
- H. Gil-Marín, F. Schmidt, W. Hu, R. Jimenez, and L. Verde. The bispectrum of  $f(R)$  cosmologies. *JCAP*, 11:019, November 2011. doi: 10.1088/1475-7516/2011/11/019.
- H. Gil-Marín, C. Wagner, J. Noreña, L. Verde, and W. Percival. Dark matter and halo bispectrum in redshift space: theory and applications. *JCAP*, 12: 029, December 2014. doi: 10.1088/1475-7516/2014/12/029.
- H. Gil-Marín et al. The power spectrum and bispectrum of SDSS DR11 BOSS galaxies - I. Bias and gravity. *MNRAS*, 451:539–580, July 2015. doi: 10.1093/mnras/stv961.

- H. Gil-Marín et al. 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. *MNRAS*, 465:1757–1788, February 2017. doi: 10.1093/mnras/stw2679.
- H. and others Gil-Marín. 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. *MNRAS*, 465:1757–1788, February 2017. doi: 10.1093/mnras/stw2679.
- E. B. Gliner. Algebraic Properties of the Energy-momentum Tensor and Vacuum-like States of  $^+$  Matter. *Soviet Journal of Experimental and Theoretical Physics*, 22:378, February 1966.
- G. Goldhaber. The Acceleration of the Expansion of the Universe: A Brief Early History of the Supernova Cosmology Project (SCP). In D. B. Cline, editor, *American Institute of Physics Conference Series*, volume 1166 of *American Institute of Physics Conference Series*, pages 53–72, September 2009. doi: 10.1063/1.3232196.
- E. J. Groth and P. J. E. Peebles. Statistical analysis of catalogs of extragalactic objects. VII - Two- and three-point correlation functions for the high-resolution Shane-Wirtanen catalog of galaxies. *ApJ*, 217:385–405, October 1977. doi: 10.1086/155588.
- Gualdi et al. Enhancing BOSS bispectrum cosmological constraints with maximal compression. *ArXiv e-prints*, June 2018a.
- D. Gualdi, M. Manera, B. Joachimi, and O. Lahav. Maximal compression of the redshift space galaxy power spectrum and bispectrum. *MNRAS*, January 2018b. doi: 10.1093/mnras/sty261.
- J. E. Gunn et al. The 2.5 m Telescope of the Sloan Digital Sky Survey. *AJ*, 131: 2332–2359, April 2006. doi: 10.1086/500975.
- A. H. Guth. Inflationary universe: A possible solution to the horizon and flatness problems. *Phys. Rev. D*, 23:347–356, January 1981. doi: 10.1103/PhysRevD.23.347.
- A. H. Guth and S.-Y. Pi. Fluctuations in the new inflationary universe. *Physical Review Letters*, 49:1110–1113, October 1982. doi: 10.1103/PhysRevLett.49.1110.
- L. Guzzo et al. A test of the nature of cosmic acceleration using galaxy redshift distortions. *Nature*, 451:541–544, January 2008. doi: 10.1038/nature06555.
- A. Hajian and T. Souradeep. Measuring the Statistical Isotropy of the Cosmic Microwave Background Anisotropy. *ApJ*, 597:L5–L8, November 2003. doi: 10.1086/379757.

- J. Hartlap, P. Simon, and P. Schneider. Why your model parameter confidences might be too optimistic. Unbiased estimation of the inverse covariance matrix. *A&A*, 464:399–404, March 2007. doi: 10.1051/0004-6361:20066170.
- I. Hashimoto, Y. Rasera, and A. Taruya. Precision cosmology with redshift-space bispectrum: a perturbation theory based model at one-loop order. *ArXiv e-prints*, May 2017.
- S. W. Hawking. The development of irregularities in a single bubble inflationary universe. *Physics Letters B*, 115:295–297, September 1982. doi: 10.1016/0370-2693(82)90373-2.
- A. Heavens. Statistical techniques in cosmology. *ArXiv e-prints*, June 2009.
- A. F. Heavens, S. Matarrese, and L. Verde. The non-linear redshift-space power spectrum of galaxies. *MNRAS*, 301:797–808, December 1998. doi: 10.1046/j.1365-8711.1998.02052.x.
- A. F. Heavens, S. Matarrese, and Licia Verde. The Nonlinear redshift-space power spectrum of galaxies. *Mon. Not. Roy. Astron. Soc.*, 301:797–808, 1998. doi: 10.1046/j.1365-8711.1998.02052.x.
- A. F. Heavens, R. Jimenez, and O. Lahav. Massive lossless data compression and multiple parameter estimation from galaxy spectra. *MNRAS*, 317:965–972, October 2000. doi: 10.1046/j.1365-8711.2000.03692.x.
- A. F. Heavens, E. Sellentin, D. de Mijolla, and A. Vianello. Massive data compression for parameter-dependent covariance matrices. *MNRAS*, 472:4244–4250, December 2017. doi: 10.1093/mnras/stx2326.
- H. Hildebrandt, M. Viola, C. Heymans, S. Joudaki, K. Kuijken, C. Blake, T. Erben, B. Joachimi, and et al. KiDS-450: cosmological parameter constraints from tomographic weak gravitational lensing. *MNRAS*, 465:1454–1498, February 2017. doi: 10.1093/mnras/stw2805.
- E. Hivon et al. Redshift distortions of clustering: a Lagrangian approach. *A&A*, 298:643, June 1995.
- K. Hoffmann, Y. Mao, H. Mo, and B. D. Wandelt. Signatures of Cosmic Reionization on the 21cm 2- and 3-point Correlation Function I: Quadratic Bias Modeling. *ArXiv e-prints*, February 2018.
- C. Howlett and W. J. Percival. Galaxy two-point covariance matrix estimation for next generation surveys. *MNRAS*, 472:4935–4952, December 2017. doi: 10.1093/mnras/stx2342.
- E. Hubble and M. L. Humason. The Velocity-Distance Relation among Extra-Galactic Nebulae. *ApJ*, 74:43, July 1931. doi: 10.1086/143323.
- B. Jain and E. Bertschinger. Second-order power spectrum and nonlinear evolution at high redshift. *ApJ*, 431:495–505, August 1994. doi: 10.1086/174502.

- M. Jarvis, G. Bernstein, and B. Jain. The skewness of the aperture mass statistic. *MNRAS*, 352:338–352, July 2004. doi: 10.1111/j.1365-2966.2004.07926.x.
- B. Joachimi, X. Shi, and P. Schneider. Bispectrum covariance in the flat-sky limit. *A&A*, 508:1193–1204, December 2009. doi: 10.1051/0004-6361/200912906.
- A. Joyce, L. Lombriser, and F. Schmidt. Dark Energy Versus Modified Gravity. *Annual Review of Nuclear and Particle Science*, 66:95–122, October 2016. doi: 10.1146/annurev-nucl-102115-044553.
- R. and others Juszkiewicz. Skewness as a probe of baryon acoustic oscillations. *MNRAS*, 429:1206–1212, February 2013. doi: 10.1093/mnras/sts409.
- N. Kaiser. On the spatial correlations of Abell clusters. *ApJ*, 284:L9–L12, September 1984a. doi: 10.1086/184341.
- N. Kaiser. On the spatial correlations of Abell clusters. *ApJ*, 284:L9–L12, September 1984b. doi: 10.1086/184341.
- N. Kaiser. Clustering in real space and in redshift space. *MNRAS*, 227:1–21, July 1987. doi: 10.1093/mnras/227.1.1.
- Marc Kamionkowski and Ari Buchalter. Weakly nonlinear clustering for arbitrary expansion histories. *Astrophys.J.*, 514:7, 1999. doi: 10.1086/306950.
- N. Katz, L. Hernquist, and D. H. Weinberg. The Clustering of High-Redshift Galaxies in the Cold Dark Matter Scenario. *ApJ*, 523:463–479, October 1999. doi: 10.1086/307744.
- I. Kayo and M. Takada. Cosmological parameters from weak lensing power spectrum and bispectrum tomography: including the non-Gaussian errors. *ArXiv e-prints*, June 2013.
- Kayo, I. and Takada, M. and Jain, B. Information content of weak lensing power spectrum and bispectrum: including the non-Gaussian error covariance matrix. *MNRAS*, 429:344–371, February 2013. doi: 10.1093/mnras/sts340.
- M. Kilbinger. Cosmology with cosmic shear observations: a review. *Reports on Progress in Physics*, 78(8):086901, July 2015. doi: 10.1088/0034-4885/78/8/086901.
- M. Kilbinger and P. Schneider. Cosmological parameters from combined second- and third-order aperture mass statistics of cosmic shear. *A&A*, 442:69–83, October 2005. doi: 10.1051/0004-6361:20053531.
- F.-S. Kitaura and S. Heß. Cosmological structure formation with augmented Lagrangian perturbation theory. *MNRAS*, 435:L78–L82, August 2013. doi: 10.1093/mnrasl/slt101.
- F.-S. Kitaura et al. The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: mock galaxy catalogues for the BOSS Final Data Release. *MNRAS*, 456:4156–4173, March 2016. doi: 10.1093/mnras/stv2826.

- Francisco-Shu Kitaura et al. The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: mock galaxy catalogues for the BOSS Final Data Release. *Mon. Not. Roy. Astron. Soc.*, 456(4):4156–4173, 2016. doi: 10.1093/mnras/stv2826.
- L. Kofman. The Origin of Matter in the Universe: Reheating after Inflation. *ArXiv Astrophysics e-prints*, May 1996.
- E. Komatsu, J. Dunkley, M. R.olta, C. L. Bennett, B. Gold, G. Hinshaw, N. Jarosik, D. Larson, M. Limon, L. Page, D. N. Spergel, M. Halpern, R. S. Hill, A. Kogut, S. S. Meyer, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. Five-Year Wilkinson Microwave Anisotropy Probe Observations: Cosmological Interpretation. *ApJS*, 180:330–376, February 2009. doi: 10.1088/0067-0049/180/2/330.
- Ofer Lahav, Per B. Lilje, Joel R. Primack, and Martin J. Rees. Dynamical effects of the cosmological constant. *Mon. Not. Roy. Astron. Soc.*, 251:128–136, 1991.
- J. M. Lamarre, J. L. Puget, F. Bouchet, P. A. R. Ade, A. Benoit, J. P. Bernard, J. Bock, and et al. The Planck High Frequency Instrument, a third generation CMB experiment, and a full sky submillimeter survey. *nar*, 47:1017–1024, December 2003. doi: 10.1016/j.newar.2003.09.006.
- R. Laureijs et al. Euclid Definition Study Report. *ArXiv e-prints*, October 2011.
- A. Lazanu, T. Giannantonio, M. Schmittfull, and E. P. S. Shellard. Matter bispectrum of large-scale structure: Three-dimensional comparison between theoretical models and numerical simulations. *Phys. Rev. D*, 93(8):083517, April 2016. doi: 10.1103/PhysRevD.93.083517.
- M. Levi et al. The DESI Experiment, a whitepaper for Snowmass 2013. *ArXiv e-prints*, August 2013.
- A. Lewis, A. Challinor, and A. Lasenby. Efficient Computation of Cosmic Microwave Background Anisotropies in Closed Friedmann-Robertson-Walker Models. *ApJ*, 538:473–476, August 2000. doi: 10.1086/309179.
- Antony Lewis, Anthony Challinor, and Anthony Lasenby. Efficient computation of CMB anisotropies in closed FRW models. *Astrophys. J.*, 538:473–476, 2000. doi: 10.1086/309179.
- M. Liguori, E. Sefusatti, J. R. Fergusson, and E. P. S. Shellard. Primordial Non-Gaussianity and Bispectrum Measurements in the Cosmic Microwave Background and Large-Scale Structure. *Advances in Astronomy*, 2010:980523, 2010. doi: 10.1155/2010/980523.
- A. Linde. Recent Progress in Inflationary Cosmology. In K. Sato, T. Suginoara, and N. Sugiyama, editors, *Cosmological Constant and the Evolution of the Universe*, page 133, 1996.

- A. D. Linde. Scalar field fluctuations in the expanding universe and the new inflationary universe scenario. *Physics Letters B*, 116:335–339, October 1982. doi: 10.1016/0370-2693(82)90293-3.
- Eric V. Linder. Cosmic growth history and expansion history. *Phys. Rev.*, D72: 043529, 2005. doi: 10.1103/PhysRevD.72.043529.
- LSST Science Collaboration et al. LSST Science Book, Version 2.0. *ArXiv e-prints*, December 2009.
- R. Lynds. The Absorption-Line Spectrum of 4c 05.34. *ApJ*, 164:L73, March 1971. doi: 10.1086/180695.
- S. J. Maddox, W. J. Sutherland, G. Efstathiou, and J. Loveday. The APM galaxy survey. I - APM measurements and star-galaxy separation. *MNRAS*, 243:692–712, April 1990.
- M. Manera and E. Gaztañaga. The local bias model in the large-scale halo distribution. *MNRAS*, 415:383–398, July 2011. doi: 10.1111/j.1365-2966.2011.18705.x.
- M. Manera et al. The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: a large sample of mock galaxy catalogues. *MNRAS*, 428:1036–1054, January 2013. doi: 10.1093/mnras/sts084.
- Marc Manera, Lado Samushia, Rita Tojeiro, Cullan Howlett, Ashley J. Ross, Will J. Percival, Hector Gil-Marín, Joel R. Brownstein, Angela Burden, and Francesco Montesano. The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: mock galaxy catalogues for the low-redshift sample. *Mon. Not. Roy. Astron. Soc.*, 447:437, 2015. doi: 10.1093/mnras/stu2465.
- Q. and others Mao. Constraining primordial non-Gaussianity with moments of the large-scale density field. *MNRAS*, 443:1402–1415, September 2014. doi: 10.1093/mnras/stu1255.
- S. Matarrese, L. Verde, and A. F. Heavens. Large-scale bias in the Universe: bispectrum method. *MNRAS*, 290:651–662, October 1997. doi: 10.1093/mnras/290.4.651.
- S. Matarrese, Licia Verde, and A. F. Heavens. Large scale bias in the universe: Bispectrum method. *Mon. Not. Roy. Astron. Soc.*, 290:651–662, 1997. doi: 10.1093/mnras/290.4.651.
- P. McDonald and A. Roy. Clustering of dark matter tracers: generalizing bias for the coming era of precision LSS. *JCAP*, 8:020, August 2009a. doi: 10.1088/1475-7516/2009/08/020.
- P. McDonald and A. Roy. Clustering of dark matter tracers: generalizing bias for the coming era of precision LSS. *JCAP*, 8:020, August 2009b. doi: 10.1088/1475-7516/2009/08/020.

- M. Milgrom. MOND theory. *Canadian Journal of Physics*, 93:107–118, February 2015. doi: 10.1139/cjp-2014-0211.
- S. and others Miyazaki. Hyper Suprime-Cam. In *Ground-based and Airborne Instrumentation for Astronomy IV*, volume 8446 of *Proceedings of the SPIE*, page 84460Z, September 2012. doi: 10.1117/12.926844.
- C. Modi, E. Castorina, and U. Seljak. Halo bias in Lagrangian space: estimators and theoretical predictions. *MNRAS*, 472:3959–3970, December 2017. doi: 10.1093/mnras/stx2148.
- I. Mohammed, U. Seljak, and Z. Vlah. Perturbative approach to covariance matrix of the matter power spectrum. *MNRAS*, 466:780–797, April 2017. doi: 10.1093/mnras/stw3196.
- V. Mukhanov. *Physical Foundations of Cosmology*. November 2005. doi: 10.2277/0521563984.
- A. Nusser and M. Davis. The Cosmological Bulk Flow: Consistency with  $\Lambda$ CDM and  $z \approx 0$  Constraints on  $\sigma_8$  and  $\gamma$ . *ApJ*, 736:93, August 2011. doi: 10.1088/0004-637X/736/2/93.
- D. Parkinson et al. The WiggleZ Dark Energy Survey: Final data release and cosmological results. *Phys. Rev. D*, 86(10):103518, November 2012. doi: 10.1103/PhysRevD.86.103518.
- C. Patrignani et al. Review of Particle Physics. *Chin. Phys.*, C40(10):100001, 2016. doi: 10.1088/1674-1137/40/10/100001.
- J. A. Peacock and A. F. Heavens. The statistics of maxima in primordial density perturbations. *MNRAS*, 217:805–820, December 1985. doi: 10.1093/mnras/217.4.805.
- F. R. and others Pearce. A Simulation of Galaxy Formation and Clustering. *ApJ*, 521:L99–L102, August 1999. doi: 10.1086/312196.
- D. W. Pearson and L. Samushia. A Detection of the Baryon Acoustic Oscillation Features in the SDSS BOSS DR12 Galaxy Bispectrum. *ArXiv e-prints*, December 2017.
- P. J. Peebles and B. Ratra. The cosmological constant and dark energy. *Reviews of Modern Physics*, 75:559–606, April 2003. doi: 10.1103/RevModPhys.75.559.
- P. J. E. Peebles. *The large-scale structure of the universe*. 1980.
- P. J. E. Peebles and E. J. Groth. Statistical analysis of catalogs of extragalactic objects. V - Three-point correlation function for the galaxy distribution in the Zwicky catalog. *ApJ*, 196:1–11, February 1975. doi: 10.1086/153390.
- P. J. E. Peebles and J. T. Yu. Primeval Adiabatic Perturbation in an Expanding Universe. *ApJ*, 162:815, December 1970. doi: 10.1086/150713.

- W. J. Percival et al. Baryon acoustic oscillations in the Sloan Digital Sky Survey Data Release 7 galaxy sample. *MNRAS*, 401:2148–2168, February 2010. doi: 10.1111/j.1365-2966.2009.15812.x.
- W. J. Percival et al. The clustering of Galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: including covariance matrix errors. *MNRAS*, 439:2531–2541, April 2014. doi: 10.1093/mnras/stu112.
- S. Perlmutter et al. Measurements of  $\Omega$  and  $\Lambda$  from 42 High-Redshift Supernovae. *ApJ*, 517:565–586, June 1999. doi: 10.1086/307221.
- J. Pielorz, J. Rödiger, I. Tereno, and P. Schneider. A fitting formula for the non-Gaussian contribution to the lensing power spectrum covariance. *A&A*, 514:A79, May 2010. doi: 10.1051/0004-6361/200912854.
- Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Bacigalupi, M. Ballardini, and et al. Planck 2018 results. VI. Cosmological parameters. *ArXiv e-prints*, July 2018.
- Planck Collaboration et al. Planck 2013 results. XVI. Cosmological parameters. *A&A*, 571:A16, November 2014. doi: 10.1051/0004-6361/201321591.
- Planck Collaboration et al. Planck 2015 results. XIII. Cosmological parameters. *A&A*, 594:A13, September 2016a. doi: 10.1051/0004-6361/201525830.
- Planck Collaboration et al. Planck 2015 results. XVII. Constraints on primordial non-Gaussianity. *A&A*, 594:A17, September 2016b. doi: 10.1051/0004-6361/201525836.
- S. Pyne, B. Joachimi, and H. V. Peiris. Weak lensing deflection of three-point correlation functions. *JCAP*, 12:043, December 2017. doi: 10.1088/1475-7516/2017/12/043.
- A. Refregier. Weak Gravitational Lensing by Large-Scale Structure. *ARA&A*, 41:645–668, 2003. doi: 10.1146/annurev.astro.41.111302.102207.
- D. M. Regan, M. M. Schmittfull, E. P. S. Shellard, and J. R. Fergusson. Universal non-gaussian initial conditions for  $n$ -body simulations. *Phys. Rev. D*, 86:123524, Dec 2012. doi: 10.1103/PhysRevD.86.123524. URL <https://link.aps.org/doi/10.1103/PhysRevD.86.123524>.
- B.and others Reid. SDSS-III Baryon Oscillation Spectroscopic Survey Data Release 12: galaxy target selection and large-scale structure catalogues. *MNRAS*, 455:1553–1573, January 2016. doi: 10.1093/mnras/stv2382.
- A. G. Riess et al. Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. *AJ*, 116:1009–1038, September 1998. doi: 10.1086/300499.
- G. Rossi. Neutrino mass from the Lyman-Alpha forest. *ArXiv e-prints*, June 2014.

- V. C. Rubin and W. K. Ford, Jr. Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions. *ApJ*, 159:379, February 1970. doi: 10.1086/150317.
- D. Saadeh, S. M. Feeney, A. Pontzen, H. V. Peiris, and J. D. McEwen. How Isotropic is the Universe? *Physical Review Letters*, 117(13):131302, September 2016. doi: 10.1103/PhysRevLett.117.131302.
- K. K. Schaffer et al. The first public release of south pole telescope data: Maps of a 95 $\hat{\text{A}}$  deg<sup>2</sup> field from 2008 observations. *The Astrophysical Journal*, 743(1):90, 2011. URL <http://stacks.iop.org/0004-637X/743/i=1/a=90>.
- R. J. Scherrer and E. Bertschinger. Statistics of primordial density perturbations from discrete seed masses. *ApJ*, 381:349–360, November 1991. doi: 10.1086/170658.
- R. J. Scherrer and D. H. Weinberg. Constraints on the Effects of Locally Biased Galaxy Formation. *ApJ*, 504:607–611, September 1998. doi: 10.1086/306113.
- Fabian Schmidt, Alberto Vallinotto, Emiliano Sefusatti, and Scott Dodelson. Weak lensing effects on the galaxy three-point correlation function. *Phys. Rev. D*, 78:043513, Aug 2008. doi: 10.1103/PhysRevD.78.043513. URL <https://link.aps.org/doi/10.1103/PhysRevD.78.043513>.
- M. Schmittfull, T. Baldauf, and U. Seljak. Near optimal bispectrum estimators for large-scale structure. *Phys. Rev. D*, 91(4):043530, February 2015. doi: 10.1103/PhysRevD.91.043530.
- M. M. Schmittfull et al. Fast estimation of gravitational and primordial bispectra in large scale structures. *Phys. Rev. D*, 88(6):063512, September 2013. doi: 10.1103/PhysRevD.88.063512.
- P. Schneider, M. Kilbinger, and M. Lombardi. The three-point correlation function of cosmic shear. II. Relation to the bispectrum of the projected mass density and generalized third-order aperture measures. *A&A*, 431:9–25, February 2005. doi: 10.1051/0004-6361:20034217.
- R. L. Schuhmann, B. Joachimi, and H. V. Peiris. Gaussianization for fast and accurate inference from cosmological data. *MNRAS*, 459:1916–1928, June 2016. doi: 10.1093/mnras/stw738.
- Robert L. Schuhmann, Benjamin Joachimi, and Hiranya V. Peiris. Gaussianization for fast and accurate inference from cosmological data. *Mon. Not. Roy. Astron. Soc.*, 459(2):1916–1928, 2016. doi: 10.1093/mnras/stw738.
- R. Scoccimarro. The Bispectrum: From Theory to Observations. *ApJ*, 544:597–615, December 2000. doi: 10.1086/317248.
- R. Scoccimarro and R. K. Sheth. PTHALOS: a fast method for generating mock galaxy distributions. *MNRAS*, 329:629–640, January 2002. doi: 10.1046/j.1365-8711.2002.04999.x.

- R. Scoccimarro, E. Sefusatti, and M. Zaldarriaga. Probing primordial non-Gaussianity with large-scale structure. *Phys. Rev. D*, 69(10):103513, May 2004. doi: 10.1103/PhysRevD.69.103513.
- R. Scoccimarro et al. Nonlinear Evolution of the Bispectrum of Cosmological Perturbations. *ApJ*, 496:586–604, March 1998. doi: 10.1086/305399.
- Roman Scoccimarro, Stephane Colombi, James N. Fry, Joshua A. Frieman, Eric Hivon, and Adrian Melott. Nonlinear evolution of the bispectrum of cosmological perturbations. *Astrophys. J.*, 496:586, 1998. doi: 10.1086/305399.
- Roman Scoccimarro, H. M. P. Couchman, and Joshua A. Frieman. The Bispectrum as a Signature of Gravitational Instability in Redshift-Space. *Astrophys. J.*, 517:531–540, 1999. doi: 10.1086/307220.
- D. Scott, J. Silk, and M. White. From Microwave Anisotropies to Cosmology. *Science*, 268:829–835, May 1995. doi: 10.1126/science.268.5212.829.
- E. Sefusatti, M. Crocce, S. Pueblas, and R. Scoccimarro. Cosmology and the bispectrum. *Phys. Rev. D*, 74(2):023522, July 2006. doi: 10.1103/PhysRevD.74.023522.
- Emiliano Sefusatti et al. Cosmology and the Bispectrum. *Phys. Rev.*, D74:023522, 2006. doi: 10.1103/PhysRevD.74.023522.
- E. Sellentin and A. F. Heavens. Parameter inference with estimated covariance matrices. *MNRAS*, 456:L132–L136, February 2016. doi: 10.1093/mnras/llv190.
- R. K. Sheth, K. C. Chan, and R. Scoccimarro. Nonlocal Lagrangian bias. *Phys. Rev. D*, 87(8):083002, April 2013. doi: 10.1103/PhysRevD.87.083002.
- Y. Shtanov, J. Traschen, and R. Brandenberger. Universe reheating after inflation. *Phys. Rev. D*, 51:5438–5455, May 1995. doi: 10.1103/PhysRevD.51.5438. URL <https://link.aps.org/doi/10.1103/PhysRevD.51.5438>.
- P. Simon et al. CFHTLenS: a Gaussian likelihood is a sufficient approximation for a cosmological analysis of third-order cosmic shear statistics. *MNRAS*, 449:1505–1525, May 2015. doi: 10.1093/mnras/stv339.
- P. and others Simon. CFHTLenS: higher order galaxy-mass correlations probed by galaxy-galaxy-galaxy lensing. *MNRAS*, 430:2476–2498, April 2013. doi: 10.1093/mnras/stt069.
- Z. Slepian et al. The large-scale 3-point correlation function of the SDSS BOSS DR12 CMASS galaxies. *ArXiv e-prints*, December 2015.
- Z. Slepian et al. The large-scale three-point correlation function of the SDSS BOSS DR12 CMASS galaxies. *MNRAS*, 468:1070–1083, June 2017a. doi: 10.1093/mnras/stw3234.
- Z. Slepian et al. 469:1738–1751, August 2017b. doi: 10.1093/mnras/stx488.

- A. Slosar et al. The acoustic peak in the Lyman alpha forest. *JCAP*, 10:019, October 2009. doi: 10.1088/1475-7516/2009/10/019.
- G. F. Smoot et al. Structure in the COBE differential microwave radiometer first-year maps. *ApJ*, 396:L1–L5, September 1992. doi: 10.1086/186504.
- Song, Yong-Seon and Taruya, Atsushi and Oka, Akira. Cosmology with anisotropic galaxy clustering from the combination of power spectrum and bispectrum. *JCAP*, 1508(08):007, 2015. doi: 10.1088/1475-7516/2015/08/007.
- V. and others Springel. First results from the IllustrisTNG simulations: matter and galaxy clustering. *MNRAS*, 475:676–698, March 2018. doi: 10.1093/mnras/stx3304.
- A. A. Starobinsky. Dynamics of phase transition in the new inflationary universe scenario and generation of perturbations. *Physics Letters B*, 117:175–178, November 1982. doi: 10.1016/0370-2693(82)90541-X.
- G. S. Stinson, J. Bailin, H. Couchman, J. Wadsley, S. Shen, S. Nickerson, C. Brook, and T. Quinn. Cosmological galaxy formation simulations using smoothed particle hydrodynamics. *MNRAS*, 408:812–826, October 2010. doi: 10.1111/j.1365-2966.2010.17187.x.
- R. A. Sunyaev and Y. B. Zeldovich. Small-Scale Fluctuations of Relic Radiation. *Astrophysics and Space Science*, 7:3–19, April 1970. doi: 10.1007/BF00653471.
- D. S. Swetz et al. Overview of the Atacama Cosmology Telescope: Receiver, Instrumentation, and Telescope Systems. *ApJS*, 194:41, June 2011. doi: 10.1088/0067-0049/194/2/41.
- M. Takada and B. Jain. Cosmological parameters from lensing power spectrum and bispectrum tomography. *MNRAS*, 348:897–915, March 2004. doi: 10.1111/j.1365-2966.2004.07410.x.
- M. Takada et al. Extragalactic science, cosmology, and Galactic archaeology with the Subaru Prime Focus Spectrograph. *PASJ*, 66:R1, February 2014. doi: 10.1093/pasj/pst019.
- A. Taruya, K. Koyama, and T. Matsubara. Signature of primordial non-Gaussianity on the matter power spectrum. *Phys. Rev. D*, 78(12):123534, December 2008. doi: 10.1103/PhysRevD.78.123534.
- A. Taruya, T. Nishimichi, and S. Saito. Baryon acoustic oscillations in 2D: Modeling redshift-space power spectrum from perturbation theory. *Phys. Rev. D*, 82(6):063522, September 2010. doi: 10.1103/PhysRevD.82.063522.
- Atsushi Taruya, Kazuya Koyama, and Jiro Soda. Quasi-nonlinear evolution of stochastic bias. *The Astrophysical Journal*, 510(2):541, 1999. URL <http://stacks.iop.org/0004-637X/510/i=2/a=541>.
- A. Taylor and B. Joachimi. Estimating cosmological parameter covariance. *MNRAS*, 442:2728–2738, August 2014. doi: 10.1093/mnras/stu996.

- M. Tegmark, A. N. Taylor, and A. F. Heavens. Karhunen-Loève Eigenvalue Problems in Cosmology: How Should We Tackle Large Data Sets? *ApJ*, 480: 22–35, May 1997. doi: 10.1086/303939.
- M. Tegmark et al. The Three-Dimensional Power Spectrum of Galaxies from the Sloan Digital Sky Survey. *ApJ*, 606:702–740, May 2004. doi: 10.1086/382125.
- Max Tegmark, Andy Taylor, and Alan Heavens. Karhunen-Loeve eigenvalue problems in cosmology: How should we tackle large data sets? *Astrophys. J.*, 480:22, 1997. doi: 10.1086/303939.
- M. Tellarini, A. J. Ross, G. Tasinato, and D. Wands. Galaxy bispectrum, primordial non-Gaussianity and redshift space distortions. *JCAP*, 6:014, June 2016. doi: 10.1088/1475-7516/2016/06/014.
- The Dark Energy Survey Collaboration. The Dark Energy Survey. *ArXiv Astrophysics e-prints*, October 2005.
- J. L. and others Tinker. The Large-scale Bias of Dark Matter Halos: Numerical Calibration and Model Tests. *ApJ*, 724:878–886, December 2010. doi: 10.1088/0004-637X/724/2/878.
- D. Toussaint and F. Wilczek. Elementary examples of baryon number generation. *Physics Letters B*, 81:238–240, February 1979. doi: 10.1016/0370-2693(79)90532-X.
- Licia Verde et al. Large scale bias in the universe. 2. Redshift space bispectrum. *Mon. Not. Roy. Astron. Soc.*, 300:747–756, 1998. doi: 10.1046/j.1365-8711.1998.01937.x.
- Licia Verde et al. Large-scale structure, the cosmic microwave background and primordial non-gaussianity. *Monthly Notices of the Royal Astronomical Society*, 313(1):141, 2000. doi: 10.1046/j.1365-8711.2000.03191.x. URL [+http://dx.doi.org/10.1046/j.1365-8711.2000.03191.x](http://dx.doi.org/10.1046/j.1365-8711.2000.03191.x).
- D. H. Weinberg, M. J. Mortonson, D. J. Eisenstein, C. Hirata, A. G. Riess, and E. Rozo. Observational probes of cosmic acceleration. *Physics Reports*, 530: 87–255, September 2013. doi: 10.1016/j.physrep.2013.05.001.
- D. H. Weinberg et al. Cosmology with the Lyman-alpha Forest. *ArXiv Astrophysics e-prints*, October 1998.
- S. Weinberg. Cosmological production of baryons. *Physical Review Letters*, 42: 850–853, March 1979. doi: 10.1103/PhysRevLett.42.850.
- M. and others White. The Clustering of Massive Galaxies at  $z \sim 0.5$  from the First Semester of BOSS Data. *ApJ*, 728:126, February 2011. doi: 10.1088/0004-637X/728/2/126.
- S. D. M. White and M. J. Rees. Core condensation in heavy halos - A two-stage theory for galaxy formation and clustering. *MNRAS*, 183:341–358, May 1978. doi: 10.1093/mnras/183.3.341.

- R. W. Wilson and A. A. Penzias. Measurement of the Flux of Five Sources at 4080 Mc/sec. *AJ*, 70:697, November 1965. doi: 10.1086/109803.
- E. L. Wright, S. S. Meyer, C. L. Bennett, N. W. Boggess, E. S. Cheng, M. G. Hauser, A. Kogut, C. Lineweaver, J. C. Mather, G. F. Smoot, R. Weiss, S. Gulkis, G. Hinshaw, M. Janssen, T. Kelsall, P. M. Lubin, S. H. Moseley, Jr., T. L. Murdock, R. A. Shafer, R. F. Silverberg, and D. T. Wilkinson. Interpretation of the cosmic microwave background radiation anisotropy detected by the COBE Differential Microwave Radiometer. *ApJ*, 396:L13–L18, September 1992. doi: 10.1086/186506.
- A. Zablocki and S. Dodelson. Extreme data compression for the CMB. *Phys. Rev. D*, 93(8):083525, April 2016. doi: 10.1103/PhysRevD.93.083525.
- Alan Zablocki and Scott Dodelson. Extreme data compression for the CMB. *Phys. Rev.*, D93(8):083525, 2016. doi: 10.1103/PhysRevD.93.083525.
- Y. B. Zel'dovich. Special Issue: the Cosmological Constant and the Theory of Elementary Particles. *Soviet Physics Uspekhi*, 11:381–393, March 1968. doi: 10.1070/PU1968v011n03ABEH003927.
- C. Zunckel, D. Huterer, and G. D. Starkman. Testing the statistical isotropy of large scale structure with multipole vectors. *Phys. Rev. D*, 84(4):043005, August 2011. doi: 10.1103/PhysRevD.84.043005.
- F. Zwicky. Die Rotverschiebung von extragalaktischen Nebeln. *Helvetica Physica Acta*, 6:110–127, 1933.

“And the ship went out into the High Sea and passed into the West, until at last on a night of rain Frodo smelled a sweet fragrance on the air and heard the sound of singing that came over the water. And then it seemed to him that as in his dream in the house of Bombadil, the grey rain-curtain turned all to silver glass and was rolled back, and he beheld white shores and beyond them a far green country under a swift sunrise.”

— J.R.R. Tolkien, *The Return of the King*