

# **UNIVERSITÀ DEGLI STUDI DI PADOVA**

**Dipartimento di Fisica e Astronomia "Galileo Galilei"**

**Master Degree in Physics**

**Final Dissertation**

## **The Integrated Angular Bispectrum of Weak Lensing**

**Thesis supervisor**  
**Prof. Michele Liguori**

**Thesis co-supervisor**  
**Dr. Gabriel Jung**

**Candidate**  
**Leonardo Rebeschini**

**Academic Year 2021/2022**



**T**RY not to become a man of success.  
Rather become a man of value.

---

A. Einstein



# Contents

<b>1</b>	<b>Cosmological Framework</b>	<b>5</b>
1.1	Einstein's Equations and FLRW Metric . . . . .	6
1.2	Friedmann Equations . . . . .	7
1.3	The matter power spectrum . . . . .	9
<b>2</b>	<b>Gravitational Lensing</b>	<b>13</b>
2.1	Introduction . . . . .	13
2.2	History . . . . .	13
2.3	Gravitational Optics . . . . .	14
2.4	More General Systems . . . . .	18
2.5	Applications . . . . .	20
<b>3</b>	<b>The Shape of Non-Gaussianity</b>	<b>23</b>
3.1	Introduction . . . . .	23
3.2	Gaussianity and the power spectrum . . . . .	24
3.3	Bispectrum . . . . .	24
3.3.1	Equilateral and flattened (folded) triangles . . . . .	25
3.3.2	Squeezed triangles . . . . .	25
<b>4</b>	<b>Weak Lensing Statistics</b>	<b>33</b>
4.1	Position-Dependent Power Spectrum from Convergence Maps . . . . .	33
4.1.1	Flat-Sky Treatment . . . . .	33
4.1.2	All-Sky Formulation . . . . .	36
4.2	The integrated bispectrum estimator . . . . .	41
4.2.1	Definition . . . . .	41
4.2.2	Mask . . . . .	42
4.2.3	Linear correction . . . . .	42
4.3	Covariance estimation . . . . .	42
<b>5</b>	<b>Computational Implementation</b>	<b>45</b>
5.1	Maps and Methods . . . . .	45
5.1.1	Weak lensing maps . . . . .	45
5.1.2	N-body simulations and Ray-tracing . . . . .	46
5.1.3	Healpix map's structure . . . . .	48
5.1.4	Patches and masks . . . . .	48
5.1.5	Rotations . . . . .	52
5.1.6	Flat sky projections . . . . .	53

5.2	Power spectrum estimation . . . . .	55
5.2.1	Full-sky . . . . .	56
5.2.2	Flat-sky . . . . .	56
5.2.3	Power spectra results . . . . .	58
5.3	Monopole term estimation . . . . .	60
5.3.1	Full-sky . . . . .	60
5.3.2	Flat-sky . . . . .	60
5.4	Integrated bispectrum estimator: 2 pipelines . . . . .	61
5.4.1	Slow pipeline . . . . .	61
5.4.2	Fast pipeline . . . . .	62
5.4.3	Computational time summary . . . . .	62
<b>6</b>	<b>Results from All-sky Simulations</b>	<b>63</b>
6.1	Introduction . . . . .	63
6.2	Signal to noise ratio and Optimality . . . . .	63
6.3	Observed integrated bispectrum results . . . . .	64
6.3.1	$IB_{\ell}^{obs}$ compared with theoretical signal . . . . .	64
6.3.2	$IB_{\ell}^{obs}$ with more localized patches . . . . .	65
6.4	Timing summary . . . . .	67
<b>7</b>	<b>Conclusions</b>	<b>69</b>
7.1	This work . . . . .	69
7.2	Towards Covariance matrix . . . . .	70
<b>A</b>	<b>Handmade flat-sky power spectrum estimator - code</b>	<b>73</b>
<b>B</b>	<b>Needlets Patches</b>	<b>77</b>
	<b>Bibliography</b>	<b>80</b>

# Introduction

On May 29, 1919, during a solar eclipse, the deflection of light rays of stars due to the Sun's gravitational field was measured [12], marking the first successful test of the theory of general relativity (GR; Einstein 1916). Then, the first discovery of extra-galactic gravitational lensing was obtained in 1979, with the detection of a doubly-imaged quasar lensed by a galaxy [53]. Lensing distortions have been known since 1987 with the observation of giant arcs – strongly distorted galaxies behind massive galaxy clusters [48]. Three years later in 1990, weak gravitational lensing was detected for the first time as statistical tangential alignments of galaxies behind massive clusters [51]. It took another 10 years until, in 2000, coherent galaxy distortions were measured in blind fields, showing the existence of weak gravitational lensing by the large-scale structure, or cosmic shear [21], [52], [54]. And so, nearly 100 years after its first measurement, the technique of gravitational lensing has evolved into a powerful tool for challenging GR on cosmological scales.

All observed light from distant galaxies is subject to gravitational lensing. This is because light rays propagate through a universe that is inhomogeneous due to the ubiquitous density fluctuations at large scales. These fluctuations create a tidal gravitational field that causes light bundles to be deflected differentially. As a result, images of light-emitting galaxies that we observe are distorted. The direction and amount of distortion is directly related to the size and shape of the matter distribution projected along the line of sight. The deformation of high-redshift galaxy images in random lines of sight therefore provides a measure of the large-scale structure (LSS) properties, which consists of a network of voids, filaments, and halos. The larger the amplitude of the inhomogeneity of this cosmic web is, the larger the deformations are.

The typical distortions of high-redshift galaxies by the cosmic web are on the order of a few percent, much smaller than the width of the intrinsic shape and size distribution. Thus, for an individual galaxy, the lensing effect is not detectable, placing cosmic shear into the regime of *weak gravitational lensing*. The presence of a tidal field acting as a gravitational lens results in a coherent alignment of galaxy image orientations. This alignment can be measured statistically as a correlation between galaxy shapes.

Cosmic shear is a very versatile probe of the LSS. It measures the clustering of the LSS from the highly non-linear, non-Gaussian sub-megaparsec (Mpc) regime, out to very large, linear scales of more than a hundred Mpc. By measuring galaxy shape correlations between different redshifts, the evolution of the LSS can be traced, enabling us to detect the effect of dark energy on the growth of structure. Together with the ability to measure the geometry of the Universe, cosmic shear can potentially distinguish between dark energy and modified gravity theories [16]. Since gravitational lensing is not sensitive to the dynamical state of the intervening masses, it yields a direct measure of the total matter, dark plus luminous. By adding information about the distribution of galaxies, cosmic shear can shed light on the

complex relationship between galaxies and dark matter.

Since the first detection over a few square degrees of sky area a decade and a half ago, cosmic shear has matured into an important tool for cosmology. Current surveys span hundreds of square degrees, and thousands of square degrees more to be observed in the near future. Cosmic shear is a major science driver of large imaging surveys from both ground and space. The traditional statistics to use for comparison of theory with data are two-point statistics - correlation functions or power spectra. These can be sufficient statistics provided that the field under consideration is a random Gaussian field, but the non-linear growth of structure by gravitational instability means that the field is non-Gaussian, and the two-point statistics do not capture all of the cosmological information contained in the field. Thus, to fully exploit the data requires going beyond the power spectrum to probe also the non-Gaussian properties of the field. There are several ways to do this, from sophisticated Bayesian forward-modelling techniques, which incorporate a gravity model [38] and which apply the likelihood at the field level, likelihoodfree inference [17], or by analysing higher-order summary statistics. The addition of the bispectrum to the power spectrum can lead to significant reduction in errors [49] [23] [42] and better control of systematics [40].

There are alternative approaches to including non-Gaussian information, such as with skew-spectra [33] [31] and Minkowski functionals [32], both of which have been developed for weak lensing.

For higher-order statistics, there are several challenges, one of which is the very large number of three-point functions (bispectra, in harmonic space) that can be considered. In addition to this is the formidable challenge of making accurate theoretical predictions for these statistics [3] and to compute their sampling distribution. In order to address the first complication, we consider the integrated bispectrum, which involves computing only power spectra, in patches on the sky, but which probes the squeezed limit of the bispectrum. This approach has been first proposed in the context of Large Scale Structure [7] and has a wide range of applications like galaxy clustering, 21 cm and weak lensing studies.

Here we consider its extension to 2D random fields on the sphere developed in [18] for CMB non-Gaussianity (NG) analyses. We consider the spin-0 convergence field, rather than the spin-2 cosmic shear field, as this field is easier to deal with, whilst still incorporating many of the same challenges as the cosmic shear field, which is the usual field studied with weak lensing. Cosmic shear and the inclusion of NG contributions in the covariance will be the subjects of future studies.

## Goal

Build a pipeline for the integrated angular bispectrum estimator implementing the flat-sky approximation in the position dependent power spectrum estimator, in order to have a high enough speed to apply it to many thousands of simulations in order to extract the covariance matrix.

## Thesis structure

In **Ch.1** we introduce the cosmological framework, recalling Einstein's equations, FLRW metric, Friedmann equations, the matter power spectrum and its time evolution. In **Ch.2** we

introduce gravitational lensing formalism and its various applications. In **Ch.3** we discuss the shape of non-Gaussianity analyzing the effect in real space of different bispectrum shapes, in particular focusing on squeezed configurations; in **Ch.4** we introduce the statistical tools needed to define the integrated angular bispectrum estimator and the theory behind its descriptions, with both a flat-sky and a full-sky treatment. In **Ch.5** we discuss the computational implementation of flat-sky approximation within the pipeline. In **Ch.6** we show results from all sky simulations and in **Ch.7** we draw the conclusions of this work.



# Chapter 1

## Cosmological Framework

When the scientific interest towards the study of the Universe brought the necessity of a scientific theory able to describe the whole Universe, it became clear that in order to pursue this goal a guiding principle was needed; then, at the begins of the 20th century, the Cosmological Principle has been introduced. This principle claims that on large enough scales the Universe can be considered homogeneous and isotropic from the point of view of a comoving observer<sup>1</sup>. This means that on such large scales (more precisely, on length scales larger than hundreds of Mpc) there is not a preferred direction or orientation in the space. Although at first this was just a principle introduced to simplify scientific theories, later on some observations confirmed this assumption [47]. Nowadays it is thought that the model that best describes a Universe with these components is the  $\Lambda$  Cold Dark Matter (CDM) model or  $\Lambda$ CDM model. The main components of the present day Universe are:

- Dark Energy ( $\sim 68\%$ ), it is responsible for the accelerated expansion of the Universe;
- Dark Matter ( $\sim 27\%$ ), it is composed of non-interacting massive particles that have an equation of state dust-like:  $p = w\rho$  with  $w = 0$ , whose nature has not been clarified yet;
- Baryonic Matter ( $\sim 5\%$ ), it is the ordinary matter produced during the Big Bang Nucleosynthesis.

The  $\Lambda$ CDM model is a six free parameters model, this means that six parameters are needed in order to parametrise the  $\Lambda$ CDM model. According to the Planck mission [37], combining Planck temperature with Planck lensing data, the values of these parameters are written in table 1.1:

---

<sup>1</sup>A comoving observer is an observer that moves with the cosmic flow.

$H_0 [c]$	$67.66 \pm 0.42$
$\tau$	$0.0561 \pm 0.0071$
$n_s$	$0.9665 \pm 0.0038$
$\ln(10^{10} A_s)$	$3.047 \pm 0.014$
$\Omega_b h^2$	$0.02242 \pm 0.00014$
$\Omega_m h^2$	$0.14240 \pm 0.00087$

Table 1.1: Cosmological parameters associated to the  $\Lambda$ CDM model with their values from the Planck mission.

## 1.1 Einstein's Equations and FLRW Metric

The model introduced to describe the Universe is based on Einstein's equations (EE) of general relativity (GR):

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 8\pi GT_{\mu\nu} \quad (1.1)$$

where  $G_{\mu\nu}$  is the Einstein tensor describing the geometrical part in the left-hand-side, with  $R_{\mu\nu}$  and  $R$  being the Ricci tensor and the Ricci scalar, derived from the metric tensor  $g_{\mu\nu}$ . EE can be derived from the variation with respect to the metric of the total action  $S_{TOT} = S_{HE} + S_m$ , where

$$S_{HE} = \frac{1}{16\pi G} \int d^4x \sqrt{-g} R \quad (1.2)$$

is the Hilbert-Einstein action that accounts for gravity, while

$$S_m = \int d^4x \sqrt{-g} \mathcal{L}_m \quad (1.3)$$

is the action associated to all the other particles (scalar fields, fermions, gauge bosons, ...). In particular, from the variation  $\frac{\delta S_{HE}}{\delta g^{\mu\nu}}$  it derives the geometric LHS of Eq.(1.1) while the stress-energy tensor  $T_{\mu\nu}$  on the RHS is defined as

$$T_{\mu\nu} = -\frac{2}{\sqrt{-g}} \frac{\delta S_m}{\delta g^{\mu\nu}}. \quad (1.4)$$

This tensor can be computed focusing on its global properties and treat it as a perfect fluid; this lead to the very well known form (true for a generic reference frame)

$$T_{\mu\nu} = u_\mu u_\nu (\rho + P) + P g_{\mu\nu} \quad (1.5)$$

where  $u_\mu$  is the 4-velocity of the fluid in its reference frame.

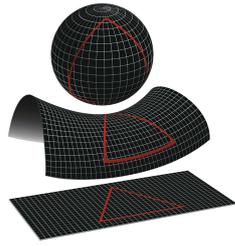
There are other two important relations that must be taken into account every time that Einstein's equations enter in the game; these are the Bianchi identity  $D_\mu G^{\mu\nu} = 0$  and the continuity equation  $D_\mu T^{\mu\nu} = 0$ , which indeed are consequence of one another, given the Einstein's equations.

It is also important to remember that eq.(1.1) can be either used to find metrics that satisfy some symmetries or, using some specific spacetime and constraints, to find some system. If we proceed in first way, looking for a metric that describes an expanding universe with

curvature  $\kappa$  and which is symmetric under rotation and spatial translation, assuming the Cosmological Principle: isotropy and spatially homogeneity<sup>2</sup>, we find the FLRW metric

$$ds^2 = -c^2 dt^2 + a^2(t) \left[ \frac{dr^2}{1 - \kappa r^2} + r^2 d\Omega^2 \right], \quad (1.6)$$

where  $t$  is the cosmic time, whereas the time coordinate of an observer comoving with the cosmic fluid; the scale factor  $a(t)$  spans the expansion of the universe;  $d\Omega^2 = d\theta^2 + \sin^2\theta d\phi^2$  is the infinitesimal solid angle and the curvature  $\kappa^3$  can be either positive (open universe), null (flat universe) or negative (closed universe)<sup>4</sup>. Usually one parametrizes  $\kappa$  as



$$\kappa = \begin{cases} -1 & \text{open} \\ 0 & \text{flat} \\ +1 & \text{close} \end{cases} \quad (1.7)$$

Figure 1.1: Close, open and flat universe representations.

It is important to underline that the set  $(r, \theta, \phi)$  is made of comoving coordinates, i.e. they do not depend on the evolution and remain constant, so that, once they are multiplied by the scale factor, the physical coordinates are recovered<sup>5</sup>. From the symmetry properties of eq. (1.6), it can be found the very easy form for the stress-energy tensor

$$T_{\mu\nu} = \text{diag}[\rho(t), P(t), P(t), P(t)], \quad (1.9)$$

where  $\rho(t)$  and  $P(t)$  are respectively the energy density and the isotropic pressure of the fluid, which do not depend on  $\vec{x}$  because of isotropy and homogeneity.

## 1.2 Friedmann Equations

The dynamics of the expanding universe can be explicitly seen by unfolding the Einstein's equations

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 8\pi GT_{\mu\nu} - \Lambda g_{\mu\nu}, \quad (1.10)$$

<sup>2</sup>We do not ask for the time translation invariance because we want a universe able to change throughout time.

<sup>3</sup>Which is indicated with the same character we will use for the convergence field, here the curvature is a constant while the other is a field.

<sup>4</sup>One can show that these three cases represent respectively the geometry of an hyperboloid, a plane and a sphere. It is intuitive that in all of these the "landscape" around a specific point is completely isotropic and it is the same in every point one chooses, hence they indeed are isotropic and homogeneous.

<sup>5</sup>Considering the comoving coordinates, one can show that the coordinate separation between two points remain constant in time. Then the spatial slices of the metric in the parenthesis gets rescaled by the scale factor, in such a way that for any distance  $\lambda$ , it holds

$$\lambda_{\text{physical}} = a(t) \cdot \lambda_{\text{comoving}}. \quad (1.8)$$

where we have added for the sake of completeness  $\Lambda$ , the cosmological constant. In eq. (1.10) appears the Ricci tensor  $R_{\mu\nu}$ , which is the contraction of the first and third index of the Riemann tensor equal to

$$R^\lambda_{\sigma\mu\nu} = \partial_\mu \Gamma^\lambda_{\sigma\nu} - \partial_\nu \Gamma^\lambda_{\sigma\mu} + \Gamma^\lambda_{\mu\rho} \Gamma^\rho_{\nu\sigma} - \Gamma^\lambda_{\nu\rho} \Gamma^\rho_{\mu\sigma} ; \quad (1.11)$$

the Ricci scalar  $R$ , which is the contraction of the two remaining indices of the Ricci tensor, and the Christoffel symbols  $\Gamma^\mu_{\nu\lambda}$ , defined as

$$\Gamma^\mu_{\nu\lambda} = \frac{1}{2} g^{\mu\rho} (\partial_\lambda g_{\rho\nu} + \partial_\nu g_{\rho\lambda} - \partial_\rho g_{\nu\lambda}) . \quad (1.12)$$

Thus, one can find the Friedmann equations by writing explicitly the  $_{00}$  and  $_{ij}$  components of the Einstein's equations and by exploiting the conservation laws  $\nabla_\mu T^{\mu\nu} = 0$  (in this specific case the  $\nabla$  refers to the covariant derivative) finding

$$H^2 + \frac{\kappa}{a^2} = \frac{8\pi G}{3} \rho + \frac{\Lambda}{3} , \quad (1.13)$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3} (\rho + 3P) + \frac{\Lambda}{3} , \quad (1.14)$$

$$\dot{\rho} + 3H(\rho + P) = 0 , \quad (1.15)$$

where the dot indicates the derivative w.r.t. cosmic time and  $H$  is the Hubble rate defined as  $H = \dot{a}/a$ . These equation are indeed dependent from one another through the Bianchi identities so that only two of them are independent.

To close the system, one needs an extra relation linking  $P(t)$  and  $\rho(t)$ , named equation of state  $P(\rho)$ . The simplest choice is

$$P = \omega\rho \quad \text{with} \quad \omega = \text{constant} , \quad (1.16)$$

where  $\omega$  depend on the energy content considered

$$\omega = \begin{cases} 0 & \text{dust or pressureless matter} \\ \frac{1}{3} & \text{radiation} \\ -1 & \Lambda \text{ cosmological constant} \end{cases} . \quad (1.17)$$

Solving this system of equations, one finds the explicit expression of the scale factor  $a(t)$ , which in FLRW has the usual solution with a singularity back in time

$$a(t) = a_* \left( \frac{t}{t_*} \right)^{\frac{2}{3(\omega+1)}} , \quad (1.18)$$

where  $*$  indicates a reference scale<sup>6</sup>.

For what regards the energy density  $\rho$ :

$$\rho = \rho_* \left( \frac{a}{a_*} \right)^{-3(1+\omega)} \quad \Rightarrow \quad \rho \propto \begin{cases} a^{-3} & \omega = 0 \\ a^{-4} & \omega = \frac{1}{3} \\ \text{constant} & \omega = -1 \end{cases} . \quad (1.19)$$

---

<sup>6</sup>From this relation one can appreciate the fact that only ratios of scale factors are physical, assuming a spatially flat universe. In fact, the coordinate can always be rescaled by a constant without any physical consequence.

Whereas in the case of  $\kappa \neq 0$  the normalization of  $a(t)$  becomes physical, since in the Friedmann equations the term  $\propto \kappa$  cannot be rescaled freely.

These dependency can be intuitively derived if one thinks at how energy gets diluted with the expansion of the universe: in the case of pressureless matter (e.g. we can think of a bunch of protons), the energy density will scale as volume<sup>-1</sup>, so  $a^{-3}$ , since the number of particles in that volume will not change; for radiation, the energy gets an extra  $a^{-1}$  factor because the wave length of radiation will also be diluted by expansion; finally  $\Lambda$  is by definition a constant, so even if the universe expands, its value will not change.

Keeping in mind this scaling relation and going back in time in the evolution history of the Universe, it is possible to identify three main eras through the correspondent values of the scale factor:

$$\begin{cases} \text{Dark energy domination:} & a_\Lambda < a < a_0, \\ \text{Matter domination:} & a_{eq} < a < a_\Lambda, \\ \text{Radiation domination:} & 0 < a < a_{eq}. \end{cases} \quad (1.20)$$

Observations seem to show that the dominance of the cosmological constant starts at  $z \sim 1$ .

## 1.3 The matter power spectrum

In order to explain the origin and the evolution of the cosmic structures, we have to take into account primordial matter density perturbations  $\delta(x)$ . These perturbations came from quantum fluctuations of the scalar field called inflaton that filled the Universe in the first stages of its evolution. These density fluctuations, due to gravitational instability, became larger and larger until they generate the cosmic structures we see today, as the Universe during the inflationary period underwent a remarkable expansion of about 60-efolds [24]. They can be considered quasi-Gaussian random fields and, as a consequence, can be treated from a statistical point of view as if they were exactly Gaussian; this means that these fields are defined by their mean  $\langle \delta(x) \rangle$  that is zero by definition and their covariance or two-point correlation function  $\xi(r) = \langle \delta(x)\delta(x+r) \rangle$ . This field lives in a Robertson-Walker Universe, that geometrically is a spacetime with the three-dimensional spatial component homogeneous and isotropic, this property translates in a definition of a density fluctuation  $\delta(x)$  homogeneous and isotropic itself with the two-point correlation function that depends only on the  $r$  coordinate with  $r = |x_2 - x_1|$ . Moving to Fourier space<sup>7</sup>:

$$\delta(x) = \frac{1}{(2\pi)^3} \int d^3k e^{i\mathbf{k}\cdot\mathbf{x}} \delta_{\mathbf{k}} \quad (1.21)$$

and the two-point correlation function will be:

$$\xi(r) = \langle \delta(\mathbf{x})\delta(\mathbf{x} + \mathbf{r}) \rangle = \frac{1}{(2\pi)^6} \int d^3k_1 \int d^3k_2 e^{i(\mathbf{k}_1\cdot\mathbf{x} + \mathbf{k}_2\cdot(\mathbf{x}+\mathbf{r}))} \langle \delta_{\mathbf{k}_1} \delta_{\mathbf{k}_2} \rangle. \quad (1.22)$$

If the random field is homogeneous and isotropic, also its statistical ensemble will have the same properties, this means that it is possible to write:

<sup>7</sup>Later  $\delta$  will indicate the angular size of the cartesian projection, do not confuse it with the mass fluctuation field.

$$\langle \delta_{\mathbf{k}_1} \delta_{\mathbf{k}_2} \rangle = (2\pi)^3 \delta^{(3)}(\mathbf{k}_1 + \mathbf{k}_2) P(k), \quad (1.23)$$

where the Dirac delta  $\delta^{(3)}(\mathbf{k}_1 + \mathbf{k}_2)$  ensures translation invariance and the function  $P(k)$  called Power Spectrum guarantees rotational invariance.

It is possible to show that the power spectrum of our statistical system is simply the Fourier transform of the two-point correlation function

Moving from cartesian coordinates to spherical coordinates in Fourier space the explicit expression of the two-point correlation function will be function of  $k, \theta_k$  and  $\phi_k$ .

After an integration over the angles it will be:

$$\xi(r) = \frac{1}{2\pi^2} \int dk k^3 P(k) \frac{\sin(kr)}{kr}. \quad (1.24)$$

Now, we focus on the linear regime and study the evolution of density perturbations at this scale.

The equations that describe the evolution of a fluid inside a Universe with a non-static background (remember that Universe evolves) can be written as:

$$\frac{\partial \rho}{\partial t} + 3H\rho + \frac{1}{a} \nabla \cdot (\rho \mathbf{v}) = 0 \quad (1.25)$$

$$\begin{aligned} \xi(r) &= \frac{1}{(2\pi)^3} \int d^3 k e^{i\mathbf{k} \cdot \mathbf{r}} P(k), \\ \xi(r) &= \langle \delta(\mathbf{x}) \delta(\mathbf{x} + \mathbf{r}) \rangle = \left\langle \frac{1}{(2\pi)^6} \int d^3 k_1 \int d^3 k_2 e^{i\mathbf{k}_1 \cdot \mathbf{x}} e^{i\mathbf{k}_2 \cdot (\mathbf{x} + \mathbf{r})} \delta_{\mathbf{k}_1} \delta_{\mathbf{k}_2} \right\rangle \\ &= \frac{1}{(2\pi)^6} \int d^3 k_1 \int d^3 k_2 e^{i\mathbf{k}_1 \cdot \mathbf{x}} e^{i\mathbf{k}_2 \cdot (\mathbf{x} + \mathbf{r})} \delta_{\mathbf{k}_1} \delta_{\mathbf{k}_2} (2\pi)^3 \delta^{(3)}(\mathbf{k}_1 + \mathbf{k}_2) P(k) \\ &= \frac{1}{(2\pi)^3} \int d^3 k_1 e^{-i\mathbf{k}_1 \cdot \mathbf{r}} P(k) = \frac{1}{(2\pi)^3} \int d^3 k_1 e^{i\mathbf{k}_1 \cdot \mathbf{r}} P(k) \end{aligned} \quad (1.26)$$

$$\begin{aligned} \frac{\partial \mathbf{v}}{\partial t} + H\mathbf{v} + \frac{1}{a} (\mathbf{v} \cdot \nabla_{\mathbf{x}}) \mathbf{v} &= -\frac{1}{a} \nabla_{\mathbf{x}} \Phi \\ \nabla_{\mathbf{x}}^2 \Phi &= 4\pi G a^2 \delta \rho \end{aligned} \quad (1.27)$$

Eq. (1.25) is the continuity equation, eq. (1.26) is the Euler equation and eq. (1.27) is Poisson's equation.  $H(t)$  is the Hubble constant,  $a(t)$  the scale factor and  $\Phi$  is the Newtonian gravitational potential and  $\mathbf{v}$  is the peculiar velocity. We want to write these equation for the evolution of the initial density fluctuation  $\delta$  in the linear regime, this means that  $\delta \ll 1$  and we can linearise our system,

$$\delta \rho = \rho - \bar{\rho} \quad \text{and} \quad \rho = (1 + \delta) \bar{\rho} \quad (1.28)$$

we obtain:

$$\begin{aligned} \frac{\partial \delta}{\partial t} + \frac{1}{a} \nabla_{\mathbf{x}} \cdot \mathbf{v} &= 0 \\ \frac{\partial \mathbf{v}}{\partial t} + H\mathbf{v} &= -\frac{1}{a} \nabla_{\mathbf{x}} \Phi \\ \nabla_{\mathbf{x}}^2 \Phi &= 4\pi G a^2 \delta \bar{\rho} \end{aligned} \quad (1.29)$$

### 1.3. The matter power spectrum

---

In order to find a solution to this system of equations we go to Fourier space where our vector fields will be  $\delta(\mathbf{k}, t)$ ,  $\mathbf{v}(\mathbf{k}, t)$  and  $\Phi(\mathbf{k}, t)$ .

Now, combining the three equations we get:

$$\ddot{\delta} + 2\frac{\dot{a}}{a}\dot{\delta} - 4\pi G\delta\bar{\rho} = 0. \quad (1.30)$$

If we are in a matter-dominated Universe, then  $a(t) \propto t^{2/3}$ ,  $H(t) = 2/3t$  and  $\bar{\rho} = 1/6\pi Gt^2$  and it is possible to rewrite the second order differential equation as:

$$\ddot{\delta} + \frac{4}{3t}\dot{\delta} - \frac{2}{3t^2}\delta = 0. \quad (1.31)$$

With a substitution we find two different solutions:

$$\delta \propto \begin{cases} t^{2/3} \propto a(t) \\ t^{-1} \propto H(t) \end{cases} \quad (1.32)$$

In particular we introduce  $D_+(t)$  that is the growing mode factor and  $D_-(t)$  that is the decaying mode factor and we have:

$$\delta(\mathbf{x}, t) = D_+(t)\delta(\mathbf{x}) + D_-(t)\delta(\mathbf{x}). \quad (1.33)$$

Since  $D_- \propto t^{-1}$  it becomes negligible in the early stages of the evolution of the Universe, the interesting term is  $D_+(t) \propto a(t) \propto t^{2/3}$  that describes the growing mode of the initial density perturbation. As these density perturbations are small, the linear regime approach is able to explain their physical evolution; when the perturbations due to gravitational instability becomes larger, the linear regime is no longer the right approach and we have to pass to the non-linear regime in order to explain structure formation 3.2.



# Chapter 2

## Gravitational Lensing

### 2.1 Introduction

Photons travel along null geodesics of the spacetime metric, and the presence of sources of mass/energy in their path induces deflections in the direction of the light bundles. This phenomenon of gravitational light deflection is called Gravitational Lensing and it is well described in the context of General Relativity. Light from distant galaxies is deflected on its way toward the observer through an inhomogeneous Universe and light bundles are differentially distorted by tidal fields of the Large Scale Structure components. The interesting property of the gravitational lensing is that it can be considered as a probe of the total matter (dark matter and visible matter) content of the Universe without having to take into account dark matter probes [45].

### 2.2 History

The idea that gravity might bend light did not originate with Einstein, it was already considered by Newton in his publication about *Optics*, and by Laplace. Following Newton, light rays should have experienced the gravity force.

In 1804 Soldner attempted to apply the Newtonian gravity to calculate the deflection of the light ray by the Sun, under the assumption that light is composed by massive particles. He found that, according to the classical theory of gravity, the deflection angle  $\alpha$  of a photon with respect to its unperturbed path due to the presence of a point-like body of mass  $M$  would be given by

$$\alpha_N = \frac{2GM}{bc^2}, \quad (2.1)$$

where  $G$  is the gravity universal constant,  $b$  the impact parameter (minimum distance of the light ray from the mass  $M$ ) and  $c$  the velocity of light in vacuum. The argument to obtain such result in the Newtonian limit is quite simple and refers to the classical process of scattering on a centrally symmetric field force with inverse quadratic power-law with distance. Unfortunately, this calculation has a number of problems associated with it. Newton's theory simply cannot be applied to massless particles: they feel no gravitational force (because the force depends on their mass) and they have no inertia. Unaware of Soldner's calculation,

in 1907 Einstein began to think about the possible bending of light. By this stage, he had already formulated the equivalence principle, but it was to be another eight years before the general theory of relativity was completed. He realised that the equivalence principle in itself required light to be bent by gravitating bodies. But he assumed that the effect was too small ever to be observed in practice, so he shelved the calculation.

In 1911, still before the general theory was ready, he returned to the problem. What he did in this calculation was essentially to repeat the argument based on Newtonian theory, but incorporating the equation  $E = mc^2$ . Although photons do not have mass, they certainly have energy, and Einstein's theory says that even pure energy has to behave in some ways like mass. Using this argument, and spurred on by the realisation that the light deflection he was thinking about might after all be measurable, he calculated the bending of light from background stars by the Sun. These parameters yield a deflection of 0.87 seconds of arc; for reference, the angle in the sky occupied by the Sun is around half a degree. This answer is precisely the same as the Newtonian value obtained more than a century earlier by Soldner. The predicted deflection is tiny, but according to the astronomers Einstein consulted, it could just about be measured.

In 1915, with the full general theory of relativity in hand, Einstein returned to the light-bending problem. And he soon realised that in 1911 he had made a mistake. The correct answer was not the same as the Newtonian result, but twice as large. Einstein had neglected to include all effects of curved space in the earlier calculation. The origin of the factor two is quite straightforward when one looks at how a Newtonian gravitational potential distorts the metric of space-time. In flat space  $ds^2 = c^2 dt^2 - dl^2$  and null geodesics in this case are straight lines. Around a spherical distribution of mass  $M$  the metric changes so that, in the weak field limit, it becomes

$$ds^2 = \left(1 + \frac{2GM}{rc^2}\right) c^2 dt^2 - \left(1 - \frac{2GM}{rc^2}\right) dl^2. \quad (2.2)$$

Since the corrections in the parenthesis are small, one can solve the equation  $ds^2 = 0$  by expanding each bracket in a power series. Einstein's original calculation had included only the first term, which corresponds to the  $R_{00}$  part of the field equations. The second doubles the net deflection. Not only does energy gravitate, so does momentum and this appears in the second term in the metric. The angular deflection predicted by Einstein's equations in the Newtonian limit is therefore

$$\alpha_E = \frac{4MG}{bc^2}, \quad (2.3)$$

which yields 1.74 arcsec for  $M = M_\odot$  and  $b = R_\odot$ . Not only is this easier to measure, being larger, but it also offers the possibility of a definitive test of the theory, since it differs from the Newtonian value.

The 1919 British expeditions to Sobral and Principe were to prove this last calculation to be right. This was the first proof of GR [12].

## 2.3 Gravitational Optics

In general it is a difficult problem to determine the trajectories of light rays in curved spacetimes. However, in the cosmological setting, we can simplify the task by applying some

assumptions [43].

- For a start we assume that the global background geometry is well described by the Robertson–Walker metric we introduced in chapter 1.
- We make use of a Newtonian approximation for the light trajectories. We assume that a light ray travels unperturbed from a background source until it is very close to the lens, whereupon it is deflected by some angle we shall assume to be small. It then follows an unperturbed trajectory from the lens to the observer (so called Born approximation).
- We need to require that the effective gravitational potential of the lens  $\Phi$  is such that  $|\Phi^2| \ll c^2$ .
- The lens is moving with respect to a cosmological frame with a velocity  $v \ll c$ .

If these conditions apply, then the deflection produced by the lens is going to be small and it can be computed as

$$\hat{\alpha} = \frac{2}{c^2} \int \nabla_{\perp} \Phi dl \quad (2.4)$$

where the gradient of the Newtonian potential is taken perpendicular to the light path and the integral is taken along photon trajectory. With the simplification mentioned above, the gradient can be taken to be perpendicular to the original (unperturbed) light ray rather than the actual (perturbed) one. In this case we only need to consider the impact parameter  $b$  of the light ray as it crosses the lens plane. The relevant potential for a point lens can be written

$$\Phi(b, z) = -\frac{GM}{\sqrt{b^2 + z^2}} \quad (2.5)$$

where  $z$  is the distance along the path. With this for of the potential we find

$$\nabla_{\perp} \Phi(b, z) = \frac{GMb}{(b^2 + z^2)^{3/2}}. \quad (2.6)$$

The deflection angle is then

$$\hat{\alpha} = \frac{4GM}{c^2 b}. \quad (2.7)$$

Assuming that the lens is infinitely thin compared with the distances from source to lens and from observer to lens, we can consider the lens to be a mass sheet lying in a plane usually called the *lens plane*. The relevant property of the sheet is its surface mass density,  $\Sigma$ , where

$$\Sigma(\boldsymbol{\xi}) = \int \rho(\boldsymbol{\xi}, z) dz \quad (2.8)$$

in which the integral is taken over the photon path as before. It can be shown that the net deflection (now written as a vector to show its direction in the lens plane) is given by

$$\hat{\boldsymbol{\alpha}} = \frac{4G}{c^2} \int \frac{(\boldsymbol{\xi} - \boldsymbol{\xi}') \Sigma(\boldsymbol{\xi})}{|\boldsymbol{\xi} - \boldsymbol{\xi}'|^2} d^2 \boldsymbol{\xi}' \quad (2.9)$$

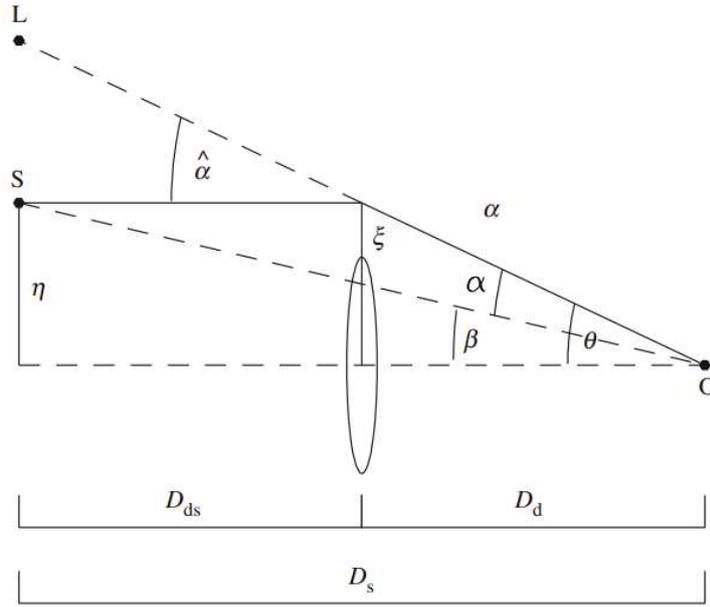


Figure 2.1: Gravitational lensing. A light ray travels from the source  $S$  to the observer  $O$  passing the lens at an impact parameter  $\xi$ . The transverse distance from the optic axis is  $\eta$ . The light ray is deflected through an angle  $\hat{\alpha}$ ; the angular separations of source and image from the optic axis are denoted  $\beta$  and  $\theta$ , respectively. The angular-diameter distances between observer and source, observer and lens and lens and source are  $D_s$ ,  $D_d$  and  $D_{ds}$ , respectively.

If the distribution of mass in the lens plane is circularly symmetric, then the deflection angle points towards the centre of symmetry and has modulus

$$\hat{\alpha} = \frac{4GM(\xi)}{c^2\xi} \quad (2.10)$$

where  $\xi$  is the distance from the centre of the lens and  $M(\xi)$  is the mass enclosed within a radius  $\xi$  so defined:

$$M(\xi) = 2\pi \int_0^\xi \Sigma(\xi') \xi' d\xi'. \quad (2.11)$$

We can now put this altogether to look at the geometry of a general lensing system as shown in Figure 2.1. The figure introduces the reduced deflection angle  $\alpha$ , which is related to  $\hat{\alpha}$  via

$$\alpha = \frac{D_{ds}}{D_s} \hat{\alpha}. \quad (2.12)$$

From the diagram, assuming small angles everywhere, we get

$$\theta D_s = \beta D_s - \alpha D_{ds} \quad (2.13)$$

so that

$$\beta = \theta - \alpha(\theta). \quad (2.14)$$

This is called the *lens equation*; it relates the angular position of images and sources. Note that angular-diameter distances must be used in this and the following.

As an example let us look at a case with constant surface mass density  $\Sigma$  in the lens plane. From equation (2.10) we obtain

$$\alpha(\theta) = \frac{D_{\text{ds}}}{D_s} \times \frac{4G}{c^2 \xi} \times \Sigma \pi \xi^2 = \frac{4\pi G \Sigma}{c^2} \frac{D_d}{D_{\text{ds}}} D_s \theta \quad (2.15)$$

where  $\xi = D_d \theta$ . In this case we can define a critical surface mass density

$$\Sigma_* = \frac{c^2}{4\pi G} \frac{D_s}{D_d D_{\text{ds}}}, \quad (2.16)$$

where  $D$  is defined by

$$D = \frac{D_d D_s}{D_{\text{ds}}}. \quad (2.17)$$

The interpretation of the critical density  $\Sigma_*$  is that the deflection angle  $\alpha(\theta) = \theta$  so that  $\beta = 0$  for any  $\theta$ . This is a perfect lens which brings all light rays to focus at a well-defined focal length. Real gravitational lenses are not perfect, but nevertheless display interesting optical properties. Lenses which have  $\Sigma > \Sigma_*$  typically produce multiple images of a background source. Now let us generalise to the case of a circular lens with an arbitrary mass profile. The lens equation (2.13) then becomes

$$\beta = \theta - \frac{D_{\text{ds}}}{D_d D_s} \frac{4GM(\theta)}{c^2 \theta}. \quad (2.18)$$

If the mass density is sufficient, then a source with  $\beta = 0$ , i.e. one that lies on the optic axis, is lensed into a ring with radius  $\theta_E$ , where

$$\theta_E^2 = \frac{4GM(\theta_E)}{Dc^2}. \quad (2.19)$$

This is called the Einstein radius.

For a point mass we obtain

$$\theta_E = \left( \frac{4GM}{Dc^2} \right)^{1/2}. \quad (2.20)$$

We can use this to rewrite the lens equation in this case as

$$\beta = \theta - \frac{\theta_E^2}{\theta} \quad (2.21)$$

which has two solutions:

$$\theta_{\pm} = \frac{1}{2} \left( \beta \pm \sqrt{\beta^2 + 4\theta_E^2} \right). \quad (2.22)$$

The two solutions correspond to two images, one lying on either side of the source. One image is always inside the Einstein ring and the other outside it. If the source is moved further from the optic axis (i.e. if  $\beta$  increases), then one image gets closer to the lens and the other gets nearer the source.

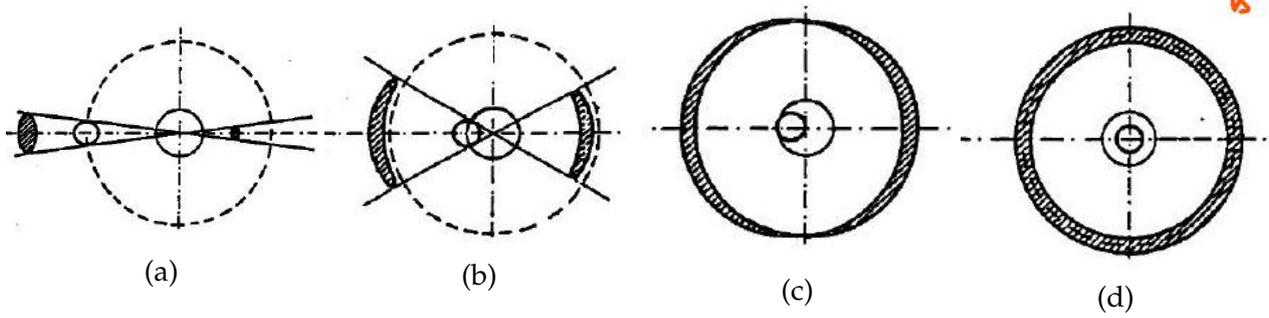


Figure 2.2: This illustrates the geometry of the deflection of light by a deflector in the center of each image. The lens is the inner circle, the source is the off-centre circle. The dashed circle marks the Einstein ring. The shaded region shows the lensed image. From (a) to (d): the source gets closer and closer to the lens. Interesting to note that for an extended source, like shown here as a white circle, the two images have apparent sizes larger than the real one. These can be obtained from the two crossing straight lines centred in the lens barycentre, the two lines being aligned with the borders of the source in its unperturbed position and defining the borders of the lensed images. To demonstrate this consider that every point in the source and lensed images can be treated according to eq (2.22), and particularly the points at the borders that are aligned that are aligned along the two lines.

## 2.4 More General Systems

The preceding section dealt with simple lens systems. In the following we shall look at some examples of how to deal with the more general case without any special symmetry. To simplify the notation let us start by defining a scaled potential  $\psi(\boldsymbol{\theta})$  by

$$\psi(\boldsymbol{\theta}) = \frac{1}{D} \frac{2}{c^2} \int \Phi(D_d \boldsymbol{\theta}, z) dz. \quad (2.23)$$

This is useful because the gradient of  $\psi$  with respect to  $\theta$  is just the deflection angle  $\boldsymbol{\alpha}$  because

$$\nabla_{\boldsymbol{\theta}} \psi = D_d \nabla_{\boldsymbol{\xi}} \psi = \frac{2}{c^2} \frac{D_{ds}}{D_s} \int \nabla_{\perp} \Phi dz = \boldsymbol{\alpha}. \quad (2.24)$$

Moreover, the Laplacian of  $\psi$  with respect to  $\theta$  is proportional to the surface mass density in the lens plane:

$$\nabla_{\boldsymbol{\theta}}^2 \psi = \frac{2}{c^2} \frac{D_d D_{ds}}{D_s} \int \nabla_{\boldsymbol{\xi}}^2 \Phi dz = \frac{2}{c^2} \frac{D_d D_{ds}}{D_s} \times 4\pi G \Sigma = 2 \frac{\Sigma}{\Sigma_*}. \quad (2.25)$$

It is then convenient to define the convergence field  $\kappa$  via

$$\kappa(\boldsymbol{\theta}) \equiv \frac{\Sigma(\boldsymbol{\theta})}{\Sigma_*}, \quad (2.26)$$

so that the Laplacian is just twice the convergence in a two-dimensional version of Poisson's equation:

$$\nabla_{\boldsymbol{\theta}}^2 \psi = 2\kappa \quad (2.27)$$

This means that we can write the potential as a function of  $\kappa$  using

$$\psi(\boldsymbol{\theta}) = \frac{1}{\pi} \int \kappa(\boldsymbol{\theta}') \log |\boldsymbol{\theta} - \boldsymbol{\theta}'| d^2\theta'. \quad (2.28)$$

Because the deflection angle is just the gradient of the potential  $\psi$  from (2.24), we can write

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) = \frac{1}{\pi} \int \kappa(\boldsymbol{\theta}') \frac{\boldsymbol{\theta} - \boldsymbol{\theta}'}{|\boldsymbol{\theta} - \boldsymbol{\theta}'|^2} d^2\theta', \quad (2.29)$$

which is equivalent to the equation (2.13) we obtained earlier.

In general the lens produces a mapping of the source plane onto the image plane. The local properties of this mapping are best specified by the Jacobian matrix

$$\mathbf{A}_{ij} = \frac{\partial\beta_i}{\partial\theta_j} = \left( \delta_{ij} - \frac{\partial\alpha_i(\boldsymbol{\theta})}{\partial\theta_j} \right) = \left( \delta_{ij} - \frac{\partial^2\psi}{\partial\theta_i\partial\theta_j} \right). \quad (2.30)$$

The Jacobian  $\mathbf{A}_{ij}$  may be thought of as the inverse of a magnification tensor  $M_{ij}$ . The local distortion of an image due to the lens given by the determinant of  $\mathbf{A}$ . If a solid angle  $\delta\beta^2$  of the source becomes  $\delta\theta^2$  in the image, then

$$\frac{\partial\theta^2}{\partial\beta^2} = \det M = \frac{1}{\det \mathbf{A}}. \quad (2.31)$$

This is a general form of equation (2.21).

The general properties of the mapping from source to image can be described somewhat more simply than the general form (2.21). First define a notation such that

$$\psi_{ij} \equiv \frac{\partial^2\psi}{\partial\theta_i\partial\theta_j}. \quad (2.32)$$

Using (2.27) we find that

$$\kappa = \frac{1}{2} (\psi_{11} + \psi_{22}). \quad (2.33)$$

We are in *weak lensing* regime if  $\kappa < 1$  and in *strong lensing* regime if  $\kappa > 1$ .

We can also use the elements of  $\psi_{ij}$  to construct components of a shear tensor. First define

$$\gamma_1 = \frac{1}{2} (\psi_{11} - \psi_{22}) \equiv \gamma \cos(2\phi) \quad (2.34)$$

and

$$\gamma_2 = \psi_{12} = \psi_{21} \equiv \gamma \sin(2\phi). \quad (2.35)$$

Using these definitions we can write

$$\mathbf{A} = \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{pmatrix} \quad (2.36)$$

which can also be written

$$\mathbf{A} = (1 - \kappa) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \gamma \begin{pmatrix} \cos 2\phi & \sin 2\phi \\ \sin 2\phi & -\cos 2\phi \end{pmatrix}. \quad (2.37)$$

This notation is useful because it allows a simple visual interpretation of the effects of lensing. A pure convergence  $\kappa$  corresponds to an isotropic magnification of the source in such a way that a circular source becomes a larger but still circular image. The components  $\gamma_1$  and  $\gamma_2$  represent shear in such a way that

$$\gamma = \sqrt{\gamma_1^2 + \gamma_2^2} \quad (2.38)$$

represents the magnitude of the shear and  $\phi$  its orientation. A non-zero shear transforms a circular source into an elliptical image.

In some places the mapping between source and image plane becomes singular. These singularities are normally called caustics and they lead to interesting optical effects owing to the non-uniqueness of the mapping between image and source planes to which they correspond. Basically a given (extended) lens will generate a set of caustics in the source plane. When a source crosses such a caustic a new pair of images is produced in the image. An extended lens can produce many images, depending on the mass distribution in the lens plane, while a point-mass lens only produces two. Near the caustics the shape of the images can be complicated, producing near-circular giant arcs. These can be very bright, owing to the magnification effect which is formally infinite at a caustic.

The consequences of these can be spectacular but complicated and, generally, considerable modelling is needed to understand the complex images obtained.

## 2.5 Applications

The gravitational effects on light propagation offer us an ideal tool to measure the total gravitating mass content in cosmological objects and structures [20], which is of course a very uncertain task, as we know; moreover gravitational lensing offers also other studies and applications:

- The case for strongly lensed high-redshift sources and the gravitational telescopes [15],
- The fraction of strongly lensed objects as a function of the cosmological parameters (constraints on  $\Omega_\Lambda$ ) [8],
- Time delays due to lensing and the estimate of  $H_0$  [14],
- Micro-lensing and the search for massive dark objects in the Milky Way halo,
- Variability of high-redshift quasars [34],
- Weak-lensing and the Large Scale Structure [41], [36], see figure 2.3.

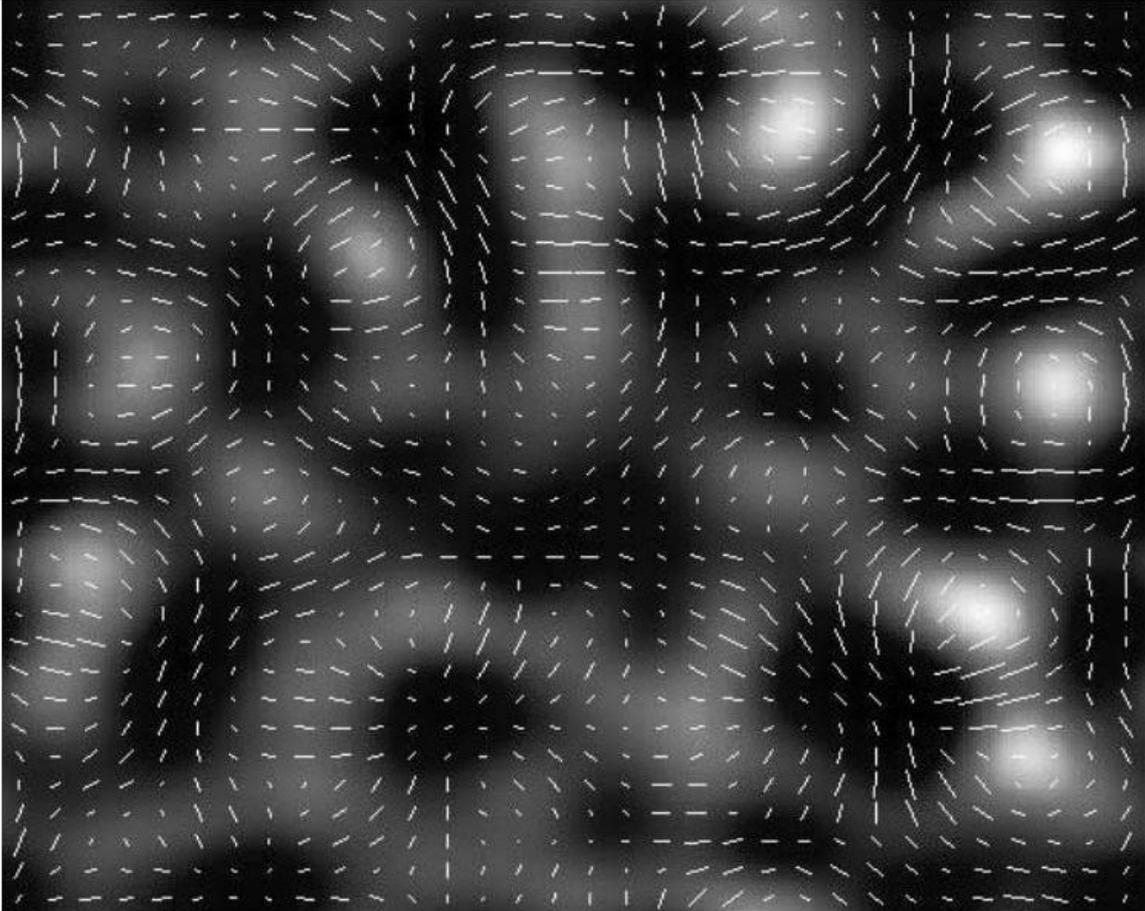


Figure 2.3: Simulation of the weak lensing distortion induced by large-scale structure. The pattern of density perturbations is shown as a greyscale picture upon which lines are superimposed representing the size and angle of the distortions (convergence and shear field)



# Chapter 3

## The Shape of Non-Gaussianity

### 3.1 Introduction

Most of the fields that we observe in cosmology are significantly non-Gaussian: the large-scale matter density is strongly non-Gaussian due to non-linear growth (1.32), and even the CMB is significantly non-Gaussian on small scales due to non-linear effects. Current evidence suggests that primordial fluctuations at the beginning of the hot big bang were very close to Gaussian, but primordial non-Gaussianity is still a possibility and, if observed, would be a way to rule out wide classes of early-universe models. Understanding non-Gaussianity is therefore crucial to extract the most information from cosmological observations. Since the initial fluctuations were close to Gaussian, the non-Gaussianities can often be treated perturbatively, in which case the bispectrum and trispectrum contain most of the additional information (though at very late times where strongly non-linear processes are involved a deeper analysis may be required). Here we will focus on non-Gaussianity as described by a bispectrum. We will show what fields with various qualitatively different types of non-Gaussianity look like in real space, and give some general results for the form of the squeezed bispectrum and its angular decomposition.

Here we do not aim to review the details of observational analysis or physical modelling of non-Gaussianities, we will mostly focus on generalities rather than specific models, though we mention specific cases of cosmological interest. For simplicity we restrict to scalar fields in flat space; the generalization to 2D fields on the sphere (e.g. the full-sky CMB or weak lensing convergence field), a non-flat background, or tensor fields (e.g. weak lensing shear and CMB polarization) is conceptually simple, though can be technically significantly more complicated.

We will assume a standard statistically isotropic and homogeneous background cosmology. In this case the wavevectors describing a particular non-Gaussian configuration of modes must sum to zero: for a bispectrum there are three modes with wavevectors that close to form a triangle; by "shape", we will mean a particular configuration of fixed-length wavevectors, e.g. a specific bispectrum triangle. A general bispectrum gives the full wavenumber and configuration dependence, i.e. the amount of signal expected in each possible shape. For example an 'orthogonal' bispectrum does not correspond to a particular triangle shape, but instead a particular wavenumber dependence of the signal in each shape. Instead of showing the magnitude of the signal as a function of the wavevectors, we instead focus on

what the signal looks like in real space for specific configurations. This can be a useful aid to understanding which kinds of physical process generate the different shapes (see [2] and [28]).

## 3.2 Gaussianity and the power spectrum

Before discussing non-Gaussianity, it is worth quickly remembering the key features of Gaussian fields. In particular we are usually interested in statistically isotropic and homogenous universe models, and hence in fields that have these symmetries. For simplicity we shall focus mainly on scalar fields in 2D flat space, for example a slice through the matter density field or a small patch of the weak lensing convergence field, or of the CMB, but almost everything generalizes to other cases such as full-sky observations. Assuming we can measure a field  $T(\mathbf{x})$  as a function of position, in flat space these can be Fourier transformed and written as

$$T(\mathbf{x}) = \frac{1}{(2\pi)^{N/2}} \int d\mathbf{k} T(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}} \quad (3.1)$$

where  $N$  is the number of dimensions. Statistical homogeneity and isotropy means that the statistical properties of the field must be unchanged under translations and rotations  $T(\mathbf{x}) \rightarrow T(\mathbf{x}')$ , so  $\langle T(\mathbf{x})T(\mathbf{x}') \rangle$  can only be a function of the invariant separation between the points  $|\mathbf{x} - \mathbf{x}'|$ . This implies that the covariance of the field is determined by a power spectrum depending only on  $k \equiv |\mathbf{k}|$ :

$$\langle T(\mathbf{k}_1) T(\mathbf{k}_2) \rangle = \delta(\mathbf{k}_1 + \mathbf{k}_2) P(k_1). \quad (3.2)$$

For a small patch of the CMB, the power spectrum is just  $C_l$ , where  $l = k$ . The delta function says that modes with different wavevectors are completely uncorrelated: knowing the sign of  $T(\mathbf{k}_1)$  tells you nothing about the likely sign of  $T(\mathbf{k}_2)$ . From the power spectrum the only thing we know is the variance of each individual mode, which from the assumption of isotropy is the same independent of the orientation of the mode.

A purely Gaussian statistically homogeneous and isotropic field is fully described statistically by its power spectrum. However more interesting fields are possible that are also statistically homogeneous and isotropic, with the non-Gaussian statistics described by a series of higher-point correlation functions.

## 3.3 Bispectrum

The first non-Gaussian signal to consider is a bispectrum, corresponding to a three-point correlation, or in Fourier space a correlation between three different mode wavevectors. We are still interested in statistically homogeneous and isotropic fields, which implies the statistics are described by a reduced bispectrum  $b(k_1, k_2, k_3)$  that depends only on the lengths of the wavevectors:

$$\langle T(\mathbf{k}_1) T(\mathbf{k}_2) T(\mathbf{k}_3) \rangle = \frac{1}{(2\pi)^{N/2}} \delta(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) b(k_1, k_2, k_3) \quad (3.3)$$

From now on when referring to the bispectrum we will mean the reduced bispectrum as defined here; an analogous definition applies on the full sky (4.1.2). The delta function

here means that the 3-mode correlation is zero unless the wavevectors sum to zero: they form a triangle. If there is a non-zero bispectrum, modes with different wavevectors are not independent: if we measure  $T(\mathbf{k}_1)$  and  $T(\mathbf{k}_2)$ , the sign of the bispectrum  $b(k_1, k_2, k_3)$  then tells us which sign of  $T(\mathbf{k}_3)$  is more likely. Positive sign gives positive skewness (tail of very high values), negative sign gives negative skewness (tail of very low values). What this looks like in real space depends on the shape of the triangle (the relative lengths of the different wavevectors).

### 3.3.1 Equilateral and flattened (folded) triangles

The first case we consider is equilateral triangles, where the lengths of the three sides of the triangle are the same,  $k_1 = k_2 = k_3$ . If there's a non-zero equilateral bispectrum, what does the field look like in real space? To answer this we can consider taking the  $T(\mathbf{k}_1)$  and  $T(\mathbf{k}_2)$  components of the field, and then ask what the bispectrum tells us about  $T(\mathbf{k}_3)$ . Depending on the relative sign of  $T(\mathbf{k}_3)$ , a field consisting of these three modes looks rather different - see Fig. 3.1. The sign of the bispectrum tells us which sign of  $T(\mathbf{k}_3)$  is more likely, in other words whether we are more likely to have small regions of concentrated overdensity ( $b > 0$ ) or regions of concentrated underdensity ( $b < 0$ ). As can easily be imagined, such patterns can be obtained by locally moving matter around, for example concentrated overdensities can form by gravitational collapse, and thus equilateral non-Gaussianity is likely to be present in any field undergoing local non-linear dynamical processes.

A bispectrum is determined by three wavevectors which always lie in a plane. In 3D, the modes we are considering correspond to plane waves, and the concentrated overdensities correspond to filaments. These are precisely what form during the growth of large-scale structure, as shown in the famous simulation of Fig. 3.2. Since it is the overdensities that are concentrated, not the underdensities, the non-linear large-scale structure density field will have a large positive equilateral component to its bispectrum.

Of course exactly equilateral triangles are a very special case, but there are many shapes that are close to equilateral and these will also look similar, but correspond to slightly elliptical concentrated overdensities or underdensities. As the bispectrum triangle becomes more flattened, these turn into a line, or in 3D concentrated overdensity or underdensity pancakes (planes); see Fig. 3.3. Note that shapes that are qualitatively distinct in 3D may not be after projection into 2D: for example if a purely equilateral shape is present in 3D, projecting down to 2D will give flattened contributions when the line of sight lies close to the plane of the triangle (slicing a 3D filament along its length gives a line of overdensity).

### 3.3.2 Squeezed triangles

Squeezed triangles correspond to having one wavevector much shorter than the other two: in other words one large-scale mode and two much shorter-scale modes. The bispectrum is invariant under permutations of  $k_1, k_2, k_3$ , so for squeezed triangles it is convenient to adopt the convention that we permute indices so that  $k_1 \leq k_2 \leq k_3$ , and  $k_1$  therefore labels the large-scale mode. Sometimes people refer to the "squeezed-limit", meaning the limit

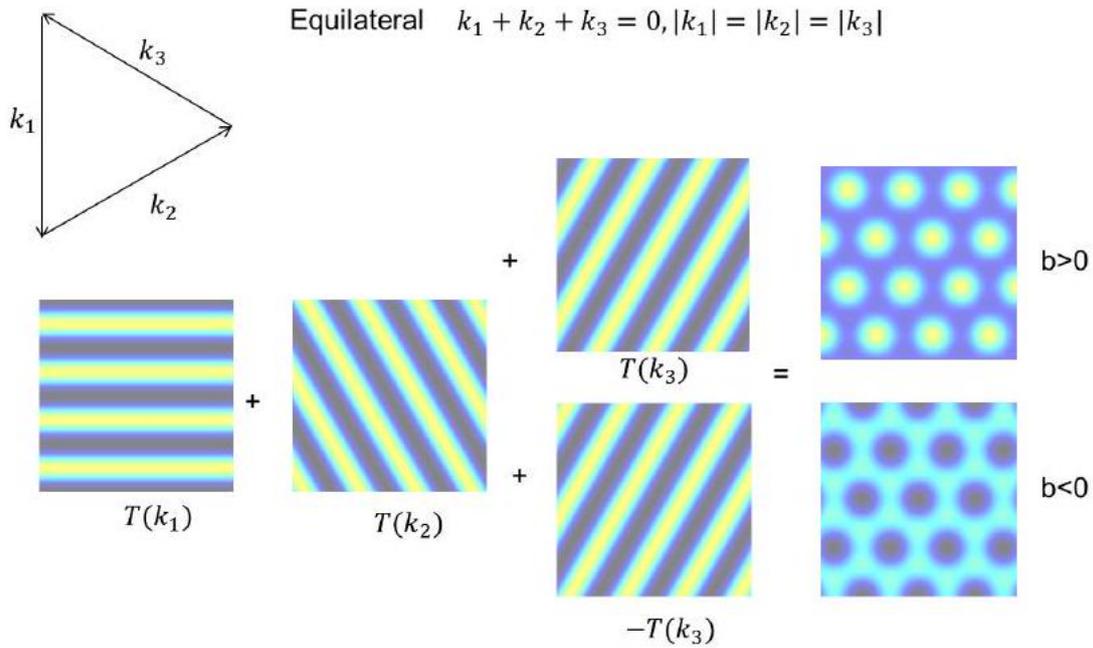


Figure 3.1: Equilateral bispectrum: a field can be decomposed into plane-wave modes, and the three components with wavevectors that form an equilateral triangle may have different relative signs. The sign of the bispectrum tells you which combination of signs is more likely (on average gives a positive or negative product of the three modes). A positive reduced bispectrum corresponds to being likely to have waves combining to have strong overdensities surrounded by larger areas of milder underdensity. A negative equilateral bispectrum corresponds to being likely to have concentrated underdensities surrounded by areas of milder overdensity. Note that in 3D the figures extend into the page, and hence the positive bispectrum corresponds to concentrated overdense filaments surrounded by larger areas of milder underdensity.

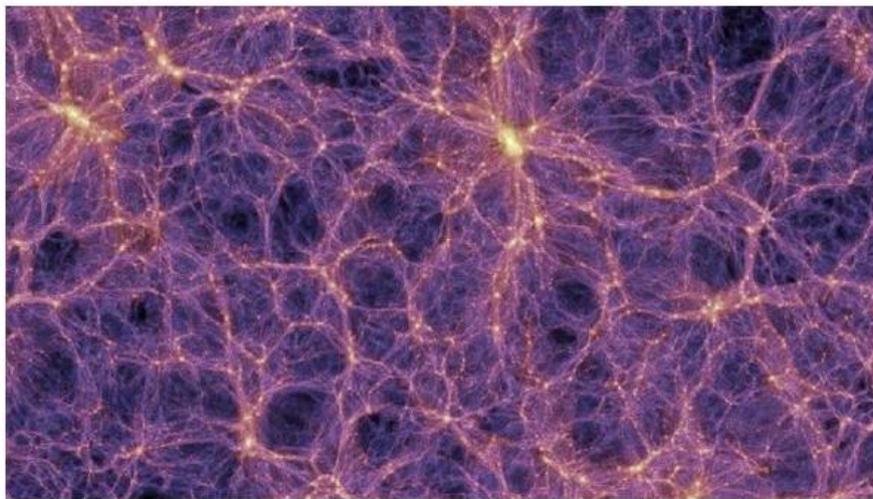


Figure 3.2: A snapshot of non-linear large-scale structure from the millennium simulations. Dynamical non-linear collapse of very dense filaments (surrounded by milder underdensities, voids) generates a large positive roughly equilateral density bispectrum.

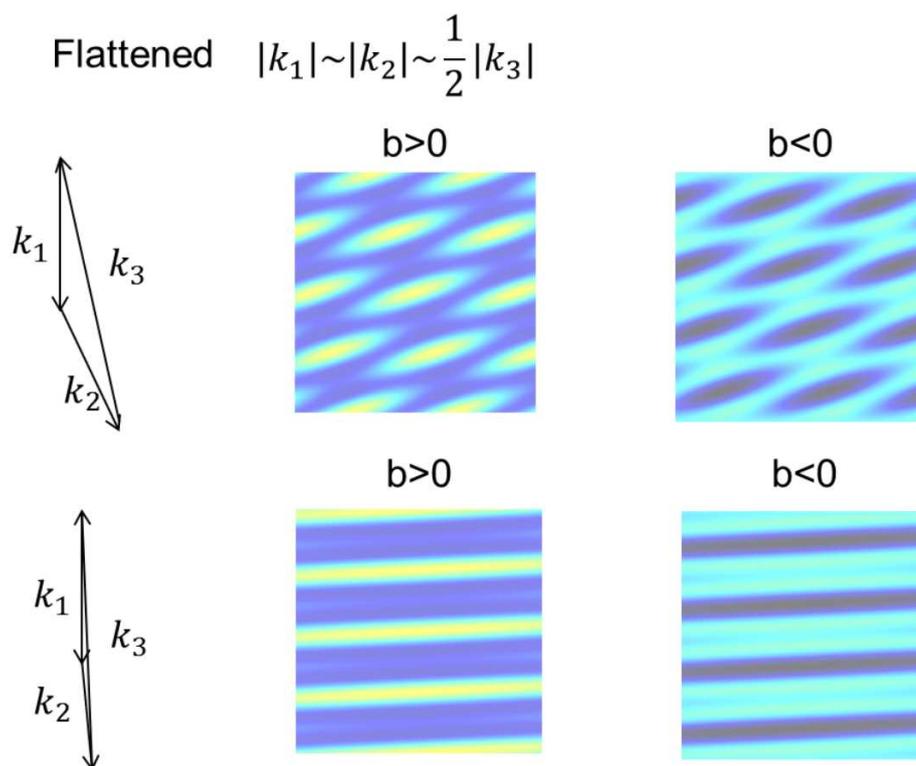


Figure 3.3: As an approximately equilateral bispectrum triangle flattens, the round areas of overdensity become flattened into pancakes. In 3D a positive flattened bispectrum with  $k_1 = k_2 = k_3/2$  corresponds to being likely to have overdense pancakes with larger mildly underdense planes in between.

as  $k_1 \rightarrow 0$ , but this is not really observationally relevant as very super-horizon modes are unobservable. By squeezed we will mean triangles with  $k_2, k_3 \gg k_1$ , but the wavelength of  $k_1$  not much larger than the horizon size today, so that the mode is still observationally relevant. First it is helpful to consider what a combination of two small-scale modes with  $\mathbf{k}_2 \sim -\mathbf{k}_3$  looks like: as shown in Fig. 3.4 the waves destructively interfere in some regions leaving little small-scale structure, but in other regions they reinforce each other giving a large small-scale signal. So this looks like a large-scale modulation in the small-scale power, where the wavevector of the modulation is given by  $-(\mathbf{k}_2 + \mathbf{k}_3)$ . In a Gaussian field the signs of all the modes are independent, so of course there is no modulation on average. However if there is a non-zero squeezed bispectrum, there is a correlation between this modulation and the large-scale modes; for example see Fig. 3.5a positive squeezed bispectrum means that where there's a large-scale overdensity there's likely to be more small-scale structure, and where there is an large-scale underdensity there is likely to be less small-scale structure.

### 3.3.2.a Angular dependence and the squeezed approximation

So far we have described the bispectrum triangles in terms of the lengths of the sides. However for squeezed triangles in particular it can be useful to describe the triangles in a different way. Since the two small-scale modes are of similar wavelength, is natural to use a

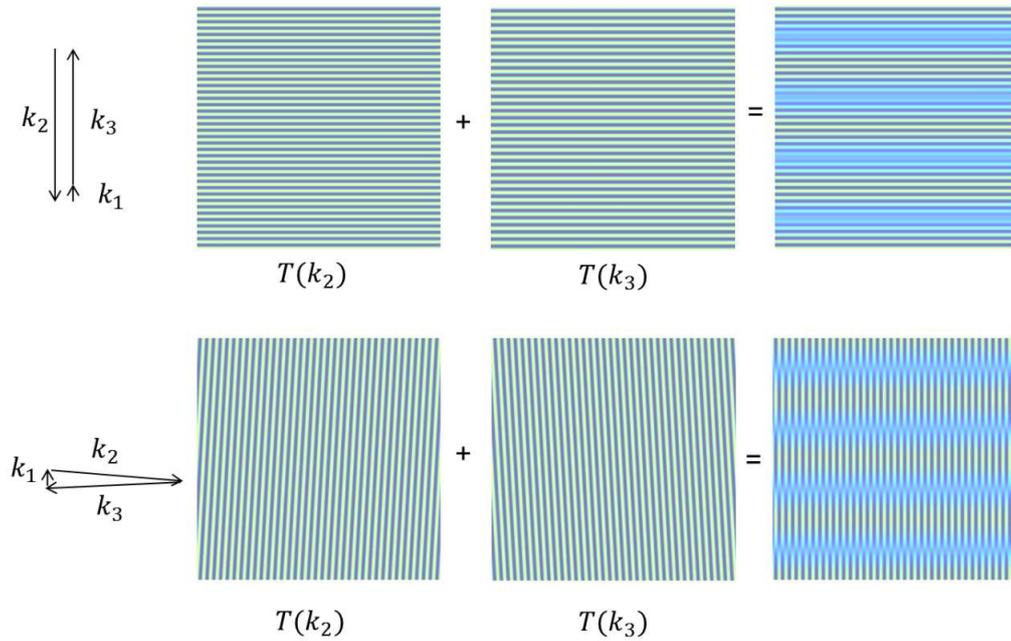


Figure 3.4: Two short-scale modes combine giving interference patterns. These look like a large-scale modulation in the amplitude of the small-scale modes, with the modulation having wavevector  $\mathbf{k}_1 = -\mathbf{k}_2 - \mathbf{k}_3$ .

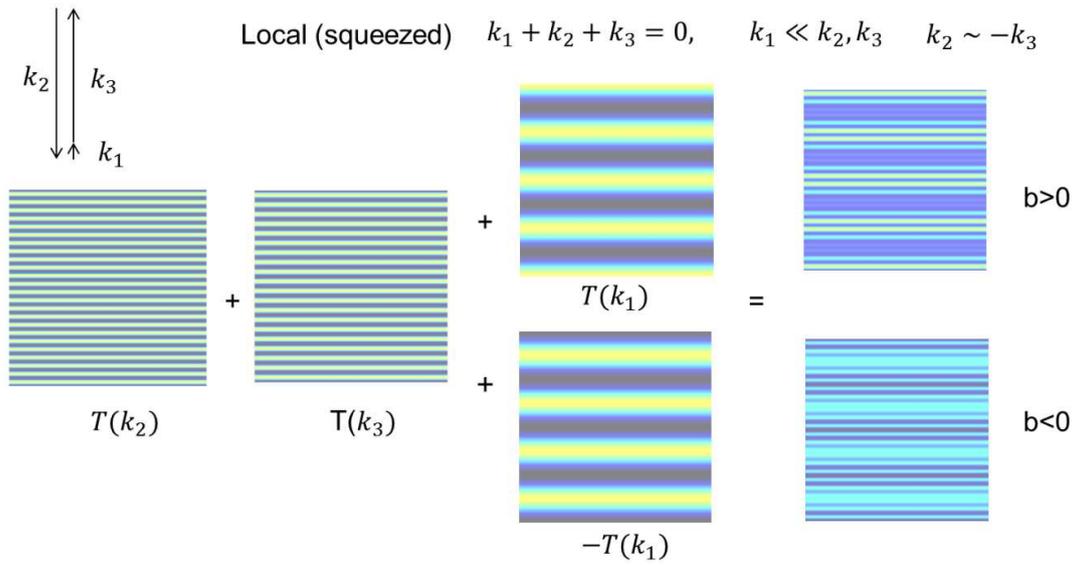
single number  $k \equiv |\mathbf{k}_2 - \mathbf{k}_3|/2$  to quantify the scale of the small-scale modes [27], as shown in Fig. 3.6. The remaining free parameter can then be taken to be the angle  $\phi$  between the large-scale and small-scale wavevectors, so that a triangle is fully described by the three numbers  $k_1, k, \phi$ . Squeezed triangles have  $k_1 \ll k$ , and in particular we shall refer to the squeezed limit as having leading corrections of  $\mathcal{O}((k_1/k)^2)$ . Note that exchanging  $k_2 \leftrightarrow k_3$  is equivalent to  $\phi \leftrightarrow \pi - \phi$  with the same  $k$ .

It can also be useful to decompose the bispectrum depending on its angular dependence, i.e. writing<sup>8</sup>

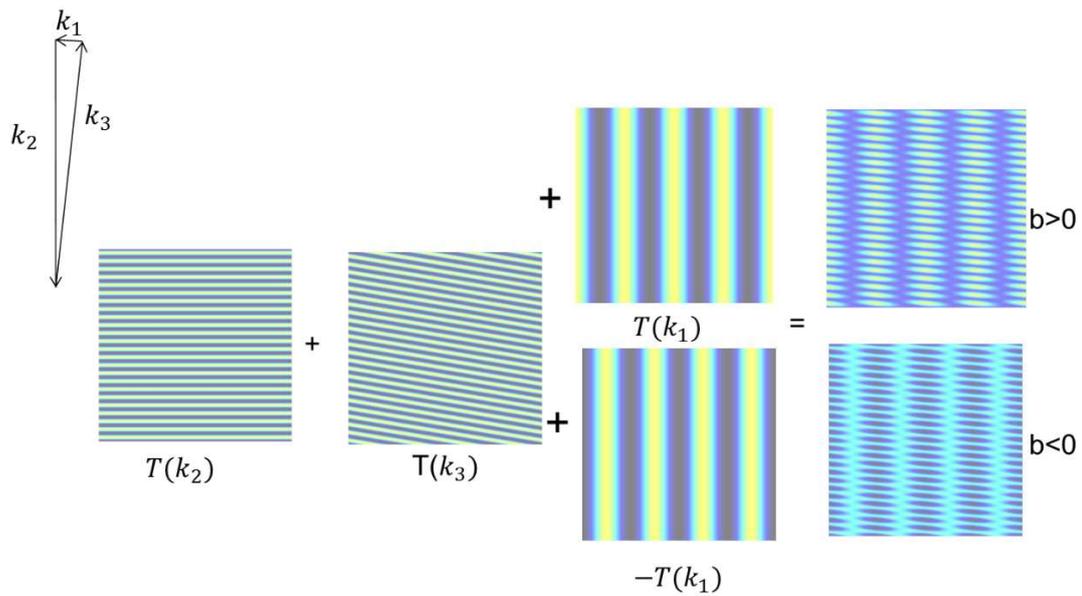
$$b(k_1, k, \phi) = \sum_m b_{k_1, k}^m e^{im\phi}. \quad (3.4)$$

If the squeezed bispectrum is independent of  $\phi$ , i.e. the orientation between the large and small-scale modes, it is called isotropic. In this case the bispectrum is fully determined by the  $m = 0$  component  $b_{k_1, k}^0$ . For 3D or statistically parity-invariant 2D fields there can be no  $\sin(m\phi)$  dependence, and from rotational invariance the odd  $m$  components should vanish, but in general there can be an angular dependence  $\cos(m\phi)$  with even  $m$ . This generally enters from  $\mathbf{k}_1 \cdot \mathbf{k}_2$  and  $\mathbf{k}_1 \cdot \mathbf{k}_3$  dependence of the bispectrum, for example from gradient contractions, and hence angular dependence enters in the combination  $k_1 \cos(\phi)$ . A bispectrum expansion in  $k_1/k$  therefore has angular dependence entering via powers of the dimensionless parameter  $\epsilon_\phi \equiv (k_1/k) \cos(\phi)$ . Since any bispectrum must be symmetric under  $k_2 \leftrightarrow k_3$  (and therefore  $\phi \leftrightarrow \pi - \phi$ ), only even powers of  $\epsilon_\phi$  can enter, and higher even powers of  $\cos(\phi)$  (and hence

<sup>8</sup>For bispectra in 3D, one could also expand into more directly orthogonal spherical harmonic  $Y_{L0}$  modes; the argument is much the same so for simplicity I will stick with the 2D modes.



(a)



(b)

Figure 3.5: Squeezed bispectrum: two small-scale modes with nearly-equal wavelength ( $k_2 \sim k_3$ ) interfere with each other, giving some regions with lots of small-scale power and others with destructive interference giving little small-scale power. The sign of the bispectrum tells you whether a region of high small-scale power is more likely to be associated with a large-scale overdensity or a large-scale underdensity. If the correlation is independent of the relative orientation (upper and lower figures have the same signal) the bispectrum is isotropic, but in general it is not.

higher even  $m$ ) are suppressed by proportionally more powers of  $k_1/k$ . Thus the squeezed

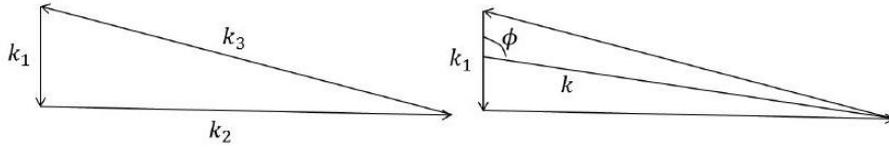


Figure 3.6: A bispectrum triangle can be described by the lengths of the three sides, or alternatively by the length of the shortest side (corresponding to the large-scale mode), the length-scale of the long sides (short-scale modes) defined by  $k = |\mathbf{k}_2 - \mathbf{k}_3|/2$ , and the angle  $\phi$  that measures the relative orientation of the long and short-scale modes.

expansion of a smooth scalar bispectrum is typically of the form

$$b(k_1, k_2, k_3) = A(k_1, k) + [B(k_1, k) + C(k_1, k) \cos(2\phi)] \frac{k_1^2}{k^2} + D(k_1, k, \phi) \mathcal{O}(k_1^4/k^4), \quad (3.5)$$

where  $A, B, C, D$  encode the scale-dependence of the particular physics involved. Hence unless there is very strong scale dependence the leading term is isotropic, and the next leading term only has isotropic and quadrupolar contributions. For some bispectra  $A = 0$  and the leading term then in general has both isotropic and quadrupolar components. Physically the reason higher angular dependence does not appear is because a small patch on a large-scale scalar modulating field will be accurately described by a field value and gradient, with the field value giving an isotropic change to the small-scale power, and the gradient defining the local basis for the quadrupolar dependence of the small-scale power. The leading term can only have anisotropic contributions if the modulation is not scalar.

For example gravitational lensing of the CMB generates both  $m = 0$  part of the bispectrum (corresponding to large-scale lenses isotropically magnifying and de-magnifying the CMB), and also an  $m = 2$  component from lensing shear (see e.g. refs. [27] [10] [5]). Primordial bispectra involving gravitational waves can also generate small  $m = 2$  components, since the modulation of the small-scale modes will depend on their orientation with respect to the anisotropic distortion produced by a large-scale tensor mode [29]. Similarly inflation models with vector fields can generate anisotropic bispectra [1]. Purely scalar local modulations are expected to give isotropic bispectra [6].

The decomposition of the squeezed bispectrum into angular moments is conceptually useful for distinguishing different physical effects. In particular, in the squeezed limit bispectra with different  $m$  are orthogonal<sup>9</sup>: an estimator for  $b_{k_1, k}^m$  should not be liable to confusion with a bispectrum  $b_{k_1, k}^{m'}$  if  $m \neq m'$ . For example this is partly what allows CMB lensing to be easily distinguished from an isotropic local primordial bispectrum: the  $m = 0$  part due to magnification is a source of confusion, but the  $m = 2$  lensing signal is distinctive and allows the lensing to be isolated and subtracted. The angular decomposition is also useful when considering secondary processing of primordial bispectra: if a statistically isotropic small-scale process affects the primordial modes it will not mix bispectra of different  $m$ . For example under gravitational lensing any primordial squeezed  $b_{k_1, k}^m$  is blurred out in  $k$  due to random small-scale lensing deflections, but the angular  $m$  dependence does not

<sup>9</sup>In 2D, see [35] [13]. For bispectra in 3D, one can expand in spherical harmonics  $Y_{L0}$ , and the modes with different  $L$  will then be orthogonal.

change [35] [13]. Since all bispectra are expected only to have the lowest  $m$  components in the squeezed limit, this also suggests that a general modal bispectrum decomposition can efficiently capture the angular dependence with only two angular modes for squeezed shapes, and since corrections are quadratic triangles do not need to be very squeezed for the squeezed approximation to be quite accurate.

Models in which statistical isotropy is broken can also generate anisotropic bispectra in the more general sense that the bispectrum signal can then depend on the orientation of the triangle [22] [11].

### 3.3.2.b General form of the squeezed bispectrum

Consider the case where the field being observed  $\tilde{T}$  can be calculated from some non-linear function of a set of purely Gaussian fields  $\{X_i\}$  that we can write as a vector  $\mathbf{X}$ . For example the Gaussian fields might be combinations of small scale and large-scale perturbation modes, different inflation field perturbations, lensing potentials, linear matter densities, velocity fields giving redshift distortions, etc. Since we are assuming underlying Gaussian fields, they are fully described by their covariance  $C$  (a matrix of power spectra). If we are interested in the bispectrum of a field  $\tilde{T}(X)$  (where the tilde denotes that it is non-linear), it can therefore be calculated by integrating out the Gaussian fields

$$\left\langle \tilde{T}(\mathbf{k}_1) \tilde{T}(\mathbf{k}_2) \tilde{T}(\mathbf{k}_3) \right\rangle = \int d\mathbf{X} \frac{\exp\left(-\frac{1}{2}\mathbf{X}^\dagger \mathbf{C}^{-1} \mathbf{X}\right)}{|2\pi\mathbf{C}|^{1/2}} \tilde{T}(\mathbf{k}_1) \tilde{T}(\mathbf{k}_2) \tilde{T}(\mathbf{k}_3) \quad (3.6)$$

For a squeezed bispectrum the large-scale field  $T(\mathbf{k}_1)$  is at a much larger scale than the two small-scale fields  $T(\mathbf{k}_2)$ ,  $T(\mathbf{k}_3)$ , and hence is often well approximated as being Gaussian (a linear function of the underlying Gaussian fields). For example in large-scale structure the large-scale mode in the bispectrum will be nearly linear if  $k_1 \lesssim 0.05 \text{Mpc}^{-1}$ , similarly in the CMB large scales are accurately linear but smaller scales have significant non-linearities due to lensing and other effects. The approximation  $\tilde{T}(\mathbf{k}_1) \approx T(\mathbf{k}_1)$  is the linear short-leg approximation [27], which is very accurate in some cases. Writing  $T(\mathbf{k}_1)$  as a linear combination of  $X_i(\mathbf{k}_1)$  in (3.6),  $T(\mathbf{k}_1) = M_i X_i(\mathbf{k}_1)$ , and then writing  $X_i(\mathbf{k}_1)$  as a functional derivative of the exponent,

$$\begin{aligned} \left\langle T(\mathbf{k}_1) \tilde{T}(\mathbf{k}_2) \tilde{T}(\mathbf{k}_3) \right\rangle &= \int d\mathbf{X} \frac{\exp\left(-\frac{1}{2}\mathbf{X}^\dagger \mathbf{C}^{-1} \mathbf{X}\right)}{|2\pi|^{1/2}} M_i X_i(\mathbf{k}_1) \tilde{T}(\mathbf{k}_2) \tilde{T}(\mathbf{k}_3) \\ &= - \int d\mathbf{X} M_i C_{ij}(k_1) \frac{\delta}{\delta X_j(\mathbf{k}_1)^*} \left( \frac{\exp\left(-\frac{1}{2}\mathbf{X}^\dagger \mathbf{C}^{-1} \mathbf{X}\right)}{|2\pi\mathbf{C}|^{1/2}} \right) \tilde{T}(\mathbf{k}_2) \tilde{T}(\mathbf{k}_3) \\ &= M_i C_{ij}(k_1) \int d\mathbf{X} \frac{\exp\left(-\frac{1}{2}\mathbf{X}^\dagger \mathbf{C}^{-1} \mathbf{X}\right)}{|2\pi|^{1/2}} \frac{\delta}{\delta X_j(\mathbf{k}_1)^*} \left( \tilde{T}(\mathbf{k}_2) \tilde{T}(\mathbf{k}_3) \right) \end{aligned} \quad (3.7)$$

Then since  $C_{ij} = P_{X_i X_j}$ , we have  $M_i C_{ij} = P_{T X_j}$  and hence

$$\left\langle \tilde{T}(\mathbf{k}_1) \tilde{T}(\mathbf{k}_2) \tilde{T}(\mathbf{k}_3) \right\rangle \approx \left\langle T(\mathbf{k}_1) \tilde{T}(\mathbf{k}_2) \tilde{T}(\mathbf{k}_3) \right\rangle = P_{T X_i}(k_1) \left\langle \frac{\delta}{\delta X_i(\mathbf{k}_1)^*} \left( \tilde{T}(\mathbf{k}_2) \tilde{T}(\mathbf{k}_3) \right) \right\rangle \quad (3.8)$$

Thus the bispectrum depends on the correlation of the large-scale field with the modulating fields  $P_{TX_i}(k_1)$ , and is proportional to the response of the small-scale non-linear modes to changes in the large-scale modulation. The result of 3.8 is fully non-perturbative and only relies on the linear short-leg approximation, not an extreme squeezed limit (if necessary it can be generalized out of the linear short-leg approximation by including higher derivative terms). In the case of CMB lensing it is possible to calculate the response term essentially exactly non-perturbatively [27]; more generally as  $k_1/k \rightarrow 0, k_2 \rightarrow k_3$ , and the response term just describes how the small-scale power spectrum changes with a different large-scale background modulation.

# Chapter 4

## Weak Lensing Statistics

This chapter is devoted to the analytical description of 2D field in flat space and on the sphere, its angular power spectrum and angle averaged integrated bispectrum. In the final section we describe the covariance matrix and we explain the need to estimate it analyzing thousands simulations.

### 4.1 Position-Dependent Power Spectrum from Convergence Maps

A relatively new observable has been developed in recent years that relies on the fact that the power spectrum measured from a survey sub-volume correlates with the mean of the same observable in the same sub-volume. This correlation gives a direct estimate of the bispectrum in the squeezed limit.

In this section we will present the analytical results relevant to the position-dependent power spectrum from 2D surveys with an emphasis particularly on 2D weak lensing surveys.

#### 4.1.1 Flat-Sky Treatment

Measurement of the bispectrum from weak lensing surveys is difficult due to non-ideal sky coverage caused by masking of regions close to bright objects as well as irregular survey boundaries. The IB proposed here tries to bypass these complexities by concentrating on the squeezed limit of the bispectrum, which can be estimated by techniques developed for estimation of the power spectrum. We will concentrate on the projected survey but generalization to tomographic bins is straight-forward.

In this section we develop the recently introduced statistics of position-dependent power spectrum to the case of weak lensing. We consider 2D weak lensing surveys but extension to 2D projected galaxy surveys can be done in a straight-forward manner. We will use the small angle approximation before generalising to the all-sky case in the following section.

Let us consider a weak lensing convergence map  $\kappa(\hat{\Omega})$  at a angular position of the sky  $\hat{\Omega}$  defined over a patch of the sky  $\hat{\Omega} = (\vartheta, \varphi)$ , where  $\vartheta$  and  $\varphi$  are right ascension and declination

respectively. We will divide the entire patch into equal area sub-patches. We will focus on one such sub-patch centered around the angular position  $\hat{\Omega}_0 = (\vartheta_0, \varphi_0)$ . The local average of  $\kappa$  on a sub-patch of the sky can differ from its global value of zero, where for a sub-patch the local average is:

$$\begin{aligned}\bar{\kappa}(\hat{\Omega}_0) &:= \frac{1}{\alpha} \int d^2\hat{\Omega} \kappa(\hat{\Omega}) W(\hat{\Omega} - \hat{\Omega}_0); & \alpha &:= \int W(\hat{\Omega} - \hat{\Omega}_0) d^2\hat{\Omega} \\ W(\hat{\Omega}) &:= \Theta(\vartheta - \vartheta_0) \Theta(\varphi - \varphi_0)\end{aligned}\quad (4.1)$$

Here,  $W$  describes the sky-patch and  $\Theta$  represents the one-dimensional top-hat function i.e.  $\Theta(\vartheta - \vartheta_0) = 1$  if  $|\vartheta - \vartheta_0| < \vartheta_S$  and zero otherwise and similarly  $\Theta(\varphi - \varphi_0) = 1$  if  $|\varphi - \varphi_0| < \varphi_S$  and zero otherwise. In this notation,  $\vartheta_S$  and  $\varphi_S$  represent half-width of a sub-patch along the  $\vartheta$  and  $\varphi$  directions while  $\alpha$  is the effective area of a sub-patch. We will assume that all sub-patches are of the same size and  $\alpha$  is independent of  $\hat{\Omega}_0$ <sup>10</sup>. In 2D, we will denote the Fourier wave-number as  $\mathbf{l}$  and use the following convention for Fourier transform:

$$\kappa(\mathbf{l}) := \int d^2\hat{\Omega} \exp(-i\mathbf{l} \cdot \hat{\Omega}) \kappa(\hat{\Omega}); \quad \kappa(\hat{\Omega}) := \int \frac{d^2\mathbf{l}}{(2\pi)^2} \exp(i\mathbf{l} \cdot \hat{\Omega}) \kappa(\mathbf{l}). \quad (4.2)$$

The power spectrum  $P^\kappa$  and bispectrum  $B^\kappa$  in 2D are defined using the following conventions:

$$\langle \kappa(\mathbf{l}_1) \kappa(\mathbf{l}_2) \rangle := (2\pi)^2 \delta_{2D}(\mathbf{l}_1 + \mathbf{l}_2) P^\kappa(l_1); \quad l = |\mathbf{l}| \quad (4.3a)$$

$$\langle \kappa(\mathbf{l}_1) \kappa(\mathbf{l}_2) \kappa(\mathbf{l}_3) \rangle := (2\pi)^2 \delta_{2D}(\mathbf{l}_1 + \mathbf{l}_2 + \mathbf{l}_3) B^\kappa(\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3) \quad (4.3b)$$

The angular brackets represent the ensemble average<sup>11</sup>. Here,  $\delta_{2D}$  is the Dirac delta function in 2D. The window  $W$  describing a patch can be used to extend the limits of angular integration. Thus, the flat-sky (local) Fourier transform takes the following form:

$$\begin{aligned}\kappa(\mathbf{l}; \hat{\Omega}_0) &:= \int d^2\hat{\Omega} \kappa(\hat{\Omega}) W(\hat{\Omega} - \hat{\Omega}_0) \exp(-i\mathbf{l} \cdot \hat{\Omega}) \\ &= \int \frac{d^2\mathbf{l}_1}{(2\pi)^2} \kappa(\mathbf{l} - \mathbf{l}_1) W(\mathbf{l}_1) \exp(-i\mathbf{l}_1 \cdot \hat{\Omega}_0)\end{aligned}\quad (4.4)$$

We use  $W(\mathbf{l})$  to denote the Fourier transform of  $W(\hat{\Omega})$ . Notice that the Fourier coefficient  $\kappa(\mathbf{l}; \hat{\Omega}_0)$  for  $\mathbf{l} = 0$  (monopole) is identical to  $\bar{\kappa}(\hat{\Omega}_0)$  defined above. The local convergence power spectrum  $P^\kappa(l; \hat{\Omega}_0)$  in this fraction of sky is given by (we will denote the global power spectrum as  $P^\kappa(l)$ ):

$$\begin{aligned}P^\kappa(\mathbf{l}; \hat{\Omega}_0) &= \frac{1}{\alpha} \int \frac{d^2\mathbf{l}_1}{(2\pi)^2} \int \frac{d^2\mathbf{l}_2}{(2\pi)^2} \kappa(\mathbf{l} - \mathbf{l}_1) \kappa(-\mathbf{l} - \mathbf{l}_2) \\ &\quad \times \exp[-i(\mathbf{l}_1 + \mathbf{l}_2) \cdot \hat{\Omega}_0] W(\mathbf{l}_1) W(\mathbf{l}_2)\end{aligned}\quad (4.5)$$

The resulting IB is defined by cross-correlating the local estimate of the power spectrum and the local average of the projected field:

<sup>10</sup>This is implemented considering equal area flat projection of the full-sky map, see **Flat projections**

<sup>11</sup>In our implementation this means averaging over  $N_{patches}$ , assuming isotropy ??

## 4.1. Position-Dependent Power Spectrum from Convergence Maps

$$\begin{aligned}
\mathcal{B}^\kappa(\mathbf{l}) &:= \left\langle P^\kappa(\mathbf{l}; \hat{\Omega}_0) \bar{\kappa}(\hat{\Omega}_0) \right\rangle \\
&= \frac{1}{\alpha^2} \int \frac{d^2 \hat{\Omega}_0}{2\pi} \int \frac{d^2 \mathbf{l}_1}{(2\pi)^2} \int \frac{d^2 \mathbf{l}_2}{(2\pi)^2} \int \frac{d^2 \mathbf{l}_3}{(2\pi)^2} \langle \kappa(\mathbf{l} - \mathbf{l}_1) \kappa(-\mathbf{l} - \mathbf{l}_2) \kappa(-\mathbf{l}_3) \rangle \\
&\quad \times W(\mathbf{l}_1) W(\mathbf{l}_1 + \mathbf{l}_3) W(\mathbf{l}_3) \exp\left[-i(\mathbf{l}_1 + \mathbf{l}_2 + \mathbf{l}_3) \cdot \hat{\Omega}_0\right]
\end{aligned} \tag{4.6}$$

The power spectrum  $P^\kappa$  and the average  $\bar{\kappa}$  used in eq (4.6) are estimated from the same patch of the sky. Many such patches are created by dividing the entire survey area. The computation of the average and the power-spectrum from a patch is far simpler than estimating the bispectrum directly. However, it is worth mentioning that such a simplification comes at a price, as the IB can only extract information about the squeezed limit of the bispectrum, if we focus on wave numbers  $l$  much larger than the wave numbers that correspond to the fundamental mode of the patch.

Next, we will specialize our discussion for weak lensing surveys. The weak lensing convergence  $\kappa$  is a line-of-sight projection of the 3D density contrast  $\delta(\mathbf{r})$ , in The Born approximation it is given by :

$$\kappa(\hat{\Omega}) := \int_0^{r_s} dr w(r) \delta(r, \hat{\Omega}); \quad w(r) := \frac{3\Omega_M H_0^2}{2} a^{-1} \frac{d_A(r) d_A(r_s - r)}{d_A(r_s)} \tag{4.7}$$

Here,  $r$  is the comoving radial distance,  $a$  represents the scale factor,  $H_0$  the Hubble parameter,  $\Omega_M$  the cosmological matter density parameter and  $d_A(r)$  is the comoving angular diameter distance. We will ignore the source distribution and assume them to be localized on a single source plane at a distance  $r = r_s$ <sup>12</sup>. Fourier decomposing  $\delta$  along and perpendicular to the line-of-sight direction we obtain:

$$\kappa(\hat{\Omega}) = \int_0^{r_s} dr \omega(r) \int \frac{dk_{\parallel}}{2\pi} \int \frac{d^2 \mathbf{k}_{\perp}}{(2\pi)^2} \exp\left[i\left(rk_{\parallel} + d_A(r)\hat{\Omega} \cdot \mathbf{k}_{\perp}\right)\right] \delta(\mathbf{k}; r) \tag{4.8}$$

In our notation,  $k_{\parallel}$  and  $\mathbf{k}_{\perp}$  are the components of the wave vector  $\mathbf{k}$  along and perpendicular to the line-of-sight direction and  $\delta(\mathbf{k})$  is the Fourier transform of the 3D density contrast  $\delta$ . We have used the following convention for the 3D FT and its inverse:

$$\delta(\mathbf{k}) := \int d^3 \mathbf{x} \exp(-i\mathbf{k} \cdot \mathbf{x}) \delta(\mathbf{x}); \quad \delta(\mathbf{x}) := \int \frac{d^3 \mathbf{k}}{(2\pi)^3} \exp(i\mathbf{x} \cdot \mathbf{k}) \delta(\mathbf{k}). \tag{4.9}$$

We have introduced  $\mathbf{x} = (r, \hat{\Omega})$  as the comoving distance and  $\mathbf{k}$  as the corresponding wave number. We will use  $\delta_{3D}$  to denote the 3D Dirac delta function. The 3D power spectrum and bispectrum for the density contrast  $\delta$  are:

$$\langle \delta(\mathbf{k}_1) \delta(\mathbf{k}_2) \rangle := (2\pi)^3 \delta_{3D}(\mathbf{k}_1 + \mathbf{k}_2) P_{3D}^\delta(k_1); \quad k = |\mathbf{k}| \tag{4.10a}$$

$$\langle \delta(\mathbf{k}_1) \delta(\mathbf{k}_2) \delta(\mathbf{k}_3) \rangle := (2\pi)^3 \delta_{3D}(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) B_{3D}^\delta(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) \tag{4.10b}$$

<sup>12</sup>We will also ignore photometric redshift errors. However, such complications are essential to link predictions to observational.

Using the Limber approximation [26] the convergence power spectrum  $P^\kappa(k)$  and bispectrum  $B^\kappa(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)$  can be expressed respectively in terms of the 3D matter power spectrum  $P_{3D}^\delta(k)$  (introduced in 1.3) and bispectrum  $B_{3D}^\delta(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)$ :

$$P^\kappa(l) = \int_0^{r_s} dr \frac{\omega^2(r)}{d_A^2(r)} P_{3D}^\delta\left(\frac{l}{d_A(r)}; r\right) \quad (4.11a)$$

$$B^\kappa(\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3) = \int_0^{r_s} dr \frac{\omega^3(r)}{d_A^4(r)} B_{3D}^\delta\left(\frac{\mathbf{l}_1}{d_A(r)}, \frac{\mathbf{l}_2}{d_A(r)}, \frac{\mathbf{l}_3}{d_A(r)}; r\right) \quad (4.11b)$$

To compute this, we use the fitting function developed in [50]. One can also include the post-Born correction [39] which becomes necessary at high redshift but has a small effect at low redshift.

In 2D the angular average of the IB, denoted as  $\bar{B}^\kappa(l)$ , can be defined as:

$$\bar{B}^\kappa(l) := \int \frac{d\varphi_l}{2\pi} B^\kappa(\mathbf{l}); \quad l = |\mathbf{l}| \quad (4.12)$$

Here  $\varphi_l$  is the angle between the vector  $\mathbf{l}$  and the  $\vartheta$  direction. Next, carrying out the  $\hat{\Omega}_0$  integral in eq (4.6) and using the resulting 2D delta function to perform the  $\mathbf{l}_2$  integral leaves us with the following expression:

$$\begin{aligned} \bar{B}^\kappa(l) = \int \frac{d\varphi_l}{2\pi} \int \frac{d^2\mathbf{l}_1}{(2\pi)^2} \int \frac{d^2\mathbf{l}_3}{(2\pi)^2} B^\kappa(\mathbf{l} - \mathbf{l}_1, -\mathbf{l} + \mathbf{l}_1 + \mathbf{l}_3, -\mathbf{l}_3) \\ \times W(\mathbf{l}_1) W(\mathbf{l}_1 + \mathbf{l}_3) W(\mathbf{l}_3). \end{aligned} \quad (4.13)$$

### 4.1.2 All-Sky Formulation

Our starting point is a 2D map  $M(\hat{\Omega})$  in pixel space for a generic field defined on the sphere, related to its spherical harmonic coefficients  $a_{\ell m}$ <sup>13</sup> by

$$M(\hat{\Omega}) = \sum_{\ell m} a_{\ell m} Y_{\ell m}(\hat{\Omega}) \quad \text{and} \quad a_{\ell m} = \int d^2\hat{\Omega} M(\hat{\Omega}) Y_{\ell m}^*(\hat{\Omega}) \quad (4.14)$$

where  $\hat{\Omega} = (\theta, \varphi)$  is the position on the celestial sphere.

A patch in the sky is selected by multiplying the map  $M(\hat{\Omega})$  by the patch window  $W(\hat{\Omega}^2 \hat{\Omega}_0)$  (patch centered at  $\hat{\Omega}_0$ ). One latter can be either a simple mask (i.e. a step function, taking values 1 inside the patch and 0 outside) or a more general beam function with no sharp edges in pixel space, like a multipole space patch as we will deal with, see 5.1). Hence, the position-dependent harmonic coefficients (i.e. the harmonic coefficients in the patch centered at  $\hat{\Omega}_0$ ) are given by

$$\tilde{a}_{\ell m}(\hat{\Omega}_0) = \int d^2\hat{\Omega} M(\hat{\Omega}) W(\hat{\Omega}, \hat{\Omega}_0) Y_{\ell m}^*(\hat{\Omega}) = \sum_{\ell_1 m_1} a_{\ell_1 m_1} K_{\ell \ell_1}^{m m_1}(\hat{\Omega}_0) \quad (4.15)$$

where we use the notation  $\sim$  to indicate that it is a patch quantity (with respect to the harmonic coefficient  $a_{\ell m}$  for the full sky) and  $K_{\ell \ell_1}^{m m_1}$  is the patch kernel defined by

<sup>13</sup>In order to refer at convergence field, it is sufficient to consider  $k_{\ell m}$  instead of  $a_{\ell m}$

#### 4.1. Position-Dependent Power Spectrum from Convergence Maps

$$K_{\ell_1 \ell_2}^{m_1 m_2}(\hat{\Omega}_0) \equiv \int d^2 \hat{\Omega} W(\hat{\Omega}, \hat{\Omega}_0) Y_{\ell_2 m_2}(\hat{\Omega}) Y_{\ell_1 m_1}^*(\hat{\Omega}) \quad (4.16)$$

The position-dependent power spectrum is then given by

$$\begin{aligned} \tilde{C}_\ell(\hat{\Omega}_0) &= \frac{1}{f_{\text{sky}}^W} \frac{1}{2\ell+1} \sum_{\ell m} \tilde{a}_{\ell m}(\hat{\Omega}_0) \tilde{a}_{\ell m}^*(\hat{\Omega}_0) \\ &= \frac{1}{f_{\text{sky}}^W} \frac{1}{2\ell+1} \sum_{\ell_1 m_1} \sum_{\ell_2 m_2} a_{\ell_1 m_1} a_{\ell_2 m_2} \sum_m (-1)^m K_{\ell \ell_1}^{m m_1}(\hat{\Omega}_0) K_{\ell \ell_2}^{-m m_2}(\hat{\Omega}_0) \end{aligned} \quad (4.17)$$

where we used  $(K_{\ell_1 \ell_2}^{m_1 m_2})^* = (-1)^{m_1+m_2} K_{\ell_1 \ell_2}^{-m_1 -m_2}$ .

The patch function can also be decomposed into spherical harmonics

$$W(\hat{\Omega}, \hat{\Omega}_0) = \sum_{\ell m} w_{\ell m}(\hat{\Omega}_0) Y_{\ell m}(\hat{\Omega}) \quad (4.18)$$

The local average value in a patch is then given by

$$\begin{aligned} \tilde{M}(\hat{\Omega}_0) &= \frac{1}{4\pi f_{\text{sky}}^W} \int d^2 \hat{\Omega} M(\hat{\Omega}) W(\hat{\Omega}, \hat{\Omega}_0) = \frac{1}{4\pi f_{\text{sky}}^W} \sum_{\ell m} \sum_{\ell' m'} a_{\ell m} w_{\ell' m'}(\hat{\Omega}_0) \int d^2 \hat{\Omega} Y_{\ell m}(\hat{\Omega}) Y_{\ell' m'}(\hat{\Omega}) \\ &= \frac{1}{4\pi f_{\text{sky}}^W} \sum_{\ell m} a_{\ell m} w_{\ell m}^*(\hat{\Omega}_0) \end{aligned} \quad (4.19)$$

where  $f_{\text{sky}}^W$  is the fraction of the sky covered by the patch and the last step is obtained using the orthogonality property of spherical harmonics.

The integrated angular bispectrum is defined as the cross-correlation of the average  $\tilde{M}$  and the position-dependent power spectrum  $\tilde{C}_\ell$  of a patch, averaged over the whole sky:

$$\text{IB}_\ell \equiv \left\langle \frac{1}{4\pi} \int d^2 \hat{\Omega}_0 \tilde{M}(\hat{\Omega}_0) \tilde{C}_\ell(\hat{\Omega}_0) \right\rangle. \quad (4.20)$$

In practice, the integrated bispectrum of a map (simulation or observation) is computed by averaging over all the patches

$$\text{IB}_\ell^{\text{obs}} = \frac{1}{N_{\text{patch}}} \sum_{\text{patch}} \tilde{M}_{\text{patch}}^{\text{obs}} \tilde{C}_{\ell, \text{patch}}^{\text{obs}} \quad (4.21)$$

where  $N_{\text{patch}}$  is the total number of patches used to divide the sky and  $\tilde{M}_{\text{patch}}^{\text{obs}}$  and  $\tilde{C}_{\ell, \text{patch}}^{\text{obs}}$  are determined for each patch after multiplying the data map by the patch window.

After substituting the position-dependent power spectrum eq. (4.17) and the mean patch value eq. (4.19) into the angular averaged integrated bispectrum definition (4.20), the expectation value of the integrated bispectrum becomes

$$\begin{aligned} \text{IB}_\ell &= \frac{1}{(4\pi)^2 (f_{\text{sky}}^W)^2} \frac{1}{2\ell+1} \sum_{\ell_1 m_1} \sum_{\ell_2 m_2} \sum_{\ell_3 m_3} \langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \rangle \\ &\times \int d^2 \hat{\Omega}_0 w_{\ell_3 m_3}^*(\hat{\Omega}_0) \sum_m (-1)^m K_{\ell \ell_1}^{m m_1}(\hat{\Omega}_0) K_{\ell \ell_2}^{-m m_2}(\hat{\Omega}_0) \end{aligned} \quad (4.22)$$

The integrated bispectrum depends explicitly on the full angular bispectrum  $\langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \rangle$ . The number of multipoles usually considered  $\ell_{\max} \sim \mathcal{O}(10^3)$  makes the brute force computation of this summation totally out of reach, a common issue to all CMB, and also weak lensing bispectrum analyses.

Assuming statistical isotropy, it is more convenient to work with a rotationally-invariant quantity called the angle-averaged bispectrum defined by<sup>14</sup>

$$B_{\ell_1 \ell_2 \ell_3} \equiv \left\langle \int d^2 \hat{\Omega} M_{\ell_1}(\hat{\Omega}) M_{\ell_2}(\hat{\Omega}) M_{\ell_3}(\hat{\Omega}) \right\rangle, \quad \text{with } M_\ell(\hat{\Omega}) = \sum_m a_{\ell m} Y_{\ell m}(\hat{\Omega}) \quad (4.23)$$

This angle-averaged bispectrum is related to the angular one by

$$\langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \rangle = h_{\ell_1 \ell_2 \ell_3}^{-1} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix} B_{\ell_1 \ell_2 \ell_3}, \quad (4.24)$$

where  $B_{\ell_1 \ell_2 \ell_3}$  is the full sky counterpart of (4.11b); the matrix is a Wigner-3  $j$  symbol and  $h_{\ell_1 \ell_2 \ell_3}$  is defined by

$$h_{\ell_1 \ell_2 \ell_3} \equiv \sqrt{\frac{(2\ell_1 + 1)(2\ell_2 + 1)(2\ell_3 + 1)}{4\pi}} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ 0 & 0 & 0 \end{pmatrix}. \quad (4.25)$$

We also use the related Gaunt integral<sup>15</sup>

$$\int d^2 \hat{\Omega} Y_{\ell_1 m_1}(\hat{\Omega}) Y_{\ell_2 m_2}(\hat{\Omega}) Y_{\ell_3 m_3}(\hat{\Omega}) = h_{\ell_1 \ell_2 \ell_3} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \quad (4.28)$$

and eq (4.18) to rewrite the patch kernel definition eq (4.16) as

$$K_{\ell_1 \ell_2}^{m_1 m_2}(\hat{\Omega}_0) = (-1)^{m_1} \sum_{\ell_3 m_3} w_{\ell_3 m_3}(\hat{\Omega}_0) h_{\ell_1 \ell_2 \ell_3} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ -m_1 & m_2 & m_3 \end{pmatrix} \quad (4.29)$$

After substituting this expression and eq (4.24) into the integrated bispectrum eq (4.22), we obtain

$$\begin{aligned} IB_\ell &= \frac{1}{(4\pi)^2 (f_{\text{sky}}^W)^2} \frac{1}{2\ell + 1} \sum_{\ell_1 \ell_2 \ell_3} \frac{B_{\ell_1 \ell_2 \ell_3}}{h_{\ell_1 \ell_2 \ell_3}} \sum_{m_1 m_2 m_3} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \\ &\times \sum_{m_4 m_5 m} (-1)^m \begin{pmatrix} \ell & \ell_1 & \ell_4 \\ -m & m_1 & m_4 \end{pmatrix} \begin{pmatrix} \ell & \ell_2 & \ell_5 \\ m & m_2 & m_5 \end{pmatrix} \int d^2 \hat{\Omega}_0 w_{\ell_3 m_3}^*(\hat{\Omega}_0) w_{\ell_4 m_4}(\hat{\Omega}_0) w_{\ell_5 m_5}(\hat{\Omega}_0). \end{aligned} \quad (4.30)$$

<sup>14</sup>Note that the reduced bispectrum  $B_{\ell_1 \ell_2 \ell_3} / h_{\ell_1 \ell_2 \ell_3}^2$ , where  $h_{\ell_1 \ell_2 \ell_3}$  is defined in eq. (4.25) is also often used in the literature.

<sup>15</sup>The symbol  $\mathcal{G}$  introduced in eq (4.28) is also known as Gaunt integral which represents the coupling of three spherical harmonics. To relate eq (4.10b) and eq (4.3b) we note that the Gaunt integral introduced in (4.28) becomes a Dirac delta function in the flat-sky limit

$$\mathcal{G}_{\ell_1 \ell_2 \ell_3}^{m_1 m_2 m_3} \approx (2\pi)^2 \delta_{2D}(\mathbf{l}_1 + \mathbf{l}_2 + \mathbf{l}_3), \quad (4.26)$$

which leads us to identify the reduced bispectrum as the flat-sky bispectrum [cite 57]:

$$b_{\ell_1 \ell_2 \ell_3}^\kappa \approx B^\kappa(\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3); \quad (4.27)$$

## 4.1. Position-Dependent Power Spectrum from Convergence Maps

The remaining integral depends exclusively on the choice of patches and its dependence on several multipole numbers  $m$ 's still makes the integrated bispectrum impossible to compute in practice. Similarly to other bispectrum estimators, the solution comes under the form of separability of the integrated bispectrum in  $\ell$ -space, which we achieved using a class of simple azimuthally symmetric patches.

### 4.1.2.a Exploiting multipole space patches

We now exploit the use of azimuthally symmetric patches having the following simple  $m$ -dependence:

$$w_{\ell m}(\hat{\Omega}_0) = w_{\ell} Y_{\ell m}^*(\hat{\Omega}_0) \quad (4.31)$$

Using the addition theorem of spherical harmonics

$$\sum_m Y_{\ell m}(\hat{\Omega}) Y_{\ell m}^*(\hat{\Omega}') = \frac{2\ell + 1}{4\pi} P_{\ell}(\hat{\Omega} \cdot \hat{\Omega}') \quad (4.32)$$

where  $P_{\ell}$  is a Legendre polynomial, it is straightforward to obtain from eq. (4.18) the simple and  $m$ -independent expression of the patch window function in real space

$$W(\hat{\Omega}, \hat{\Omega}_0) = \sum_{\ell} w_{\ell} \frac{2\ell + 1}{4\pi} P_{\ell}(\hat{\Omega} \cdot \hat{\Omega}_0). \quad (4.33)$$

With this type of patches, the remaining integral term in the integrated bispectrum expression (4.30) becomes a Gaunt integral, eq. (4.28).

Substituting this into eq (4.30) gives

$$\begin{aligned} IB_{\ell} &= \frac{1}{(4\pi f_{\text{sky}}^W)^2} \frac{1}{2\ell + 1} \sum_{\ell_1 \ell_2 \ell_3 \ell_4 \ell_5} B_{\ell_1 \ell_2 \ell_3} \frac{h_{\ell_1 \ell_4} h_{\ell_2 \ell_5} h_{\ell_3 \ell_4 \ell_5} w_{\ell_3} w_{\ell_4} w_{\ell_5}}{h_{\ell_1 \ell_2 \ell_3}} \\ &\times \sum_{\substack{m_1 m_2 m_3 \\ m_4 m_5 m}} (-1)^{m+m_3} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \begin{pmatrix} \ell & \ell_1 & \ell_4 \\ -m & m_1 & m_4 \end{pmatrix} \begin{pmatrix} \ell_2 & \ell_5 & m \\ m & m_2 & m_5 \end{pmatrix} \begin{pmatrix} \ell_3 & \ell_4 & \ell_5 \\ -m_3 & m_4 & m_5 \end{pmatrix}. \end{aligned} \quad (4.34)$$

In this new expression, the  $m$ -dependent part (the second line) is only a summation of Wigner  $3j$ -symbols, hence it is independent from the exact choice of patches. More importantly, this summation of Wigner  $3j$ -symbols reduces to a Wigner  $6j$ -symbol:

$$\begin{aligned} &\sum_{\substack{m_1 m_2 m_3 \\ m_4 m_5 m}} (-1)^{m+m_3} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \begin{pmatrix} \ell & \ell_1 & \ell_4 \\ -m & m_1 & m_4 \end{pmatrix} \begin{pmatrix} \ell & \ell_2 & \ell_5 \\ m & m_2 & m_5 \end{pmatrix} \begin{pmatrix} \ell_3 & \ell_4 & \ell_5 \\ -m_3 & m_4 & m_5 \end{pmatrix} \\ &= (-1)^{\ell_2 + \ell_4} \left\{ \begin{matrix} \ell_1 & \ell_2 & \ell_3 \\ \ell_5 & \ell_4 & \ell \end{matrix} \right\}. \end{aligned} \quad (4.35)$$

Substituting this and eq (4.25) into (4.34), we obtain our final expression for the integrated bispectrum:

$$IB_{\ell} = \frac{1}{(4\pi)^3 (f_{\text{sky}}^W)^2} \sum_{\ell_1 \ell_2 \ell_3 \ell_4 \ell_5} B_{\ell_1 \ell_2 \ell_3} \mathcal{F}_{\ell_1 \ell_2 \ell_3 \ell_4 \ell_5} w_{\ell_3} w_{\ell_4} w_{\ell_5} \quad (4.36)$$

where  $\mathcal{F}_{\ell_1\ell_2\ell_3\ell_4\ell_5}$  is a shorthand notation for the part of the expression depending only on multipole numbers

$$\begin{aligned} \mathcal{F}_{\ell_1\ell_2\ell_3\ell_4\ell_5} &= (-1)^{\ell_2+\ell_4} (2\ell_4 + 1) (2\ell_5 + 1) \\ &\times \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ 0 & 0 & 0 \end{pmatrix}^{-1} \begin{pmatrix} \ell & \ell_1 & \ell_4 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \ell_2 & \ell_5 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \ell_3 & \ell_4 & \ell_5 \\ 0 & 0 & 0 \end{pmatrix} \left\{ \begin{matrix} \ell_1 & \ell_2 & \ell_3 \\ \ell_5 & \ell_4 & \ell \end{matrix} \right\}. \end{aligned} \quad (4.37)$$

As stressed out at the beginning of the section, the position-dependent power spectrum is by definition interesting as a tool to study the squeezed limit of the bispectrum, where one multipole number is small compared to the two others. With the following convenient choice of patch window function

$$w_\ell = 0 \text{ for } \ell > \ell_w \text{ with } \ell_w \sim \mathcal{O}(10) \quad (4.38)$$

it is possible to remove (almost all) non-squeezed terms from the sum in eq (4.36) which implies a significant computational gain. With this choice, only the part of the bispectrum where  $\ell_3 \leq \ell_w$  is considered in the summation. The multipoles  $\ell_4$  and  $\ell_5$  also have to be less than or equal to  $\ell_w$ . Moreover, recalling that the Wigner  $3j$ -symbols are zero when the triangle inequality is not respected, the only non-vanishing terms of eq (4.36) require that  $\ell_1, \ell_2 \in [\ell - \ell_w, \ell + \ell_w]$ . Therefore, the total number of terms in the sum for a given  $\ell$  has been drastically reduced and is only of  $\mathcal{O}(\ell_w^5)$  (compared to  $\mathcal{O}(\ell_{\max}^5)$  before), which is fast to compute.

For this broad class of patches, one can determine the exact expected integrated bispectrum from the theoretical bispectrum shape.

#### 4.1.2.b Covariance

Before using these results to analyze data from simulations and observations, we also need its expected covariance matrix, which can be computed following a similar method.

First, we recall the well-known expression of the bispectrum variance, valid in the usual weak non-Gaussianity case ( $\langle B_{\ell_1\ell_2\ell_3} \rangle \approx 0$ ) corresponding to CMB observations or high redshift weak lensing maps

$$\begin{aligned} \langle B_{\ell_1\ell_2\ell_3} B_{\ell'_1\ell'_2\ell'_3} \rangle &= h_{\ell_1\ell_2\ell_3}^2 C_{\ell_1} C_{\ell_2} C_{\ell_3} \left[ \delta_{\ell_1\ell'_1} \delta_{\ell_2\ell'_2} \delta_{\ell_3\ell'_3} + \delta_{\ell_1\ell'_2} \delta_{\ell_2\ell'_1} \delta_{\ell_3\ell'_3} + \delta_{\ell_1\ell'_1} \delta_{\ell_2\ell'_3} \delta_{\ell_3\ell'_2} \right. \\ &\quad \left. + \delta_{\ell_1\ell'_2} \delta_{\ell_2\ell'_3} \delta_{\ell_3\ell'_1} + \delta_{\ell_1\ell'_3} \delta_{\ell_2\ell'_2} \delta_{\ell_3\ell'_1} + \delta_{\ell_1\ell'_3} \delta_{\ell_2\ell'_1} \delta_{\ell_3\ell'_2} \right] \end{aligned} \quad (4.39)$$

It is then straightforward to obtain the covariance,

$$\begin{aligned} \text{IC}_{\ell\ell'} &\equiv \langle B_\ell B_{\ell'} \rangle \\ &= \frac{1}{(4\pi)^6 \left(f_{\text{sky}}^W\right)^4} \sum_{\substack{\ell_1\ell_2\ell_3\ell_4\ell_5 \\ \ell'_1\ell'_2\ell'_3\ell'_4\ell'_5}} \langle B_{\ell_1\ell_2\ell_3} B_{\ell'_1\ell'_2\ell'_3} \rangle \mathcal{F}_{\ell_1\ell_2\ell_3\ell_4\ell_5} \mathcal{F}_{\ell'_1\ell'_2\ell'_3\ell'_4\ell'_5} w_{\ell_3} w_{\ell_4} w_{\ell_5} w_{\ell'_3} w_{\ell'_4} w_{\ell'_5} \\ &= \frac{1}{(4\pi)^6 \left(f_{\text{sky}}^W\right)^4} \sum_{\substack{\ell_1\ell_2\ell_3\ell_4\ell_5 \\ \ell'_4\ell'_5}} C_{\ell_1} C_{\ell_2} C_{\ell_3} h_{\ell_1\ell_2\ell_3}^2 \mathcal{F}_{\ell_1\ell_2\ell_3\ell_4\ell_5} w_{\ell_3} w_{\ell_4} w_{\ell_5} w_{\ell'_4} w_{\ell'_5} \\ &\times \left[ w_{\ell_3} \left( \mathcal{F}_{\ell_1\ell_2\ell_3\ell'_4\ell'_5} + \mathcal{F}_{\ell_2\ell_1\ell_3\ell'_4\ell'_5} \right) + w_{\ell_2} \left( \mathcal{F}_{\ell_1\ell_3\ell_2\ell'_4\ell'_5} + \mathcal{F}_{\ell_3\ell_1\ell_2\ell'_4\ell'_5} \right) + w_{\ell_1} \left( \mathcal{F}_{\ell_3\ell_2\ell_1\ell'_4\ell'_5} + \mathcal{F}_{\ell_2\ell_3\ell_1\ell'_4\ell'_5} \right) \right], \end{aligned} \quad (4.40)$$

where the last step is obtained after summing over  $\ell'_1, \ell'_2$  and  $\ell'_3$ . Note that the terms in  $w_{\ell_1}$  and  $w_{\ell_2}$  on the last line are zero if  $\ell > 2\ell_w$  (thus for most of the  $\ell$ 's considered). Note also that the sum is separable in  $(\ell_4, \ell_5)$  and  $(\ell'_4, \ell'_5)$ , hence in practice there are only  $\mathcal{O}(\ell_w^5)$  different terms to evaluate, as for the integrated bispectrum, eq (4.36).

#### 4.1.2.c NG contribution to Covariance

In the previous section, we have shown that the integrated bispectrum variance  $IC_{\ell\ell}$  is well described by eq. (4.40), where we assume to be in the regime of weak non-Gaussianity. However, in [18] we verify that this approximation no longer works well enough when we look at off-diagonal covariance terms. Weak lensing convergence maps are indeed significantly non-Gaussian. An accurate description of the bispectrum and integrated bispectrum covariances thus requires including higher-order connected correlation functions in the calculation [44]. A calculation of this type was performed in [23], using however the Limber approximation. The same approach is not sufficient here because we are mainly interested in the squeezed limit, where this approximation is not valid. While the full computation of the squeezed bispectrum covariance is beyond the scope of this work, we can computationally study the level of correlation between different configurations (i.e., off-diagonal covariance terms), see figure 4.2

## 4.2 The integrated bispectrum estimator

### 4.2.1 Definition

As anticipated, The integrated bispectrum estimator is based on three relatively simple steps: separate the celestial sphere into many equal-sized patches, determine the power spectrum (small-scale fluctuations) and the average value (large-scale mode) in each patch and compute their patch-by-patch correlation averaged over the sky. The result is called the integrated bispectrum and is by construction sensitive to the correlations between small-scale and large-scale effects, like the squeezed limit of the bispectrum (one multipole much smaller than the other two).

Implementing this method first requires the characteristics of the patches (size, shape, number) to be specified, and then sky realizations  $W_{\text{patch}}(\hat{\Omega})$  produced, where the index 'patch' denotes the exact patch considered in the full set (see figure 1 for examples). Then, one by one, these patch maps, are applied as masks to the observational data  $\kappa^{\text{obs}}(\Omega)$ . For each resulting map, we only have to compute two simple quantities; its power spectrum  $C_{\ell, \text{patch}}^{\text{obs}}$ , which is called the position-dependent power spectrum in the literature, and its average value  $\bar{\kappa}_{\text{patch}}^{\text{obs}}$ . The product of these two quantities is finally averaged over all the patches to obtain the integrated angular bispectrum estimator:

$$IB_{\ell}^{\text{obs}} = \frac{1}{N_{\text{patch}}} \sum_{\text{patch}} \bar{\kappa}_{\text{patch}}^{\text{obs}} C_{\ell, \text{patch}}^{\text{obs}} \quad (4.41)$$

where  $N_{\text{patch}}$  is the number of patches used to divide the sky. For example, a way to separate the sky into a set of equal-sized patches is to use the standard HEALPix pixelization: starting from a given data map, one degrades it to lower resolution and defines a patch by fixing

every pixel of the low resolution map to zero, except one. The power spectrum in the chosen patch is then computed at high resolution. Repeating this process for each low-resolution pixel gives a set of patches that covers uniformly the full sky.

## 4.2.2 Mask

Both CMB and LSS datasets are characterized by incomplete sky coverage, either because only part of the sky is actually observed, or because foreground contamination imposes to mask a non-negligible part of it. The missing power from the masked/non-observed regions induces a multiplicative bias on the observed quantities:

$$C_\ell^{\text{masked}} = f_{\text{sky}} C_\ell^{\text{unmasked}}, \quad B_{\ell_1 \ell_2 \ell_3}^{\text{masked}} = f_{\text{sky}} B_{\ell_1 \ell_2 \ell_3}^{\text{unmasked}} \quad \text{and} \quad IB_\ell^{\text{masked}} = f_{\text{sky}} B_\ell^{\text{unmasked}} \quad (4.42)$$

the so-called  $f_{\text{sky}}$  approximation, where  $f_{\text{sky}}$  is the fraction of the sky that is left unmasked (see [19] for the derivation of the first two equations while the last one is obtained by substituting the corrected bispectrum into eq. (4.36).

## 4.2.3 Linear correction

When using observational data from actual surveys, statistical isotropy is broken due to partial sky coverage and anisotropic noise. This creates a large spurious bispectrum in the squeezed limit due to the correlations between small-scale fluctuations (e.g. noise power spectrum) and large-scale effects (e.g. scanning pattern of the satellite). As shown in [19], the large resulting bias to the integrated bispectrum can be removed using a simple correction term  $I_\ell^{\text{obs}} \rightarrow IB_\ell^{\text{obs}} - IB_\ell^{\text{lin}}$  given by

$$IB_\ell^{\text{lin}} = \frac{1}{N_{\text{patch}}} \sum_{\text{patch}} \bar{\kappa}_{\text{patch}}^{\text{obs}} C_{\ell, \text{patch}}^{\text{MC}} \quad (4.43)$$

This mean-field correction displays a linear dependence on the observed data  $\left(\bar{\kappa}_{\text{patch}}^{\text{obs}}\right)$ , while the quadratic term  $C_{\ell, \text{patch}}^{\text{MC}}$  is a Monte-Carlo average of the position-dependent power spectrum from many simulations sharing the same experimental characteristics as the observed data. This is conceptually similar to the standard linear correction of the bispectrum originally introduced in [9].

## 4.3 Covariance estimation

The integrated bispectrum estimator build with flat-sky power spectrum estimator is a binned estimator (bin width depends on the dimension of the projection 5.11).

From [18] we recall the results of preliminary studies. A simple example is the covariance matrix of the binned bispectrum estimator. Because of the large bins (width of 100 for  $\ell > 100$ ), each triplet of bins contain many different correlated configurations. The estimated covariance matrix from 40 convergence simulations at  $z_s = 1$  is, in the squeezed limit, more than two times larger that its theoretical expectation computed in the diagonal case. While using

smaller bins (width of 10), the difference becomes much smaller for most of the configurations. This shows that, taken individually, the effect of non-diagonal terms is rather small; however, when integrating over many modes (e.g. using large bins), their summed contribution quickly becomes the dominant one.

Have been shown that the diagonal part of the integrated bispectrum covariance matrix is well-approximated by eq. 4.40, as can be also checked in figure 4.1, where the theoretical covariance to the one estimated from the usual 40 simulations are compared. In this figure, they also show that the error bar on the mean value of a bin of width 100 in multipole space is actually of the same order as the value from a single multipole  $\ell$  (except at very low  $\ell$ ), while it should be 10 times smaller if modes were uncorrelated. This confirms the large correlation between different integrated bispectrum modes. Doing a step further with the

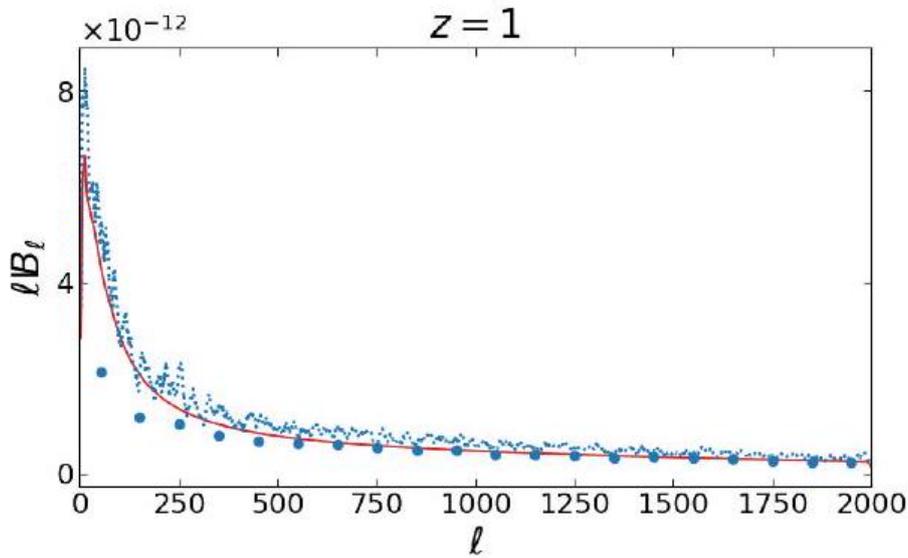


Figure 4.1: The integrated bispectrum variance  $IC_{\ell\ell}$  of the weak lensing convergence at  $z_s = 1$  using step function patches. The solid red line corresponds to the theoretical prediction. The blue dotted line is the variance estimated from 40 simulations, while the blue circles are also determined from these simulations after compressing the integrated bispectrum information to 20 bins. From [18].

integrated bispectrum and estimate its full covariance from simulations as shown in figure 4.2 (left panel). This full covariance is compared to the theoretical counterpart, in the right panel, determined in the case of a diagonal full bispectrum covariance eq. (4.39). While the theoretical covariance in the weak NG limit is close to diagonal (non-zero terms are only found where  $|\ell - \ell'| \leq 2\ell_w^{\max} = 20$ ), we can see that the actual covariance from simulations (left panel) is non-negligible everywhere. Even very different  $\ell$  and  $\ell'$  are correlated. This of course will need to be taken into account in more advanced applications, aimed for example at building a bispectrum/integrated bispectrum likelihood to measure cosmological parameters.

A direct solution to characterize the full integrated bispectrum covariance matrix is simply to estimate it from many mock simulations, as exemplified in figure 4.2. However, reaching the desired accuracy for actual parameter estimation and matrix inversion cannot be achieved with only 40 simulations, as done here for illustrative purposes. Considering the typical

number of modes in our analysis, we will actually need thousands of simulations. This is the reason we are trying to implement a faster pipeline making use of flat-sky  $P_\kappa(k, \hat{\Omega})$  estimation.

The alternative approach, namely deriving analytically the integrated bispectrum covariance matrix using eq. (4.40) presents several difficulties, like, on one side, the sheer number of terms in the sum (10 or more different multipole numbers<sup>16</sup>) and, on the other, the costly evaluation of many  $6j$ -symbols. On top of this, eq. (4.40) is also assuming that the full bispectrum covariance (4.39) is already available and we only need to extract the integrated bispectrum covariance out of it.

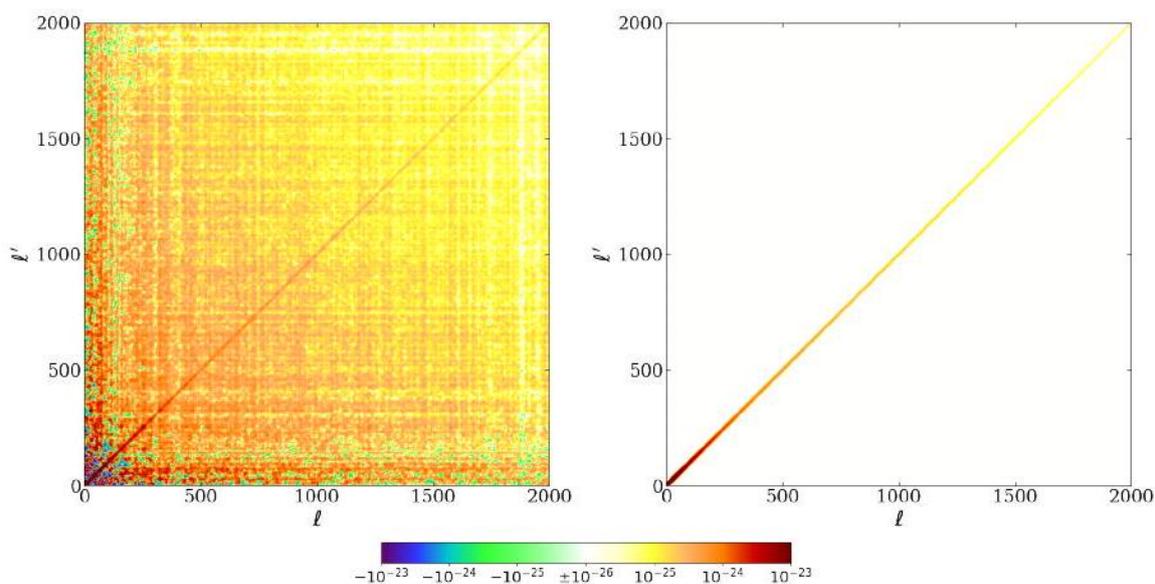


Figure 4.2: The integrated bispectrum covariance  $IC_{\ell\ell'}$  of the weak lensing convergence at  $z_s = 1$  using step function patches. On the left, the covariance is estimated from the 40 simulations. On the right, the theoretical covariance computed using eq. 4.40 if the bispectrum covariance is given by (4.39) (valid only in the weak non-Gaussianity regime. For visibility,  $IC_{\ell\ell'}$  is multiplied by the factor  $\ell\ell'$ . Note that the color scale is logarithmic, except for values smaller in absolute value than  $10^{-26}$  where it becomes linear. From [18].

<sup>16</sup>Within the flat-sky approximation the angular size of the projection and thus the angular width of the mexican needlet patch need an higher  $\ell_w$ , therefore the number of multiples to consider for theoretical signal computation increase consistently. This is the reason why up to some value of  $\ell_w^{max} \sim 25$  running the script to evaluating theoretical  $-\text{bispectrum}$  and  $-\text{variance}$  they do not give consistent results as the computation is too heavy.

# Chapter 5

## Computational Implementation

Weak lensing full sky simulations were used in [18] to validate the integrated angular bispectrum estimator described in chapter 4 for weak lensing studies and to further characterize the squeezed limit of the convergence bispectrum. It is crucial, to apply this statistic to data analysis and parameter estimation, to extract the covariance matrix. This requires applying the integrated bispectrum estimator to many thousands simulations, however the current implementation is too slow. Here we investigate a new implementation, based on the so-called flat-sky approximation, in which the power spectrum in small patches of the spherical domain is computed via a tangent-plane projection.

### 5.1 Maps and Methods

#### 5.1.1 Weak lensing maps

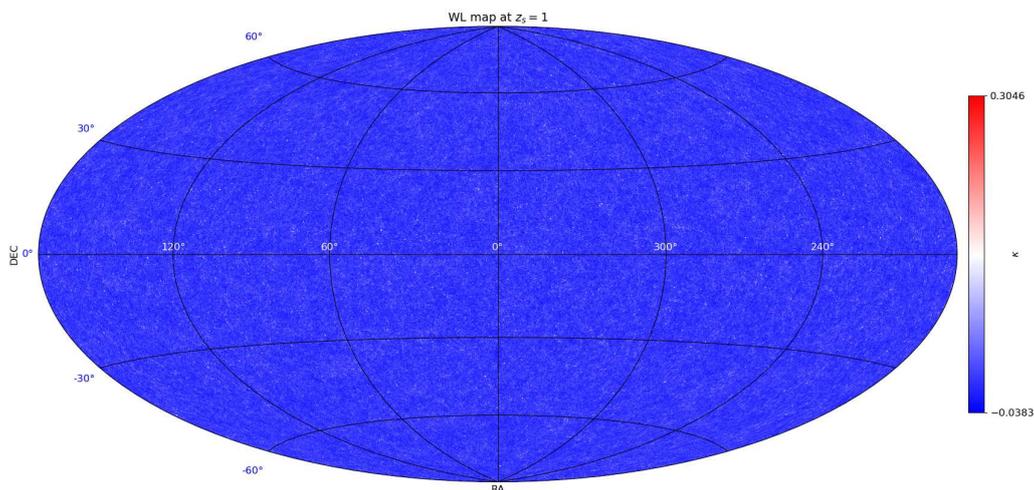


Figure 5.1: Weak lensing full-sky map at  $z_s = 1$ ,  $n_{side} = 2048$ .

Accurate simulations of the weak lensing convergence field have been built by Takahashi et al. [50], using ray-tracing through N-body simulations<sup>17</sup>, briefly discussed in the next section;

<sup>17</sup>They present 108 full-sky gravitational lensing simulation data sets generated by performing multiple-lens

We want to analyze several sets of these maps<sup>18</sup> at source redshifts ( $z_s = 1$ ).

## 5.1.2 N-body simulations and Ray-tracing

### 5.1.2.a Ray-tracing simulations: The principle

The simulations proceed by following light rays through the inhomogeneous matter distribution in the Universe [45]. The latter is generated by cosmological simulations of structure evolution. Those start at an early epoch by generating a realization of a Gaussian random field with a power spectrum according to the cosmological model considered, and follow the evolution of the density and velocity field of the matter using Newtonian gravity in an expanding Universe. The mass distribution is represented by discrete particles whose evolution in time is followed. A finite volume of the Universe is simulated this way, typically a box of comoving side-length  $L$ , for which periodic boundary conditions are applied. This allows one to use Fast Fourier Transforms (FFT) to evaluate the gravitational potential and forces from the density distribution. The box size  $L$  should be chosen such that the box contains a representative part of the real Universe, and must therefore be larger than the largest scales on which structure is expected, according to the power spectrum; a reasonable choice is  $L \gtrsim 100h^{-1}\text{Mpc}$ . The number of grid points and the number of particles that can be distributed in this volume is limited by computer memory; modern simulations work typically with  $256^3$  points and the same number of particles, though larger simulations have also been carried out; this immediately yields the size of grid cells, of order  $0.5h^{-1}\text{Mpc}$ . This comoving length, if located at a redshift of  $z \sim 0.3$  (which is about the most relevant for cosmic shear), subtends an angle of roughly  $2'$  on the sky. The finite number of particles yields the mass resolution of the simulations, which is typically  $\sim 10^{10}h^{-1}M_\odot$ , depending on cosmological parameters.

In order to obtain higher spatial resolution, force calculations are split up into near-field and far-field forces. The gravitational force due to the distant matter distribution is obtained by grid-based FFT methods, whereas the force from nearby masses is calculated from summing up the forces of individual particles; such simulations yield considerably higher resolution of the resulting mass distribution. Since the matter in these simulations is represented by massive particles, these can undergo strong interactions, leading to (unphysical) large orbital deflections. In order to avoid these unphysical strong collisions, the force between pairs of particles is modified at short distances, typically comparable to the mean separation of two particles in the simulation. This softening length defines the minimum length scale on which the results from numerical simulations can be considered reliable. Cosmological simulations consider either the dark matter only or, more recently, the hydrodynamics effects of baryons have been incorporated as well.

---

plane ray-tracing through high-resolution cosmological N-body simulations. The data sets include full-sky convergence and shear maps from redshifts  $z = 0.05$  to  $5.3$  at intervals of  $150h^{-1} \text{ Mpc}$  comoving radial distance (corresponding to a redshift interval of  $\Delta z \simeq 0.05$  at the nearby universe).

<sup>18</sup>We use the maps with  $N_{side} = 4096$  which can be downloaded here at [http://cosmo.phys.hirosaki-u.ac.jp/takahasi/allsky\\_raytracing/nres12.html](http://cosmo.phys.hirosaki-u.ac.jp/takahasi/allsky_raytracing/nres12.html). We downgrade their resolution to  $N_{side} = 2048$  and impose  $l_{max} = 2000$  for all the analyses presented in this section, unless mentioned otherwise.

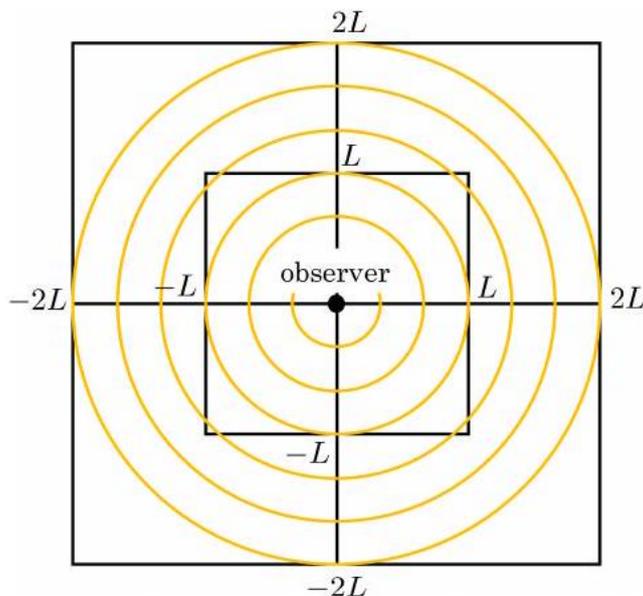


Figure 5.2: Configuration of ray-tracing simulation with cubic simulation boxes of lengths  $L, 2L, 3L, \dots$ , where  $L = 450h^{-1}\text{Mpc}$  (comoving scale), placed around the observer. The figure shows the inner two boxes with side lengths  $L$  and  $2L$ , respectively. The observer is located at the vertex of the boxes. In each box, we constructed three spherical shells with thickness of  $\Delta r = L/3 = 150h^{-1}\text{Mpc}$ ; the orange circles show the boundaries between the shells [50].

The outcome of such simulations, as far as they are relevant here, are the 3D positions of the matter particles at different (output) times or redshifts. In order to study the light propagation through this simulated mass distribution, one employs multiple lens-plane theory. First, the volume between us and sources at some redshift  $z_s$  is filled with boxes from the cosmological simulations. That is, the comoving distance  $w_s = w(z_s)$  is split up into  $n$  intervals of length  $L$ , and the mass distribution at an output time close to  $t_i = t(w = (i - 1/2)L)$  is considered to be placed at this distance. In this way, one has a light cone covered by cubes containing representative matter distributions. Since the mass distributions at the different times  $t_i$  are not independent of each other, but one is an evolved version of the earlier one, the resulting mass distribution is highly correlated over distances much larger than  $L$ . This can be avoided by making use of the statistical homogeneity and isotropy of the mass distribution: each box can be translated by an arbitrary two-dimensional vector, employing the periodicity of the mass distribution, and rotated by an arbitrary angle; furthermore, the three different projections of the box can be used for its orientation. In this way - a kind of recycling of numerical results - the worst correlations are removed.

Alternatively, one can combine the outputs from several simulations with different realizations of the initial conditions. In this case, one can use simulation boxes of different spatial extent, to match the comoving size of a big light cone as a function of redshift. That is, for a given light-cone size, only relatively small boxes are needed at low redshifts, and bigger ones at higher redshift.

Second, the mass in each of these boxes is projected along the line-of-sight, yielding a surface mass density at the appropriate comoving distance  $w_i = (i - 1/2)L$ . Each of these surface mass densities can now be considered a lens plane, and the propagation of light can be

followed from one lens plane to the next; the corresponding theory was worked out in detail by Blandford & Narayan [4]. Important to note is that the surface mass density  $\Sigma$  in each lens plane is the projection of  $\Delta\rho = \rho - \bar{\rho}$  of a box, so that for each lens plane,  $\langle \Sigma \rangle = 0$ . This multiple lens-plane approach presents a well-defined discretization of the full 3-dimensional propagation equations. Light bundles are deflected and distorted in each lens plane and thus represented as piecewise straight rays. The resulting Jacobi matrix  $\mathcal{A}$  is then obtained as a sum of products of the tidal matrices in the individual lens planes, yielding a discretized version for  $\mathcal{A}$ . The result of such simulations is then the matrix  $\mathcal{A}(\boldsymbol{\theta})$  on a predefined angular grid, as well as the positions  $\boldsymbol{\beta}(\boldsymbol{\theta})$  in the source plane. The latter will not be needed here, but have been used in studies of multiple images caused by the LSS.

### 5.1.3 Healpix map's structure

Concerning the observational component of the code, the first step is to generate the patches. One possibility would be to use HEALPix pixelization<sup>19</sup> and its hierarchical structure (see Appendix A).

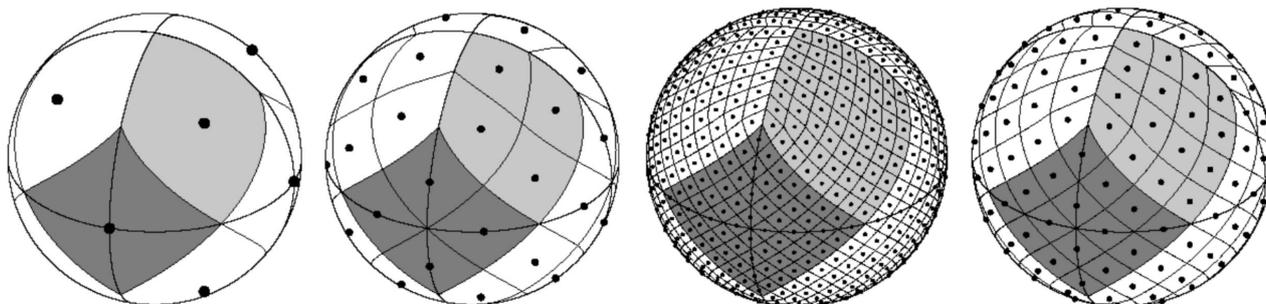


Figure 5.3: Healpix pixelization scheme,  $N_{side} = 1, 2, 3, 4$  respectively.

In figure 5.3 is shown the isotropic distribution on the sphere of pixels centers, which is a crucial aspect for the implementation as we assume statistical isotropy 4.1.2.

### 5.1.4 Patches and masks

#### 5.1.4.a Motivation

The need to divide full sky maps into small patches can be easily achieved using the HEALpix python library named `healpy`, which, thanks to its hierarchical nature, offers a simple way to divide the celestial sphere in a set of equal-sized areas by simply taking the (large) pixels of a very low resolution maps. For example to generate 192 patches, make a map with the resolution  $N_{side} = 4$  and a zero value for each of its 192 pixels. Then, fix the value of a pixel to one and upgrade the resolution to the same as the studied data map (e.g.,  $N_{side} = 2048$  for Planck) to obtain one patch. Repeat the same process for each pixel of the low resolution map to obtain the full set. However, these patches are not suited for analytical computations of the integrated bispectrum because they do not have a simple expression in multipole space (they do not allow removing  $m$ -dependence in the final integrated bispectrum expression).

<sup>19</sup>Healpy documentation <https://healpy.readthedocs.io/en/1.15.2/>

This is why, except for some results on CMB simulations (both Gaussian and non-Gaussian) cited from previous works [18] [19], we will not use this approach in the rest of the work (except for some tests in monopole estimation 5.3.2) and will instead focus on the azimuthally-symmetric, multipole space patches introduced in 5. Their simple definition in multipole space (5.1) can be translated into real space maps using eq. (4.33), requiring to compute Legendre polynomials for every pixel of the map. This takes significantly longer than the HEALpix patches approach, but the computational time remains still very reasonable, even more since in the new implementation of pipeline only one patch will be needed.

An important point to notice is that multipole space patches have non-zero values everywhere in the sky (the patch window is not a step function in pixel space), thus there is not a unique number of different patches leading to a complete, isotropic coverage of the sky. A good method to obtain such a coverage is to use a number of multipole patches again corresponding to the number of pixels in a low resolution HEALPix map (we will use  $N_{side} = 4, 8, 16$  in our analysis, corresponding to 192, 768 or 3072 pixels) and to center each of them at the center of a corresponding HEALPix pixel patch.

In previous works [19], [18], computing patches in this way was typically long time requiring, as the following step of the pipeline. Therefore, it was important to save real space maps for each patch, when simultaneously analyzing many data maps. However this task is avoided with the new implementation we are presenting here.

In the case of slow pipeline 5.4.1 we need only one patch which is rotated over the sky, corresponding to each pixel center position of a given  $N_{side}$  pixelization. Here, in order to speed it up, one could compute both power spectrum and monopole of all  $N_{maps}$  in a single patch position in the sky, and then change the pixel position rotating the patch. In this way the number of patch rotations are reduced from  $N_{patches} \times N_{maps}$  to only  $N_{maps}$  patch rotations (see 5.5). Having a lot of memory space to store even high number of patches would lead to a further improvement in computational time of this pipeline.

In the fast pipeline case 5.4.2 only one patch computation is required since the flat sky projection will be the same in every direction, thus we can use the same flat patch to mask each flat-sky projection of the weak lensing map. However, this procedure will give rise to accuracy issues in monopole estimation 5.3.2.

#### 5.1.4.b Types of patches

Here we discuss the different types of patches that can be used in this analysis: healpix patches are well defined in real space; multipole space patches like step-function patches or needlets which are well defined in harmonic space. In figure 5.3 the shape of HEALPix patches is shown, with a defined support in real space, covering the entire sphere with a well defined number of them.

To exploit the simplification in computing the theory signal (4.30) we instead use azimuthally symmetric patches having the following simple  $m$ -dependence:

$$w_{\ell m}(\hat{\Omega}_0) = w_{\ell} Y_{\ell m}^*(\hat{\Omega}_0). \quad (5.1)$$

Simple step-function windows for multipole space patches, extensively used in [18], having

$$w_{\ell} = 1 \quad \text{if} \quad \ell < \ell_{max} \quad \text{and} \quad w_{\ell} = 0 \quad \text{otherwise}, \quad (5.2)$$

However, our goal is to look at the squeezed configurations of the bispectrum only, through the integrated bispectrum estimator. In the case of [18], strong localization in real space was not necessary and in the end step-function windows were adopted. Here, we need a stronger localization in real space to implement the flat-sky projection, so we make use of Mexican needlet patches, which are defined by

$$w_\ell = \left[ \frac{\ell(\ell+1)}{B^{2j}} \right]^p e^{-\frac{\ell(\ell+1)}{B^{2j}}} \quad (5.3)$$

where  $B$ ,  $j$  and  $p$  are the needlet parameters determining its properties both in real and harmonic space [46]. For more details about needlet patches see Appendix B.

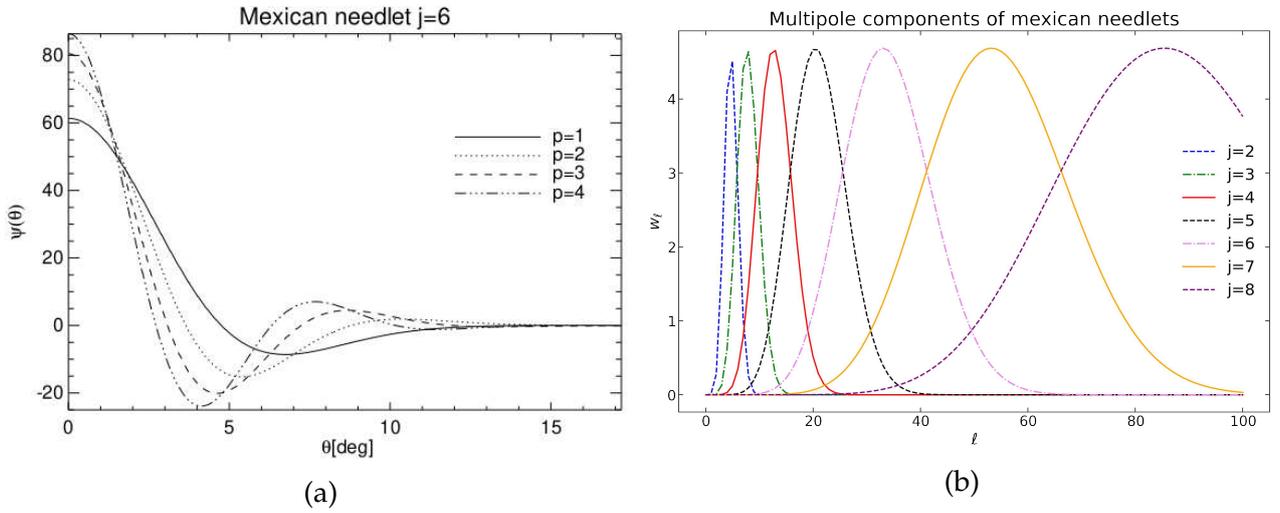


Figure 5.4: Mexican needlets patches, localization in real space (left) and in multipole space (right) for different needlets parameters  $p$  and  $j$ .

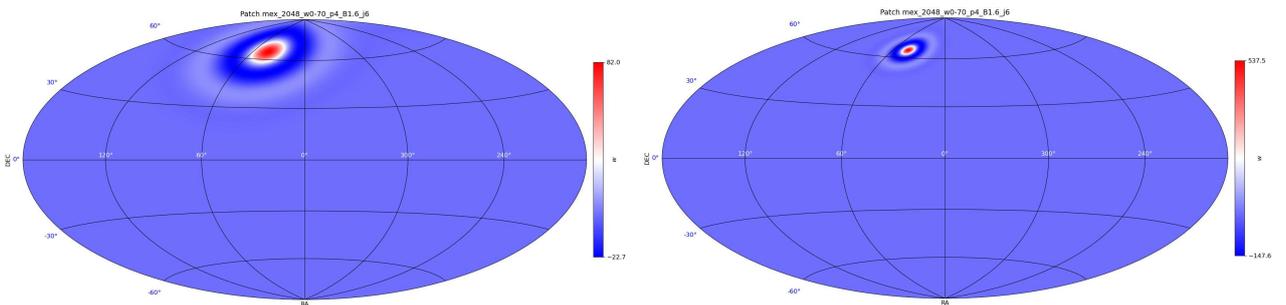


Figure 5.5: Mexican needlets patches with different patch windows and needlet parameters.

We tested the pipeline using a set of mexican needlet patches with increasingly real space localization<sup>20</sup>, characterized by different patch window function and needlets parameters:

<sup>20</sup>The angular width of a multipole space patche is proportionalto  $\ell_w^{-1}$ , thus an higher maximum multipole corresponds to an higher localization in real space.

$\ell_w$	$p$	$B$	$j$
[0, 20]	4	1.6	4
[0, 30]	4	1.6	5
[0, 70]	1	1.6	6
[0, 80]	1	1.6	7
[0, 100]	1	1.6	8

Table 5.1: Mexican needlets patches parameters, used in the analysis of the pipeline.

### 5.1.4.c Optimal number of patches

Each mexican needlet and multipole space patch, as introduced in the previous section, is defined by its central position  $\hat{\Omega}_0$  and the corresponding window function in harmonic space, see eq. (5.1). Such a patch is the largest around its center in a region where resides most of its constraining power, but it is also non-zero in the rest of the sky. Then, one needs to use enough of these patches, with a uniform distribution over the sky of their central positions, to obtain an isotropic coverage of the sky because it is important not to introduce any additional anisotropy. Unlike the HEALPix patches, where the sky is divided into a certain number of equal-sized areas and thus directly giving the number of required patches, this number is not so straightforward to get for multipole space patches<sup>21</sup>.

We recall that the method we used for the uniform repartition of their centers was simply to place each of them at the center of a corresponding HEALPix patch. This implies that we have to use a number of patch  $N_{\text{patches}}$  corresponding to the number of pixels of a low resolution map ( $N_{\text{patches}} = 192, 768, 3072$ ). It is important to keep  $N_{\text{patches}}$  as small as possible because it is proportional to the amount of computations: one power spectrum, flat projections and patch rotations to be estimated per patch, see equations (5.5) (5.6) for the precise pipeline description; while of course still getting optimal results in agreement with the theory. The effect of not using enough patches is roughly similar to a partial sky coverage in the sense it will also increase the variance because some parts of the sky are less scanned than others. In figure 5.6 we show the integrated bispectrum estimator, its standard deviation and the signal to noise ratio for different number of patches covering the sphere.

In the previous work [18] were used 192 multipole-space patches, in this work we need at least 768 patches since the flat sky approximation requires an higher localization in real space. Depending on the choice of mexican needlets parameters, we compute the integrated bispectrum estimator using 768 or 3072 patches over the sky, corresponding to  $N_{\text{side}} = 8, 16$  respectively. In figure 5.7 a plot of the sky coverage with two types of patches is shown.

### 5.1.4.d Realistic maps over the sky - Euclid mask

We analyze full-sky maps without noise, the ideal case, however the integrated bispectrum method can also be applied to more realistic observations. In [18] were used the same pseudo Euclid mask as in [32] which hides both the galactic and elliptic planes ( $f_{\text{sky}} = 0.35$ ). Then,

<sup>21</sup>The total angular width of a needlet patch is proportional to  $\ell_w^{-1}$ . This can be used to choose properly the projection width ( $2 \times \delta$ , see 5.1.5).

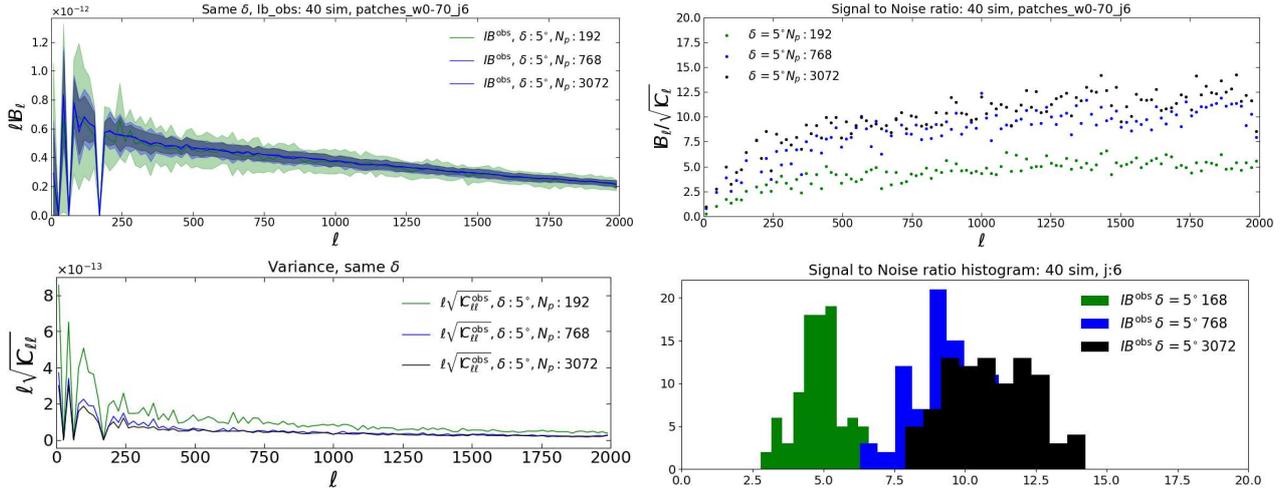


Figure 5.6: Estimator results using  $j = 6$  patches. The lower the number of patches the greater the observed variance. Quantities estimated with  $N_p = 192, 768, 3072$  respectively in green, blue, black.

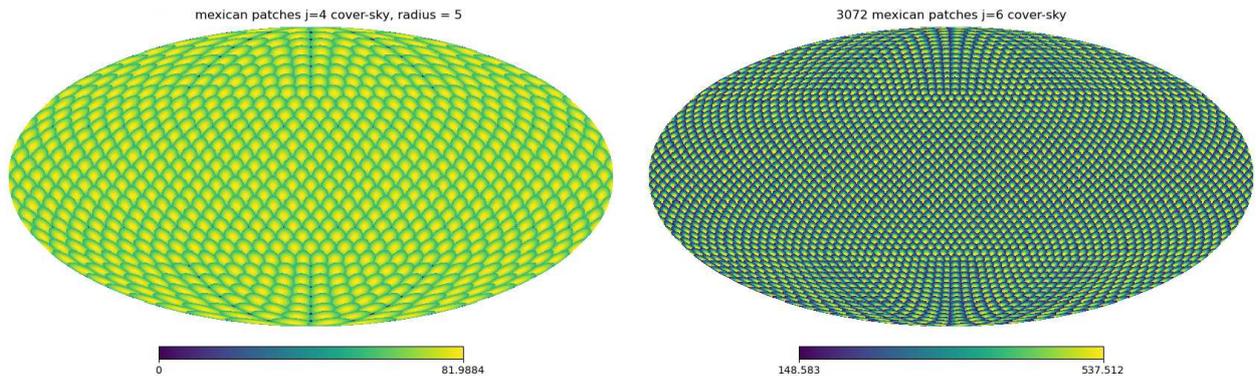


Figure 5.7: Patches sky coverage,  $j = 4$  and  $\delta = 5^\circ$  and  $N = 768$  patches (left),  $j = 6$  and  $\delta = 3^\circ$  and  $N = 3072$  patches (right). As we increase the patch localization an higher number of them is needed to cover the entire sphere.

were assumed Gaussian noise, with a noise power spectrum amplitude given by:

$$n_\ell = \frac{\sigma^2}{\bar{n}}$$

where the galaxy number density  $\bar{n}$  (typically  $\bar{n} = 30 \text{ arcmin}^{-2}$  for Euclid [25]) should be expressed in inverse steradians.

### 5.1.5 Rotations

The need to increase the numerar of patches to cover the entire sphere lead to a too high memory usage to store all patches. To avoid this we implement a map rotation, giving us the possibility to store only one and to rotate it in the desired direction.

In this way we copute each rotation using a rotator in harmonic space thanks to `rotate_map_alms()`

(5.1). Rotation acting on patch maps, which have a well defined support in multipole space, let us to fix a low  $\ell_{max}^{rot}$  (which would not possible with weak lensing maps) such that the time for one rotation is about 1 s for  $\ell_{max}^{rot} \leq 100$ .

Listing 5.1: Python code for full-sky map rotation

```

1 # Rotation def rotate_mask(patch, i, nsidepatches):
2     c=hp.pix2ang(nsidepatches, i, lonlat=True)
3     rot_patch=hp.Rotator(rot=c, inv=True)
4     patch_rotated=rot_patch.rotate_map_alms(patch, lmax=100)
5     return patch_rotated

```

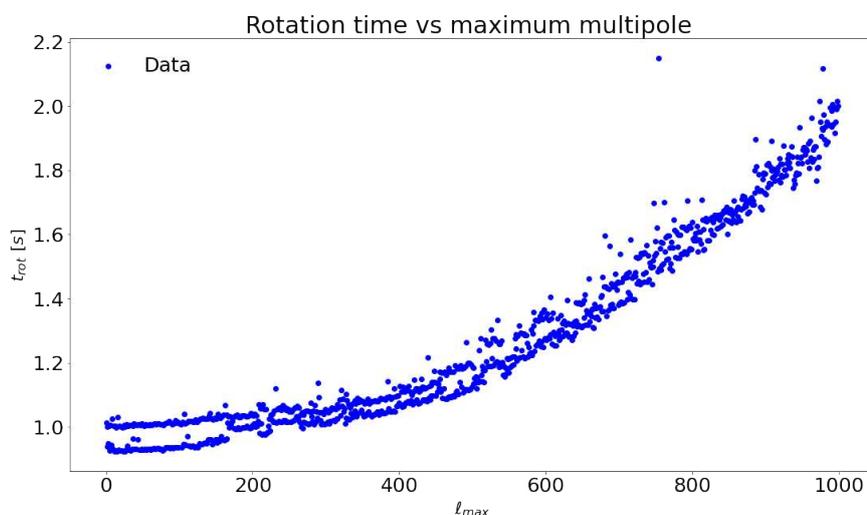


Figure 5.8: Rotation computational time tests.  $\ell_w^{max}$  on the horizontal axis, time on the vertical axis

### 5.1.6 Flat sky projections

In order to implement flat-sky power spectrum estimation, tangent plane projection of the full sky map are needed. `healpy` provides a class named `healpy.projector.CartesianProj` with a method `projmap` which, given the high resolution full-sky weak lensing map, returns the a 2D array with the projection of the map. Quantities such as the center of the projection  $c$ , number of pixel per each side  $n_{pix}$ <sup>22</sup>, angular size of the projection  $\delta$ <sup>23</sup> need to be specified. The larger angular size, the larger number of pixel is included in the projection, and so the ammount of information in a single flat map; however this also results in a loss of precision due to the tangent plane projection. The number of pixel in each side of the squared flat map is determined by proportionality, considering the number of pixels contained in the full sky region defined by the angular size, in order to minimize the deformation effect due to the projection, and optimize the pixelization mapping from curved to flat patch. The optimal

<sup>22</sup>In the code this quantity is usually called *size* or *ngrid*.

<sup>23</sup>This is half of the total angle subtended by the projected region.

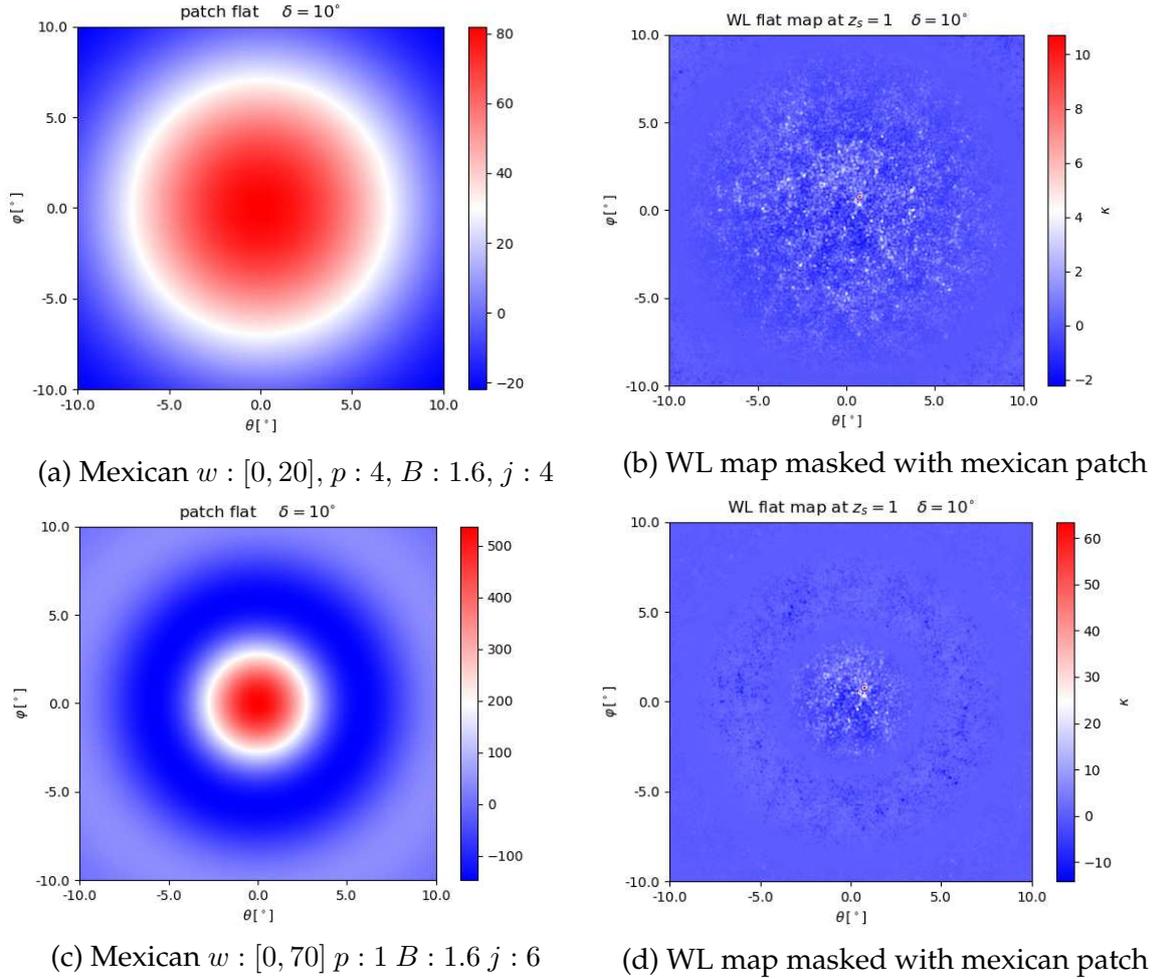


Figure 5.9: Cartesian projections with  $\delta = 10^\circ$ . Mexican needlet patches (left), weak lensing convergence maps at  $z_s = 1$  masked with the corresponding patch on the left (right).

dimension of the projection is found by:

$$A_{proj} = (2\delta)^2 \quad r = \frac{n_{pix\ 2048}}{n_{pix\ sky}} \quad n_{pix}^{opt} = (r A_{proj})^{1/2} \quad (5.4)$$

Where  $n_{pix\ sky} = 41253$  is the solid angle of the entire celestial sphere, and  $n_{pix\ 2048} = 50331648$  is the number of pixel of a  $N_{side} = 2048$  healpy map. In table 5.2 are shown some values of the angular size that have been used in this work.

Listing 5.2: Python code for Cartesian projection

```

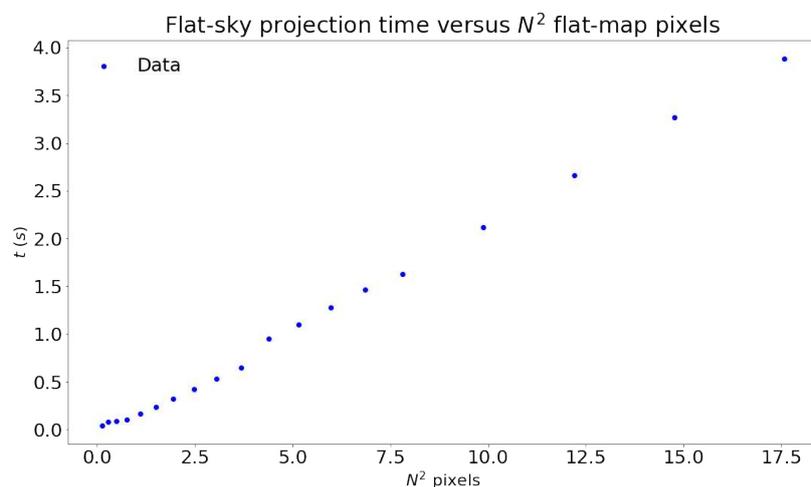
1 # cartesian projection python code
2 import numpy as np
3 import healpy as hp
4 from functools import partial
5 def get_flat_projection(map0, center, angle, ngrid):
6     angular_size = [-angle, angle]
7     proj_Cart = hp.projector.CartesianProj(rot=center, coord=G, xsize=ngrid,
8     ysize=ngrid, lonra=angular_size, latra=angular_size)
9     map0_flat = proj_Cart.projmap(map0, vec2pix_func=partial(hp.vec2pix, hp.
10 get_nside(map0)))
    
```

```
9 return map0_flat
```

On one side we believe that flat-sky approach will lead to speeding up of the pipeline, and this is proved to be true considering a single position power spectrum estimation (see 5.2). On the other side such an approach requires to project each localized patch into the tangent plane centered at HEALpix pixels centers, and this requires a certain amount of time depending on the dimension (in pixels) of the projected region; while with a full-sky approach this step was absent. Therefore we need to choose the size of the flat projection such that we gain the maximum computational time with respect to the former approach, without losing in precision in the power spectrum (and monopole) estimation.

### 5.1.6.a Projection computational time analysis

This tool scales linearly with respect to the area of the projection ( $N^2$  pixels). So it is necessary to keep the dimension of the projection as small as we still estimate the position dependent power spectrum with a good enough accuracy.



$\delta[^\circ]$	$n_{pix}^{opt}$	$t_{proj}$ [s]
5	349	0.050
7.5	524	0.062
10	699	0.074
12.5	873	0.104
15	1048	0.168
⋮	⋮	⋮
⋮	⋮	⋮
⋮	⋮	⋮

Figure 5.10: Healpy Cartesian Projection Time as function of the number of pixels in the flat map (million pixels).

Table 5.2: Angular size of the projection, optimal number of pixel per side and corresponding time required for the flat projection (right).

## 5.2 Power spectrum estimation

In this section we discuss three power spectrum estimators that have been used in this and previous works. The full-sky estimator `anafast`, implemented in the `healpy` library, is the most precise of the three we analyze here, with the disadvantage to be quite slow on high resolution maps, as we are interested on multipoles up to  $\ell_{max} = 2000$ . Instead of the costly full-sky power spectrum estimator required with the step function patches used in [18], now we exploit much more localized patches based on needlets (mexican), see 5.5. With this choice, it is then possible compute power spectra using the flat-sky approximation in each small, localized patch. This a priori should lead to large gains, both because the total number of pixels per patch is reduced by a factor  $\sim 10^2$  and because of the improved

computational scaling, allowed by FFT operations in flat-sky approximation. Here we discuss the implementation and test the time scaling of these estimators. We test the one implemented in the `pymaster` library and a second one developed by Gabriel Jung (A), both of which lead to some loss of precision at high  $\ell$  that will be discussed in 5.2.3, but a gain in computational time. We then tested the needlets parameters and window function to find best choice of needlets to recover the most precise position dependent power spectrum with respect to full sky estimation.

## 5.2.1 Full-sky

Anafast computes the power spectrum of a Healpix map, at each multipole  $\ell \in [l_{min}, l_{max}] = [2, 2000]$ ,  $l_{min} = 2$  since no removal of monopole or dipole is performed<sup>24</sup>. Within the full-sky approach and map resolution, as  $l_{max} = 2000$  each power spectrum takes  $\sim 10$  s to evaluate on a recent 8-core processor.

Listing 5.3: Python anafast estimator

```
1  ## Anafast power spectrum estimator
2  import healpy as hp
3  cl = hp.anafast(patched_map, lmax=lmax)
```

## 5.2.2 Flat-sky

Flat-sky estimators compute the power spectrum of a cartesian map, introducing a binning in multipole space in order to optimize the precision. For theoretical description of flat-sky<sup>25</sup> weak lensing convergence field power spectrum see 4.1.1.

### 5.2.2.a Binning

We will discuss the flat-sky pseudo- $C_\ell$  algorithm starting from a pixelized representation of the sky map. Let the patch of sky under inspection be contained by a rectangle of sides  $L_x$  and  $L_y$  (in units of radians), and let us discretize this rectangle by dividing it into an  $N_x \times N_y$  grid with pixels of area  $(\Delta x)^2 = \Delta x \Delta y (L_x/N_x)(L_y/N_y)$ . Each pixel in this grid is then labelled by a pair of integers  $n = (n_x, n_y)$ , and is assigned coordinates  $x_n = (n_x \Delta x, n_y \Delta y)$ . The each component of the pixelized map  $a(x)$  is therefore defined for  $n_x \in [0, N_x - 1], n_y \in [0, N_y - 1]$ .

Listing 5.4: Python code for binning

```
1  def get_ells_flat(angular_size, size):
2
3      Lx = 2*angular_size * np.pi/180.
4      Ly = 2*angular_size * np.pi/180.
5
6      l0_bins = np.arange(size) * np.pi/Lx
7      lf_bins = (np.arange(size)+1) * np.pi/Lx
```

<sup>24</sup>The input maps must be in ring-ordering. Spherical harmonics transforms are always on the full sky, if the map is masked, those pixels are set to 0. It is recommended to remove monopole from the map before running anafast to reduce boundary effects.

<sup>25</sup>Namaster flat-sky power spectrum estimation documentation is [https://namaster.readthedocs.io/en/latest/\\_downloads/bfcb0b7e72dd55ffaebc483ac560ab10/doc\\_scientific.pdf](https://namaster.readthedocs.io/en/latest/_downloads/bfcb0b7e72dd55ffaebc483ac560ab10/doc_scientific.pdf).

## 5.2. Power spectrum estimation

```

8  b = nmt.NmtBinFlat(10_bins, lf_bins)
9
10 e_ells = b.get_effective_ells()
11 return e_ells

```

In figure 5.11 we show the different bins in multipole space used for different angular size of the projection.

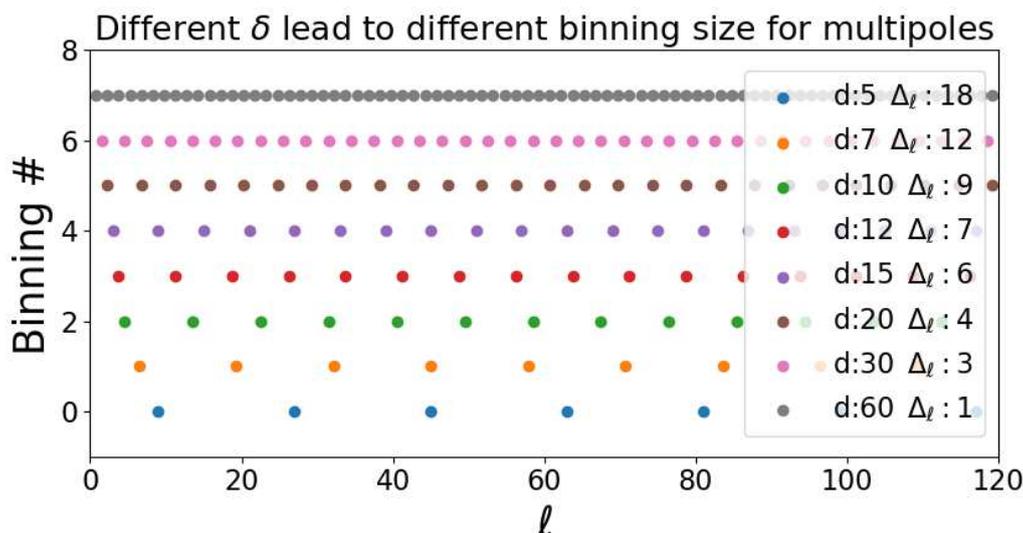


Figure 5.11: Binning examples, flat-sky power spectrum estimator. For example for an angular size equal  $10^\circ$  we have a bin-width of 9 in multipole space;  $\Delta_\ell = 1$  is recovered with<sup>a</sup>  $\delta = 60^\circ$  which however is very much computationally inefficient.

<sup>a</sup>d  $\equiv$   $\delta$  in the legend.

### 5.2.2.b Full-sky to flat-sky factor

Power spectra estimation in full-sky  $C_\ell$  and flat-sky need to be compared though an  $f_{sky}$  factor (4.42), since the selection of the squared region of pixel selected to be projected covers a small region on the sky. It is estimated as:

Listing 5.5: Python code proportionality factor

```

1  surface_sphere = 41253
2  surface_flat = (2*delta)**2
3  fac = surface_sphere/surface_flat

```

### 5.2.2.c Namaster

Flat sky power spectrum estimator, it computes the angular power spectra of two masked flat-sky fields (the same  $f_0$  in our case). Effectively, this is equivalent to computing the map FFTs and averaging over rings of wavenumber. The returned power spectrum is defined at the multipoles returned by the method `e_ells = b.get_effective_ells()`, so it is a binned estimator.

Listing 5.6: Python code for Namaster estimator

```

1  ## Namaster power spectrum estimator
2  import pymaster as nmt
3
4  def flat_sky_power_spectrum(map0, angular_size):
5      size=len(map0)
6      Lx = 2*angular_size * np.pi/180.
7      Ly = 2*angular_size * np.pi/180.
8      l0_bins = np.arange(size) * np.pi/Lx
9      lf_bins = (np.arange(size)+1) * np.pi/Lx
10
11     b = nmt.NmtBinFlat(l0_bins, lf_bins)
12     ells = b.get_effective_ells()
13
14     mask = np.ones_like(map0)
15     f0 = nmt.NmtFieldFlat(lx=Lx, ly=Ly, mask=mask, maps=[map0])
16     cl = nmt.compute_coupled_cell_flat(f0, f0, b)
17     return ells, cl[0]

```

The red line fit  $\propto N \log N$  confirms that FFTs implementations lead to a faster scaling of the estimators w.r. to anafast.

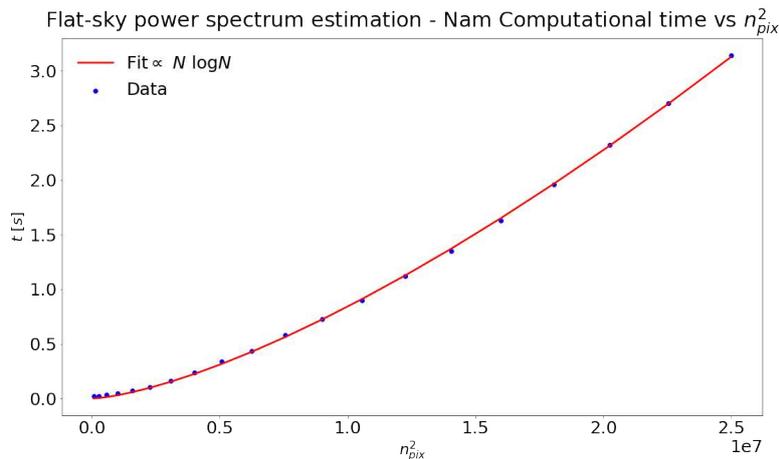


Figure 5.12: Time scaling of Namaster versus number of pixels (millions). Handmade estimator shows a similar behaviour

#### 5.2.2.d Handmade code

See appendix A

### 5.2.3 Power spectra results

In figure 5.13 flat-sky power spectrum estimators 5.2.2.c and 5.2.2.d are tested for different patches and different angular size of the projection ( $\delta = 5, 7.5, 10, 12.5, 15$ ) and compared

## 5.2. Power spectrum estimation

with the full-sky anafast estimator (green). We can see how healpix and mexican needlet with  $j = 6$  patches allow a good estimation also for smaller angular sizes of the projection, while mexican  $j = 4$  requires a larger region to recover the (reduced) full sky amplitude.

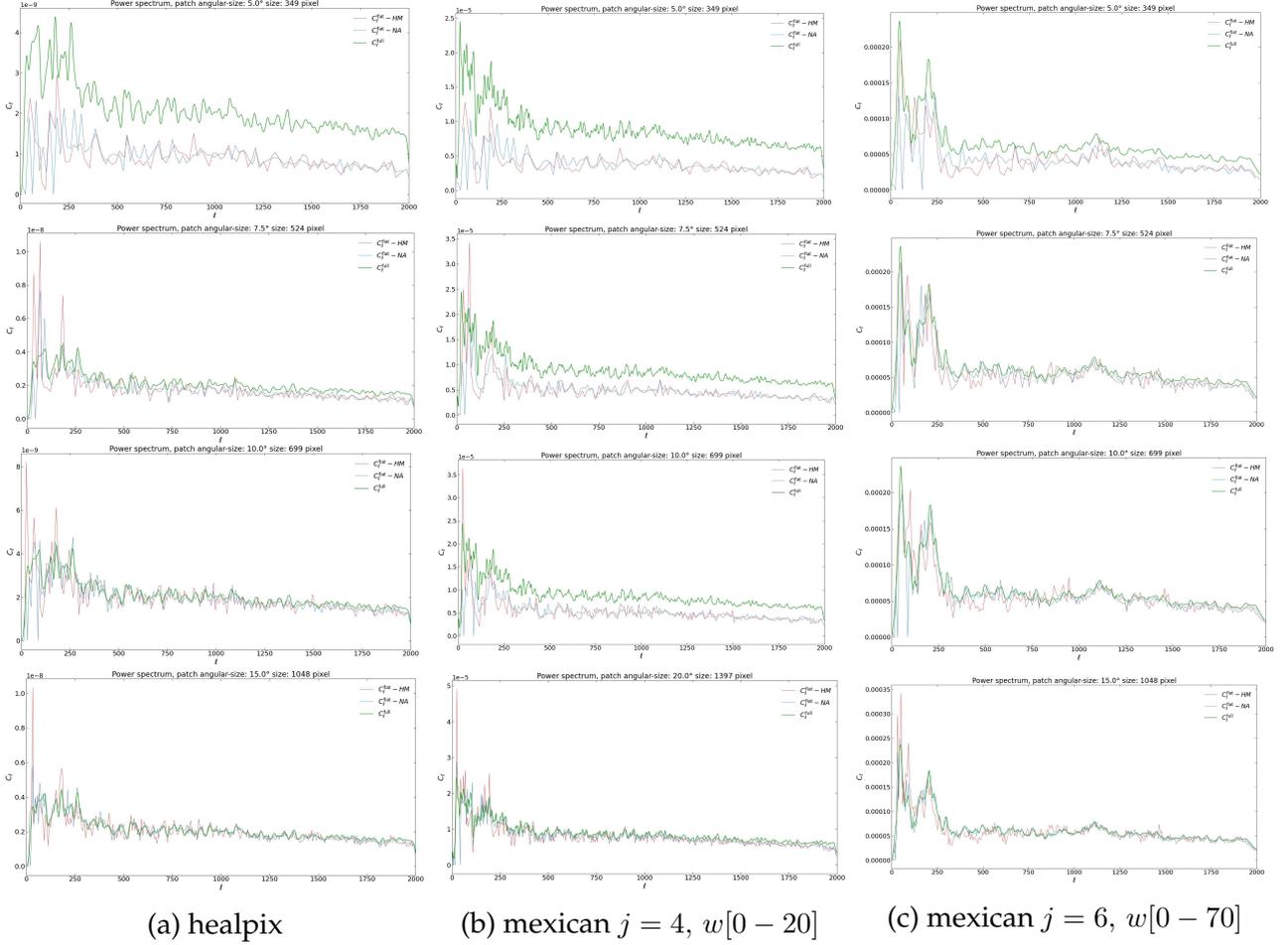


Figure 5.13: Namaster (blue) and HM (orange) power spectrum estimators tested for different types of patches, and compared with full-sky results (green). For healpix and mexican needlet with  $j = 6$   $C_\ell^{flat}$  traces the same amplitude of the full-sky counterpart  $\delta \geq 7.5^\circ$ <sup>a</sup>, while for mexican  $j = 4$  this happens for  $\delta \geq 20^\circ$

<sup>a</sup>Using  $p = 1$  instead of  $p = 4$  this is true also for smaller  $\delta$ .

## 5.3 Monopole term estimation

The second crucial quantity needed to estimate the integrated angular bispectrum is the monopole term, corresponding to the large scale modulation of the signal.

### 5.3.1 Full-sky

In the previous work [18] monopole term was evaluated simply averaging over the full-sky patched map array. In the case of 192 patches this procedure is easily implemented since each different patch can be stored in the memory. The application of the patch is done as a mask, thus multiplying map and mask arrays. Slow pipeline 5.4.1 uses this same monopole estimation procedure, exploiting patch rotation at each iteration.

```

1 # full sky monopole
2 mono_full = apply_mask(map0*patch).mean()
3 # or
4 mono_full = np.dot(map0,patch)/npix

```

### 5.3.2 Flat-sky

In this case there are some aspects to take into account. Evaluating the monopole term in flat sky using a mexican needlet brings the problem that non zero values outside the projected region are not taken into account in the average, giving a very different result from full sky value. In fact mexican needlets need to be very much localized to avoid this effect. Higher localization can be achieved selecting a larger patch window (4.18) as for example  $\ell_w \in [0, 75]$ . We tried to avoid this effect by making use of a circular mask over the patch mask. Such a mask is given by a circular one-valued mask centered at healpix pixel center with radius given by the angular size of the projection. In this way monopole evaluation improves, but how to take into account such circular mask in theory bispectrum signal is not trivial since it has not a well defined support in multipole space. Therefore this implementation has a further speed up, since no patch rotation is needed and the monopole estimation is made with two smaller arrays multiplication, however the cost its loss of precision makes us discard this pipeline implementation.

```

1 # flat sky monopole
2 mono_flat = apply_mask(map0_flat*patch_flat).mean()
3 # or

```

$n_{side}$	$t_0$	$t_1$	$t_2$
1024	0.061	0.030	0.0019
2048	0.242	0.104	0.0087
4096	0.968	0.384	0.0310

Table 5.3: Monopole estimation time for different methods of array multiplication and different array length.  $t_0$  refers to  $(v * v).mean()$ ,  $t_1$  refers to the same operation but with optimized array multiplication using *numpy*, and  $t_2$  refers to  $np.dot(v, v)/n_{pix}$ .

```
mono_flat = np.dot(map0, patch) / npix
```

## 5.4 Integrated bispectrum estimator: 2 pipelines

We have analysed the different parts of the pipeline, we now explore two different implementations.

The main step of the analysis is as simple as computing the monopole and the power spectrum of our masked flat map, multiplied by a patch function, and repeating the procedure for all previously determined patches. Hence, the amount of required computations is that of  $N_{\text{patches}}$  evaluations of a power spectrum, using the flat-sky estimator tool, and of  $N_{\text{patches}}$  tangent plane projections. In addition, as anticipated in the latter paragraph, we distinguish between two ways of estimating the monopole term: a slower but precise one using full-sky maps, and a faster but less precise one using flat-sky maps. The integrated bispectrum is then finally obtained by averaging the products between the monopole and the power spectrum in each patch.

### 5.4.1 Slow pipeline

Starting with a full sky weak lensing map and one needlet patch centered in  $c$ , the steps to compute  $N_{\text{patches}}$  power spectra are to project the patched map in flat-sky and compute the estimator (Namaster or HM); this requires one first cartesian projection of the patch map (that will be the same for each position in the sky). In order to estimate the monopole with high precision we need to average over the full-sky patched map, this requires the full sky patch to be rotated at each iteration, resulting an additional  $N_{\text{patches}}$  rotations.

Now, performing these rotations for each weak lensing map requires a very high amount of time as we need to analyze hundreds of simulations, since rotation needs  $\sim 1$  s each.

The optimized version, as anticipated in 5.1.4, reverses the loop over maps and patches in such a way that the  $N_{\text{patches}}$  rotation are performed only once, while we estimate the position dependent power spectrum and monopole for all maps in one direction in the sky at each iteration<sup>26</sup>, reducing the number of patch rotations from  $N_{\text{patches}} \times N_{\text{maps}}$  to only  $N_{\text{patches}}$ <sup>27</sup>.

In summary, the slow pipeline requires the following computations<sup>28</sup>:

$$t_{\text{ibisp}}^{\text{full}} = N_{\text{patches}} \times [N_{\text{maps}} \times (t_{\text{proj}}^{\text{map}} + t_{\text{pdps}} + t_{\text{mono}}^{\text{full}}) + t_{\text{rot}}] + t_{\text{proj}}^{\text{patch}} \quad (5.5)$$

Where the different  $t$  are specific computational times, depending on different quantities: For  $t_{\text{proj}}(\delta, n_{\text{grid}})$  see 5.2;  $t_{\text{pdps}}(\delta, n_{\text{grid}})$  see 5.2.2;  $t_{\text{mono}}^{\text{full}}(n_{\text{pix}}^{\text{full}})$  see 5.3.1;  $t_{\text{rot}}(\ell_{\text{max}})$  see 5.1.5.

<sup>26</sup>From this the name "parallel".

<sup>27</sup>Having a lot of memory space to store even high number of patches would lead to a further computational time improvement of this pipeline, about  $N_{\text{patches}}$  seconds.

<sup>28</sup>Instead of

$$N_{\text{maps}} \times N_{\text{patches}} \times (t_{\text{proj}}^{\text{map}} + t_{\text{pdps}} + t_{\text{mono}} + t_{\text{rot}}) + t_{\text{proj}}^{\text{patch}}$$

### 5.4.2 Fast pipeline

The steps are the same as the slow pipeline, except for the monopole estimation; it is computed averaging over the flat-sky patched map, thus there is not need for the  $N_{\text{patches}}$  full-sky patch rotation.

Therefore this fast pipeline requires the following computations

$$N_{\text{maps}} \times N_{\text{patches}} \times (t_{\text{proj}}^{\text{map}} + t_{\text{pdps}} + t_{\text{mono}}^{\text{flat}}) + t_{\text{proj}}^{\text{patch}} \quad (5.6)$$

Up to now the flat-sky estimation of monopole has not been enough precise in order to have results as precise as with the slow pipeline, therefore we do not delve any further in this method.

### 5.4.3 Computational time summary

The estimated times to run this pipeline for  $N_{\text{maps}} = 40$ ,  $N_{\text{patches}} = 768$ ,  $\delta = 10^\circ$  is about 75 minutes:

$$768 \times [40 \times (0.074 + 0.040 + 0.009) + 0.9] + 0.074 = 76 \text{ min.} \quad (5.7)$$

To have information about other angular sizes see table 5.4 and for other general considerations about time of computations see 6.1.

$\delta$	$n_{\text{pix}}$	$t_{\text{proj}} [\text{s}]$	$t_{\text{pdps}} [\text{s}]$
10	700	0.074	0.04
15	1050	0.17	0.09
20	1200	0.23	0.12
25	1350	0.31	0.16
30	2000	0.76	0.53
40	2500	1.32	1.01

Table 5.4: Time required for flat projection and power spectrum estimation (with namaster), for different size of the projection, with also  $t_{\text{mono}}^{2048} = 0,0087\text{s}$

# Chapter 6

## Results from All-sky Simulations

### 6.1 Introduction

Here we present 6.2 the signal to noise ratio for different patches (and sizes of the projection used); 6.3.1 how flat-sky implementation gives consistent results with the theoretical signal; 6.3.2 integrated bispectrum estimator results for higher  $\ell_w$ ; and 6.4 the expected computational time description.

### 6.2 Signal to noise ratio and Optimality

As we use more localized patches we have to check if  $N_{patches}$  used is high enough to lead to full-sky coverage of the celestial sphere (see 5.1.4.c). In figure 6.1 we show the observed integrated bispectrum, its standard deviation and the signal to noise ratio for two different patches ( $j = 6, 8$ ), estimated with angular size of the projection  $\delta = 5^\circ$ . We estimated these quantities using two different number of patches, 768 (left) and 3072 (right); we can notice how in the first case ( $j = 6$ ) we reach optimality already with  $N_p = 768$ , while for the higher localized patch ( $j = 8$ ) the higher number of patches leads to an improvement on the estimation, as can be easily seen in the  $S/N$  plot (bottom ones) in figure 6.1.

In figure 6.2 we show  $S/N$  from three different patches used; we can appreciate again how an higher number of patches leads to a better estimation of the observed signal, in particular up to  $j = 6$  a number of patches equal to 768 is sufficient and higher pixelization doesn't lead to further improvements, while for  $j = 7, 8$  we found that  $N_p = 3072$  improves significantly the signal to noise ratio. (we can not check the expected variance since the implementation for the variance is computationally too costly as  $\ell_w$  is too large.)

These behaviour can be understood remembering that different patches choices lead to different squeezed configuration entering in the IB estimator, and therefore differences in amplitude are expected. More in detail, we need also to consider the fact that  $j$  parameter determines the distribution of effective multipoles entering in the needlet patch (which determines also its angular width<sup>29</sup>), see eq. (5.4) and fig. 5.4b.

---

<sup>29</sup>The effective patch width is  $\sim \pi/\ell_w^{max.}$ , but it is important to take into account that it is defined in a definite domain in multipole space while in real space it has support on the whole sphere.

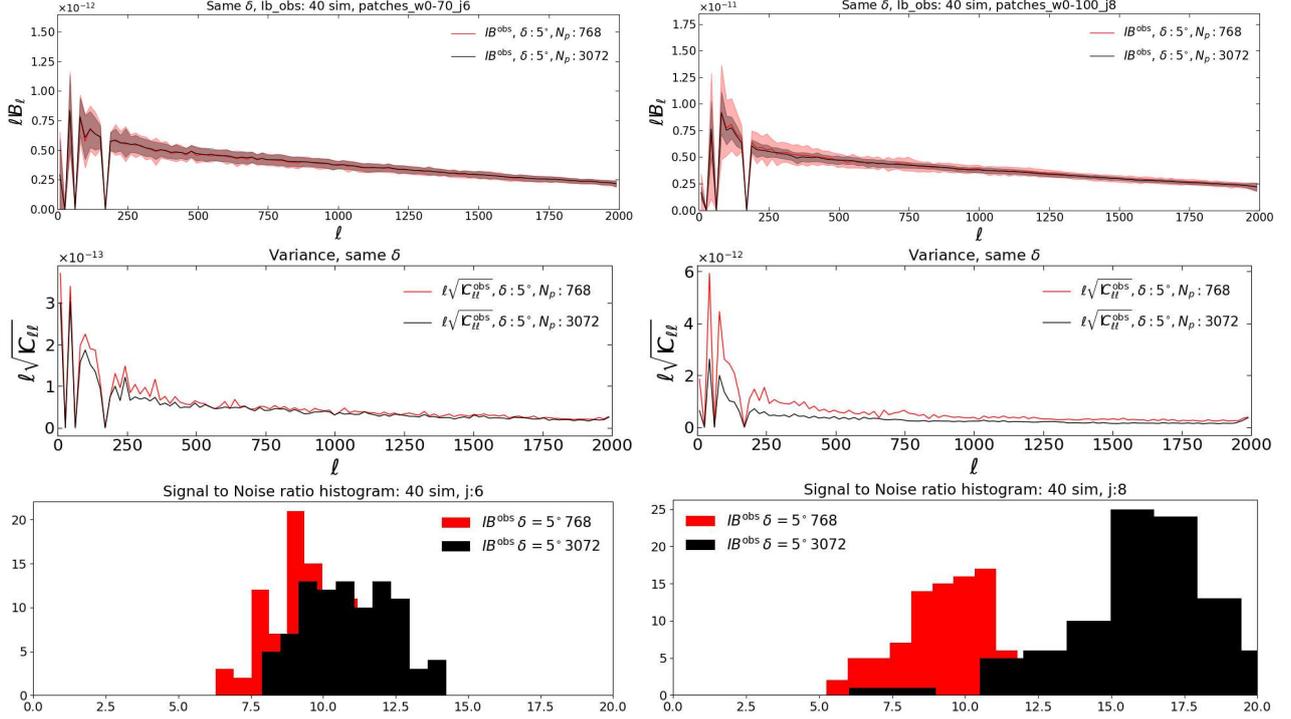


Figure 6.1: Observed integrated bispectrum, its standard deviation and the signal to noise ratio for two different patches ( $j = 6, 8$ ), estimated with angular size of the projection  $\delta = 5^\circ$ . Two different number of patches used, 768 (left) and 3072 (right). For ( $j = 6$ ) we reach optimality already with  $N_p = 768$ , while for the higher localized patch ( $j = 8$ ) the higher number of patches leads to an improvement on the estimation, indicated in  $S/N$  plots (bottom ones)

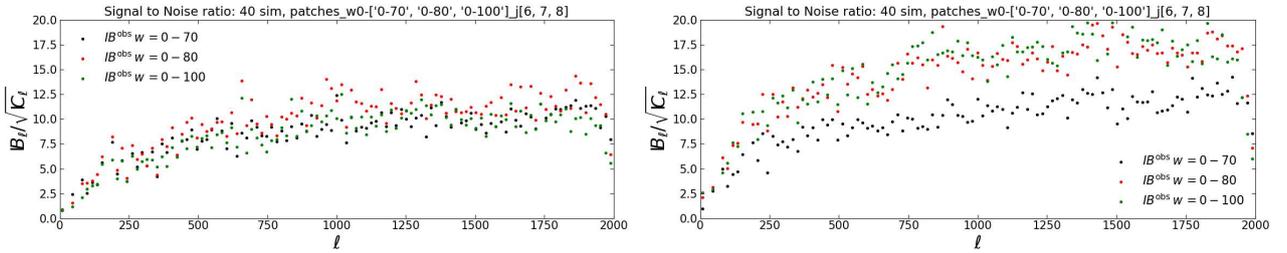


Figure 6.2:  $S/N$  ratio for a fixed  $\delta = 5^\circ$  and three different patches, 768 (left), 3072 (right). We can appreciate how an higher number of patches leads to a better estimation of the observed signal, in particular up to  $j = 6$   $N_{patches} = 768$  is sufficient and higher pixelization doesn't lead to further improvements, while for  $j = 7, 8$  we found that  $N_p = 3072$  improves significantly the  $S/N$  ratio.

## 6.3 Observed integrated bispectrum results

### 6.3.1 $IB_\ell^{obs}$ compared with theoretical signal

In this section we present the results obtained with pipeline 5.4.1; as first check we compare observed integrated bispectrum estimator results with theoretical signals (both with and without post-Born correction) and the standard deviations<sup>30</sup>; in figure 6.3a results from 40

<sup>30</sup>The observed signal and standard deviation is computed via Monte Carlo average over results from 40 simulations.

### 6.3. Observed integrated bispectrum results

simulations at  $z_s = 1$  using 768 mexican needlet patches ( $\ell_w \in [0, 20]$ ,  $j = 4$ ). At higher  $\ell_w$  theoretical signal estimation fails due to the too large amount of multipole triangular configurations to consider in eq (4.40).

As we discussed in 5.2.3 flat-sky power spectrum estimator reaches the desired accuracy, with respect to the full sky counterpart, for  $\delta \geq 20^\circ$ ; as you can see in figure 6.3b with this choice of patch we recover the expected (theoretical) amplitude for  $\delta \geq 20^\circ$  (while for more localized patches a smaller  $\delta$  can be chosen without losing in accuracy). The standard deviation obtained from 40 simulations is consistent with the Fisher matrix forecast (6.3c), it is to say that in this case with a quite large patch ( $j = 4$ ) the flat sky approximation is not as accurate as in the cases in the following paragraph. Figure (6.3d) shows clearly how binning (determined by different  $\delta$ ) does not influence the error bars.

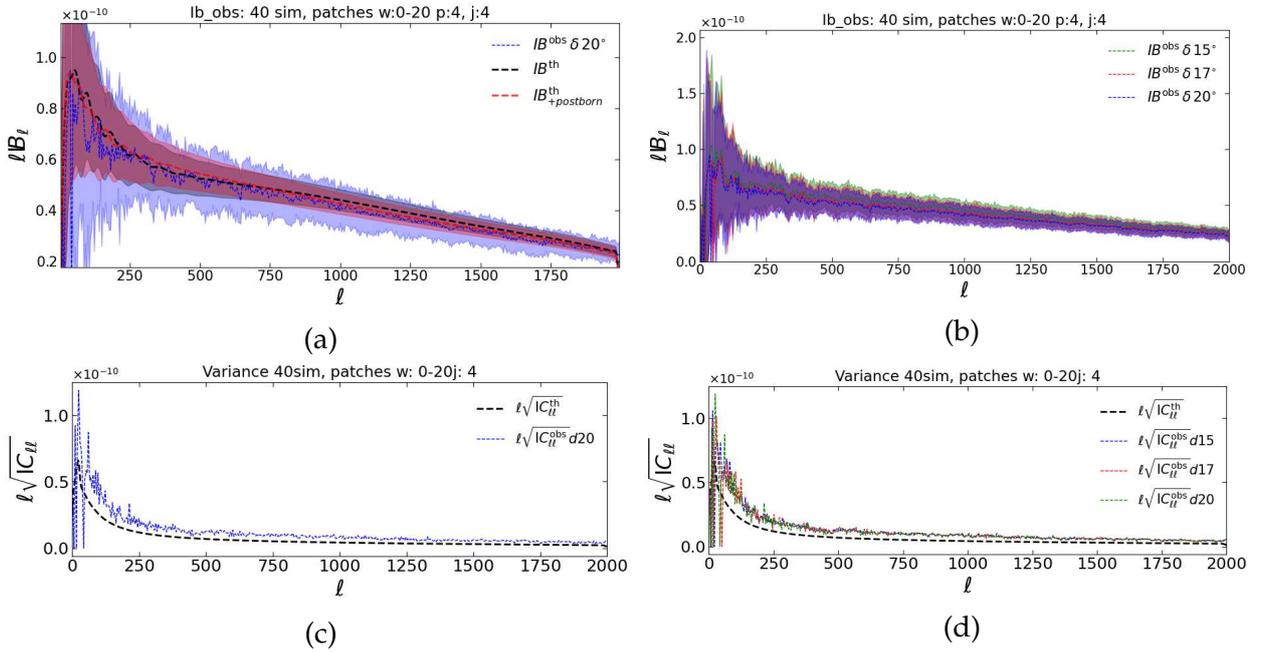


Figure 6.3: Observed integrated bispectrum estimator results with theoretical signals (both with and without post-Born correction) and the standard deviations from 40 simulations at  $z_s = 1$  using 768 mexican needlet patches ( $\ell_w \in [0, 20]$ ,  $j = 4$ ) (a). Observed integrated bispectrum estimator for different angular size of the projection (b). Standard deviation from 40 simulations and the Fisher matrix forecast (c). Standard deviation from different angular sizes of the cartesian projection compared (d).

#### 6.3.2 $IB_\ell^{\text{obs}}$ with more localized patches

Here we presents results for the remaining patches we have tested; as anticipated, now we do not show the theory signal since it is too costly computationally. In figure 6.4, 6.5 we show results for more localized patches ( $j = 5, 6$  and  $j = 7, 8$  respectively<sup>31</sup>); we see that different  $\delta$  lead to the same standard deviation of the estimator, so different angular sizes of the projection influence only the binning scheme (and the computational time), given that it is sufficiently large to have an accurate power spectrum estimation. We notice how this

<sup>31</sup>In the images for  $j = 6, 7, 8$  it is indicated  $p = 4$  while the results refers to  $p = 1$ , it's a typo.

happens with  $\delta \geq 12^\circ$  for  $j = 5$ , with  $\delta \geq 5^\circ$  for  $j = 6$  and with  $\delta \geq 3^\circ$  for  $j = 7, 8$  (while, as in the previous section, with  $\delta \geq 20^\circ$  for  $j = 4$ ).

Another effect we observe is the fact that for some values of  $\ell$  the observed signal has a fall in amplitude, depending on the angular size of the projection: it happens in some cases for  $j \geq 6$ , see figure 6.4, 6.5. However within the flat-sky approximation the estimator is more accurate for higher localization, corresponding to higher  $\ell$ .

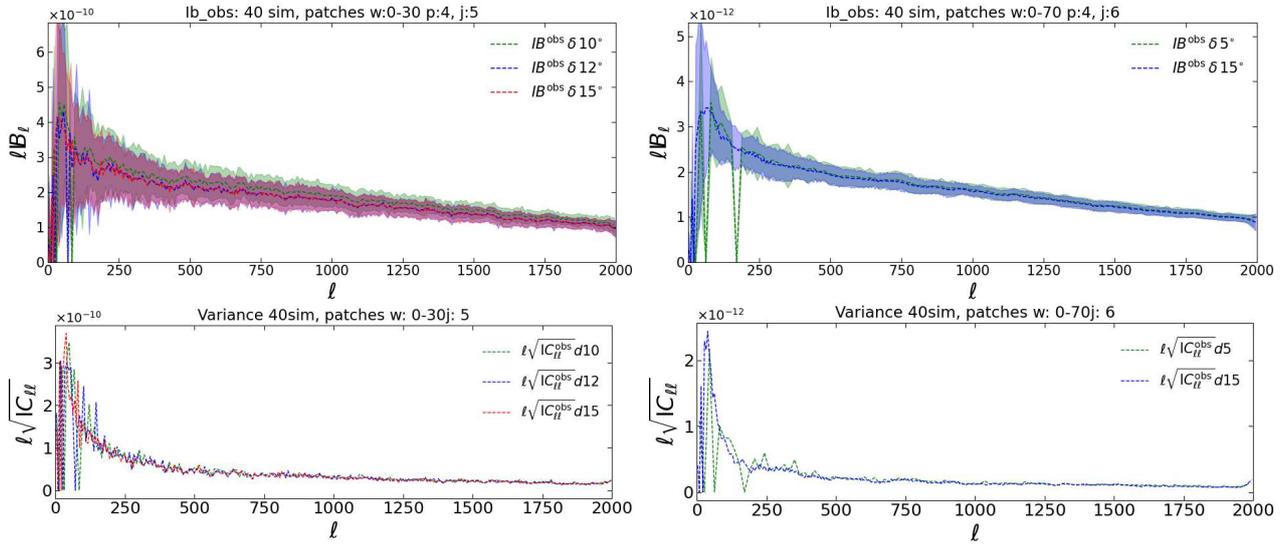


Figure 6.4: Observed integrated bispectrum estimator (top) and its standard deviation (bottom), for Mexican patches with parameters  $w[0 - 30], p = 4, B = 1.6, j = 5$  (left) and  $w[0 - 70], p = 1, B = 1.6, j = 6$  (right). Results with different  $\delta$  are represented in different colors.

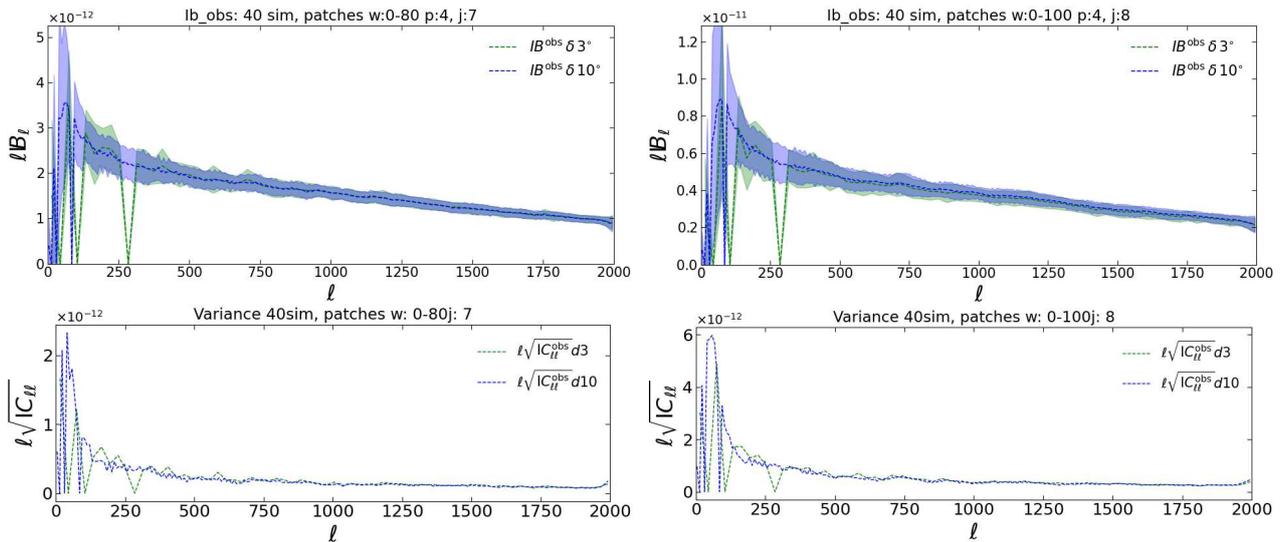


Figure 6.5: Observed integrated bispectrum estimator (top) and its standard deviation (bottom), for Mexican patches with parameters  $w[0 - 80], p = 1, B = 1.6, j = 7$  (left) and  $w[0 - 100], p = 1, B = 1.6, j = 8$  (right). Results with different  $\delta$  are represented in different colors.

## 6.4 Timing summary

Aiming to have a fast enough implementation to run with thousands of simulations, we now describe the expected computational time for the integrated bispectrum estimator within flat-sky approximation, we focus on the case of the slow <sup>32</sup> parallel <sup>33</sup> pipeline 5.4.1.

Recalling eq. (5.5) we can compute  $t_{ibsp}^{flat}$  summing the time contributions of different steps of the pipeline, multiplied by the number of patches used to cover the sky, and also for a given number of simulations. Important to notice is the fact that at small angular size of the projection the rotation time  $t_{rot}$  contributes significantly in the total time expected, while at an high number of maps its contribution becomes negligible.

Results show how the integrate bispectrum estimator pipeline with flat-sky approximation is valid if we choose an appropriate angular size of the cartesian projection (accordingly with the patch angular width); the smaller  $\delta$ , the smaller computational time is expected, however the former parameter influences also the binning width in multipole space, see 5.11, therefore we can play with it as we might be interested in estimating the covariance matrix with the minimum binning width computationally affordable). It is excluded to run with  $\delta$  too large both because the projection time becomes too large and the flat-sky approximation fails. In the following table we show the expected time to run the entire pipeline over 40 simulations, using different number of patches and three angular width as example.

$N_{patches}$	$t_{tot}^{\delta=5}$	$t_{tot}^{\delta=10}$	$t_{tot}^{\delta=15}$
768	35	76	151
3072	140	290	606

Table 6.1: Time to analyze 40 simulations, in minutes; for two sets of patches and three different angular size of the projection.

<sup>32</sup>Due to rotation needed to estimate the monopole term 5.3.1.

<sup>33</sup>At each iteration  $i = 1, 2, \dots, N_{Patches}$  computes the position dependent power spectrum estimator at  $\hat{\Omega}_i$  for all  $N_{maps}$ .



# Chapter 7

## Conclusions

### 7.1 This work

In this work, we implemented the flat-sky approximation for the position power spectrum estimation in the integrated angular bispectrum pipeline described in [18]. This was achieved by making use of a higher number of more localized patches on the spherical domain, which pick less squeezed triangles ( $\ell_3 \leq 100$ ); as the lensing signal is larger on smaller scales (where the Limber approximation is more accurate) this leads to a higher signal-to-noise ratio for the integrated signal. Thus this method measures a strong non-Gaussian signal without probing precisely the large-scale modes on which we focus in [18], but we find these two approaches complementary and addressing different questions.

Our tests, based on weak-lensing convergence simulations at redshift  $z_s = 1$ , show a good agreement between the measured non-Gaussian signal and its theoretical prediction (fig. 6.3a), where it is still computable ( $l_w \leq 20$ ). This is the ideal situation (full-sky, noiseless maps)<sup>34</sup>. In [18] was verified that the two methods (theoretical and computational) produce consistent results, displaying in both cases the same small mismatch between the measured and predicted non-Gaussian signal.

Within this new pipeline implementation we achieved the following improvement:

- Pipeline speed up of a factor  $\sim 10 - 100$ <sup>35</sup>, see (7.2).

and tests:

- Flat-sky position dependent power spectrum estimation and time scaling confirmed  $\propto N \log N$  (5.2).
- Computational time analysis for different dimension of the tangent plane projection (5.4.3).

---

<sup>34</sup>More realistic partial sky ( $f_{\text{sky}} = 0.35$ ) analyses, including Gaussian noise realizations, will be a further subject of tests.

<sup>35</sup>Depending on the number of patches used, and  $\delta$ , see 6.4

- Sky coverage with different HEALpix schemes and results comparison (5.1.4.c).
- Memory save, as only one patch is required (thanks to rotation) independently of  $N_{patches}$  used to cover the full sky. (5.1.5). 3072 patches with  $n_{side} = 2048$  would have been unaffordable to store in the memory.
- Single direction in the sky analyzed for multiple maps at the same iteration (5.4.2).

## 7.2 Towards Covariance matrix

Finally, we have explored the issue of how to precisely estimate the full integrated bispectrum covariance. This is an important point if we want to be able to use our integrated bispectrum pipeline for future applications, like cosmological parameter inference from weak lensing non-Gaussianity. In [19] was shown that, thanks to our choice of azimuthally symmetric patches, the covariance could be quickly evaluated by means of a simple semianalytical formula (4.40), valid in the weak non-Gaussianity limit of CMB analysis. However, such limit does not strictly apply to weak lensing and in [18] it was verified that this semi-analytical approach is no longer good enough for precise evaluation of the off-diagonal covariance terms (while it still holds quite well for the variance part<sup>36</sup>), see 4.2. The approach is then to evaluate the full covariance by running the estimator over many mock datasets, which requires developments to significantly speed up of the pipeline, which is the main objective of this work.

As example the total amount of time needed to run this estimator (considering  $\delta = 5^\circ$ ,  $N_p = 768$ ,  $N_{maps} = 4000$ <sup>37</sup>.) is  $\sim 40$  hours on a 8-core processor<sup>38</sup> (while it would have required 1000 hours<sup>39</sup> with the full-sky implementation, analyzing weak lensing simulations with 192 patches). Namely we have achieved a speed up of 25 times, factor which increases if we use  $\delta < 5^\circ$  and decreases otherwise. This is for a single CPU. This benchmark can be adjusted by considering a possible parallelization of the code. Assuming linear scaling (due to trivial parallelization, in which we assign independent mock realizations to different nodes), full-sky code [18] could be run in parallel over the  $N_{maps}$  whereas flat-sky slow pipeline can be run in parallel over  $N_{patches}$ . As example, considering  $\delta = 5^\circ$ ,  $N_p = 768$ ,  $N_{maps} = 4000$  and assuming to parallelize over 10 CPUs we expect  $t_{full}^{tot} = 100$  hours, and  $t_{flat}^{tot} = 4$  hours.

The choice of angular width of the projection allows both to gauge the computational time, both to change the optimal binning for the estimator (see 5.11), and accordingly the covariance binning.

<sup>36</sup>Since the actual expression for the Fisher matrix involves many computations,  $4(6) - j$  symbols, it is not computable for  $\ell_w \geq 20$ , one way to simplify its expression could be exploit the flat-sky approximation in order understand if there is a possibility to have a computable expression also at higher large-scale multipoles.

<sup>37</sup>Important to notice how the minimum number of simulations to be used to analyze is the number of modes considered, which is  $\ell_{max} = 2000$  for an unbinned estimator, while it is less for a binned one, as in our case; although in practice the idea is to analyze more simulations for numerical accuracy

<sup>38</sup>This comes out assuming that it is computationally possible to read all the map files and analyze them at the same time, requiring  $\sim 200 \text{ Mb} \times N_{map}$  of Ram; otherwise running over a smaller set of them in different times.

<sup>39</sup>Full-sky estimator takes  $\sim 15$  min for each map.

A further step would be to extend our method not only to estimate the integrated angular bispectrum of convergence  $\kappa$ , but also of the shear field  $(\gamma, \phi)$ <sup>40</sup>.

---

<sup>40</sup>Or equivalently  $(\gamma_1, \gamma_2)$ .



# Appendix A

## Handmade flat-sky power spectrum estimator - code

```
1  ###FFT
2  import numpy as np
3  import healpy as hp
4  from numba import njit, prange
5  import pyfftw
6  import matplotlib.pyplot as plt
7
8
9  #Fast fourier transform
10 def FFT2Dr(a, threads):
11
12     # align arrays
13     Ngrid = len(a)
14     a_in = pyfftw.empty_aligned((Ngrid, Ngrid), dtype="float32")
15     a_out = pyfftw.empty_aligned((Ngrid, Ngrid // 2 + 1), dtype="complex64")
16
17     # plan FFTW
18     fftw_plan = pyfftw.FFTW(
19         a_in,
20         a_out,
21         axes=(0, 1),
22         flags=("FFTW_ESTIMATE",),
23         direction="FFTW_FORWARD",
24         threads=threads,
25     )
26
27     # put input array into delta_r and perform FFTW
28     a_in[:] = a
29     fftw_plan(a_in, a_out)
30     return a_out
31
32
33 ### Pk and binning
34
35 # creating the multiple bin grid
36 @njit(parallel=True, cache=True)
37 def get_kgrid(kgrid, kmin2, kmax2):
38     ngrid = len(kgrid)
```

```

39 middle = ngrid // 2
40 for kxx in prange(ngrid):
41     kx = kxx - ngrid if (kxx > middle) else kxx
42     kx = int(kxx)
43     for kyy in range(middle + 1):
44         ky = kyy - ngrid if (kyy > middle) else kyy
45         ky = int(kyy)
46         if ky == 0 or (ky == middle and ngrid % 2 == 0):
47             if kx < 0:
48                 continue
49         val = kx * kx + ky * ky
50         if val < kmin2 or val > kmax2:
51             continue
52         else:
53             kgrid[kx, ky] = val
54
55
56 def get_num_bins(l, l_bins):
57     return np.searchsorted(l_bins[:-1], l, side="right") - 1
58
59
60 def get_bins(lbins, angle, ngrid):
61     box_size = 2 * angle * (np.pi / 180)
62     kF = 2.0 * np.pi / box_size
63     kbins = lbins / kF
64     kgrid = np.zeros((ngrid, ngrid))
65     get_kgrid(kgrid, kbins[0] ** 2, kbins[-1] ** 2)
66     kgrid = np.sqrt(kgrid)
67     kgrid_bins = get_num_bins(kgrid, kbins)
68     return kgrid_bins
69
70
71 @njit(cache=True)
72 def Pk_loop(map_k, kbins, k, Pk, nmodes, ngrid, middle):
73     for kxx in range(ngrid):
74         kx = kxx - ngrid if (kxx > middle) else kxx
75
76         for kyy in range(middle + 1): # kyy=[0,1,..,middle] --> ky>0
77             ky = kyy - ngrid if (kyy > middle) else kyy
78
79             # ky=0 & ky=middle are special (modes with (kx<0, ky=0) are not
80             # independent of (kx>0, ky=0): delta(-k)=delta*(+k))
81             if ky == 0 or (ky == middle and ngrid % 2 == 0):
82                 if kx < 0:
83                     continue
84
85             # compute |k| of the mode and its integer part
86             k_val = np.sqrt(kx * kx + ky * ky)
87             # Exclude cases not included in the binning range
88             k_index = kbins[kxx, kyy]
89             if k_index < 0:
90                 continue
91             # print(k_index)
92
93             # compute |delta_k|^2 of the mode
94             real = map_k[kxx, kyy].real

```

---

```

95     imag = map_k[kxx, kyy].imag
96     map_k2 = real * real + imag * imag
97
98     # Pk
99     k[k_index] += k_val
100    Pk[k_index] += map_k2
101    nmodes[k_index] += 1.0
102
103
104 def compute_pk(map0, angle, kbins, threads=1):
105
106     ngrid = len(map0)
107     middle = ngrid // 2
108
109     box_size = 2 * angle * (np.pi / 180)
110     kF = 2.0 * np.pi / box_size #maximum multipole
111
112     map_k = FFT2Dr(map0, threads=threads) # fourier transform the flat field (
113     map)
114
115     nbins = np.max(kbins) + 1
116     k = np.zeros(nbins, dtype=np.float64)
117     Pk = np.zeros(nbins, dtype=np.float64)
118     nmodes = np.zeros(nbins, dtype=np.float64)
119
120     # do a loop over the independent modes.
121     Pk_loop(map_k, kbins, k, Pk, nmodes, ngrid, middle)
122
123     for i in range(len(k)):
124         k[i] = (k[i] / nmodes[i]) * kF
125         Pk[i] = (Pk[i] / nmodes[i]) * (box_size / ngrid ** 2) ** 2
126
127     return k, Pk

```



# Appendix B

## Needlets Patches

In [46] different form of needlets are introduced. The introduction of Mexican needlets is due to Geller & Maveli [30]

The basic needlet function can be described in real space as follows:

$$\psi_{jk}(x) := \sqrt{\lambda_{jk}} \sum_{\ell} b_{\ell}(B, j) \sum_{m=-\ell}^{\ell} Y_{\ell m}(\xi_{jk}) \overline{Y_{\ell m}(x)} \quad (\text{B.1})$$

Here,  $x$  refers to a position  $(\theta, \phi)$  on the sphere,  $Y_{\ell m}$  are spherical harmonic functions,  $j$  is the scale (frequency) of the needlet and  $\{\lambda_{jk}\}$  is a set of cubature weights corresponding to the cubature points  $\{\xi_{jk}\}$ ; for simplicity, they can be taken to be equal to the pixel areas and the pixel centres in the HEALPix grid used for CMB analysis, i.e. we shall consider  $\lambda_{jk} = \lambda_j = 4\pi/N_j$ , where  $N_j$  is the number of pixel in the pixelization we are working with. The needlet function itself is contained in the function  $b_{\ell}(B, j)$  (or  $b_{\ell}$  for short) in harmonic space,  $B$  being one of the parameters deciding the properties of the needlet. The difference between the needlet systems we are going to discuss can thus be traced in the form of the weight function  $b_{\ell}$ .

1) **Standard needlets:** Let  $\phi(\xi)$  be an infinitely differentiable (i.e.,  $C^{\infty}$ ) function supported in  $|\xi| \leq 1$ , such that  $0 \leq \phi(\xi) \leq 1$  and  $\phi(\xi) = 1$  if  $|\xi| \leq 1/B$ ,  $B > 1$ . Define

$$b^2(\xi) = \phi\left(\frac{\xi}{B}\right) - \phi(\xi) \geq 0 \text{ so that } \forall \ell > B, \sum_{j=0}^{\infty} b^2\left(\frac{\ell}{B^j}\right) = 1$$

For standard needlets we then obtain  $b_{\ell}$  from this function  $b(\xi)$  by  $b_{\ell}(B, j) = b\left(\frac{\ell}{B^j}\right)$ . For a given scale  $j$ , the needlet function in harmonic space is centered at a multipole  $\ell^* \approx B^j$ . Thus a given scale  $j$  is mainly influenced by multipoles close to  $\ell^*$ . It is immediate to verify that  $b(\xi) \neq 0$  only if  $\frac{1}{B} \leq |\xi| \leq B$ . The main localization property of needlets is established in Narcowich et al. (2006a), where it is shown that for any  $M \in \mathbb{N}$  there exists a constant  $c_M > 0$  s.t., for every  $\xi \in \mathbb{S}^2$ :

$$|\psi_{jk}(\xi)| \leq \frac{c_M B^j}{(1 + B^j d(\xi_{jk}, \xi))^M} \text{ uniformly in } (j, k)$$

where  $d(\xi_{jk}, \xi)$  denotes the usual distance on the sphere. More explicitly, needlets are almost exponentially localized around any cubature point, which motivates their name.

2) **Bernstein Needlets:** The bound which we just provided to establish the localization properties of needlets depends on some constants  $c_M$  which we did not write down explicitly. Such constants depend on the form of the function  $b(\xi)$ , and turn out to be rather large in the case of standard needlets. We do no longer have quasi-exponentially decaying tails, however, but it is possible to establish a weaker result, namely the decay with a polynomial rate, depending on the number of bounded derivatives we are allowing for  $b(\xi)$ . It may hence seem that this construction should enjoy worse properties - but in practice this is not the case. As for standard needlets, we have  $b_\ell(B, j) = b\left(\frac{\ell}{B^j}\right)$ , but note that  $b(\xi)$  is different for Bernstein needlets. As for the standard needlets, the needlet function in harmonic space is centered at  $\ell^* \approx B^j$ .

3) **Mexican Needlets:** The construction in [30] is similar to standard needlets, insofar as a combination of Legendre polynomials with a smooth function is proposed; the main difference is that for standard needlets the kernel is taken to be compactly supported (i.e., depending only a finite number of multipoles  $\ell$ ), while the Mexican needlet construction draws information from all frequencies at any scale. More precisely, we shall consider weight functions  $b_\ell(B, j)$  of the form

$$b_\ell(B, j) = \left(\frac{\ell}{B^j}\right)^{2p} e^{-\frac{\ell^2}{B^{2j}}}$$

for  $p = 1, 2, 3, \dots$ . For instance, for  $p = 1$  the Mexican needlet takes the form

$$\psi_{jk;1}(x) = \sqrt{\lambda_j} \sum_{\ell \geq 1} \frac{\ell^2}{B^{2j}} e^{-\ell^2/B^{2j}} \frac{2\ell + 1}{4\pi} P_\ell(d(\xi_{jk}, \xi))$$

and for higher  $p$  we have

$$\psi_{jk;p}(x) := \sqrt{\lambda_{jk}} \sum_{\ell \geq 1} \left(\frac{\ell^2}{B^{2j}}\right)^p e^{-\ell^2/B^{2j}} \frac{2\ell + 1}{4\pi} P_\ell(d(\xi_{jk}, \xi))$$

Indeed, for mathematical rigour  $\ell^2$  should be replaced by the eigenvalue  $\ell(\ell + 1)$ , but for CMB data analysis the difference is negligible and we shall use  $\ell^2$  for notational simplicity. As mentioned before, Mexican needlets are not supported on a finite number of multipoles, so the discussion of their localization properties in the harmonic domain requires some care. Moreover, because we need to focus on an infinite number of spherical harmonics, from a strictly mathematical point of view exact cubature and reconstruction formulae cannot hold. Nevertheless, it must be added that the approach by [30] enjoys some undeniable strong points, some of which we list as follows:

1. Mexican needlets enjoy extremely good localization properties in the real domain; more precisely, at a fixed angular distance  $x$  their tails decay as  $\exp(-B^{2j}x^2/4)$ , as  $j$  grows to infinity.
2. By adjusting the parameter  $p$ , one has available a family of wavelets which can be optimized in terms of the desired localization properties (as we shall show below,

---

a growing  $p$  improves the localization in the harmonic domain and decreases the localization in the real domain)

3. The previously mentioned mathematical issues on the cubature points are largely negligible from a numerical point of view
4. The Monte Carlo evidence provided below proves that Mexican needlets compare favorably with standard needlets under a variety of circumstances and for many different indicators.

Concerning the last point, it is important to remark the following. It can be shown that Mexican needlets for  $p = 1$  provide a very close approximation of the widely popular Spherical Mexican Hat Wavelets (SMHW). Even in this case, though, the implementation through needlet ideas in our view yields important benefits:

- the weight function is explicitly given, making easier the implementation and the validation of numerical codes
- the localization structure in harmonic domain can be analytically studied and controlled
- the correlation structure of random Mexican needlet coefficients is explicitly given and can be used for statistical inference
- the range of scales to be considered to retain the information from the data is mathematically determined in terms of the frequencies  $j$ , rather than by an ad hoc choice of scales in the real domain as a function of angular distance.

Our results for the different types of needlets can be summarized as follows:

- **Standard needlets:** The standard needlets are much better localized in harmonic space than the Mexican needlets (but similar to Bernstein needlets). In fact, the contribution to a certain scale  $j$  is coming from a limited number of multipoles with no influence from multipoles outside this range. However, the penalty for high localization in harmonic space is that the real space localization is lower than for other needlets. The parameter  $B$  controls the localization properties: The higher the  $B$ , the higher is the real space localization and the larger are the multipole ranges included in each scale  $j$ .
- **Mexican needlets:** The Mexican needlets are much better localized in real space than any of the other types. In harmonic space however, a large (in principle infinite) multipole range contributes to each scale  $j$ . The Mexican needlets depend on the parameter  $p$ , larger  $p$  have worse real space localization properties but better harmonic space localization. For  $p = 3 - 4$ , the real space localization properties of the Mexican needlets approach those of standard needlets for high values of  $B$ . For  $p = 1$  and high  $j$ , the Mexican needlets are almost identical to the Spherical Mexican Hat Wavelets.
- **Bernstein needlets:** Bernstein needlets are in between the Mexican and standard needlets. Their real and harmonic space localization is similar to the standard needlets, sometimes it is slightly better, sometimes it is slightly worse. The Bernstein needlets depend both on the parameter  $B$  and (weakly) on a parameter  $k$ . As for the standard needlets, a higher value of  $B$  increases the number of multipoles included in each scale

$j$  and improves the real space localization. The parameter  $k$  may slightly improve or worsen the localization properties depending on the exact measure used. As an example we find that for the critical angle for a galactic cut, for  $k \geq 2$  the angle for the less stringent thresholds increases for increasing  $k$ , whereas the angle for the more stringent thresholds decreases for increasing  $k$ .

# Bibliography

- [1] Neil Barnaby, Ryo Namba, and Marco Peloso. Observable non-Gaussianity from gauge field production in slow roll inflation, and a challenging connection with magnetogenesis. , 85(12):123523, June 2012.
- [2] N. Bartolo, E. Komatsu, S. Matarrese, and A. Riotto. Non-Gaussianity from inflation: theory and observations. , 402(3-4):103–266, November 2004.
- [3] Francis Bernardeau, Camille Bonvin, Nicolas Van de Rijt, and Filippo Vernizzi. Cosmic shear bispectrum from second-order perturbations in general relativity. , 86(2):023001, July 2012.
- [4] Roger Blandford and Ramesh Narayan. Fermat’s Principle, Caustics, and the Classification of Gravitational Lens Images. , 310:568, November 1986.
- [5] Martin Bucher, Carla Sofia Carvalho, Kavilan Moodley, and Mathieu Remazeilles. CMB lensing reconstruction in real space. , 85(4):043016, February 2012.
- [6] Clare Burrage, Raquel H. Ribeiro, and David Seery. Large slow-roll corrections to the bispectrum of noncanonical inflation. , 2011(7):032, July 2011.
- [7] Chi-Ting Chiang, Christian Wagner, Fabian Schmidt, and Eiichiro Komatsu. Position-dependent power spectrum of the large-scale structure: a novel method to measure the squeezed-limit bispectrum. , 2014(5):048, May 2014.
- [8] Asantha R. Cooray. Cosmological parameters from statistics of strongly lensed radio sources. , 342:353–362, February 1999.
- [9] Paolo Creminelli, Alberto Nicolis, Leonardo Senatore, Max Tegmark, and Matias Zaldarriaga. Limits on non-Gaussianities from WMAP data. , 2006(5):004, May 2006.
- [10] Paolo Creminelli and Matias Zaldarriaga. CMB 3-point functions generated by nonlinearities at recombination. , 70(8):083532, October 2004.
- [11] Emanuela Dimastrogiovanni, Nicola Bartolo, Sabino Matarrese, and Antonio Riotto. Non-Gaussianity and Statistical Anisotropy from Vector Field Populated Inflationary Models. *Advances in Astronomy*, 2010:752670, January 2010.
- [12] F. W. Dyson, A. S. Eddington, and C. Davidson. A Determination of the Deflection of Light by the Sun’s Gravitational Field, from Observations Made at the Total Eclipse of May 29, 1919. *Philosophical Transactions of the Royal Society of London Series A*, 220:291–333, January 1920.

- [13] Duncan Hanson, Kendrick M. Smith, Anthony Challinor, and Michele Liguori. CMB lensing and primordial non-Gaussianity. , 80(8):083004, October 2009.
- [14] David Harvey. A 4 per cent measurement of  $H_0$  using the cumulative distribution of strong lensing time delays in doubly imaged quasars. , 498(2):2871–2886, October 2020.
- [15] Seyed-Yashar Hezavehe. *Strongly Lensed High Redshift Dusty Star Forming Galaxies Discovered in Wide Area Millimetre Surveys*. PhD thesis, McGill University, Canada, April 2013.
- [16] Wayne Hu. Weak lensing of the CMB: A harmonic approach. , 62(4):043007, August 2000.
- [17] Niall Jeffrey, Justin Alsing, and François Lanusse. Likelihood-free inference with neural compression of DES SV weak lensing map statistics. , 501(1):954–969, February 2021.
- [18] Gabriel Jung, Toshiya Namikawa, Michele Liguori, Dipak Munshi, and Alan Heavens. The integrated angular bispectrum of weak lensing. , 2021(6):055, June 2021.
- [19] Gabriel Jung, Filippo Oppizzi, Andrea Ravenni, and Michele Liguori. The integrated angular bispectrum. , 2020(6):035, June 2020.
- [20] Nick Kaiser and Gordon Squires. Mapping the Dark Matter with Weak Gravitational Lensing. , 404:441, February 1993.
- [21] Nick Kaiser, Gillian Wilson, and Gerard A. Luppino. Large-Scale Cosmic Shear Measurements. *arXiv e-prints*, pages astro-ph/0003338, March 2000.
- [22] Mindaugas Karčiauskas, Konstantinos Dimopoulos, and David H. Lyth. Anisotropic non-Gaussianity from vector field perturbations. , 80(2):023509, July 2009.
- [23] Issha Kayo, Masahiro Takada, and Bhuvnesh Jain. Information content of weak lensing power spectrum and bispectrum: including the non-Gaussian error covariance matrix. , 429(1):344–371, February 2013.
- [24] Edward W. Kolb and Michael S. Turner. *The early universe*, volume 69. 1990.
- [25] R. Laureijs, J. Amiaux, S. Arduini, J. L. Auguères, J. Brinchmann, R. Cole, M. Cropper, C. Dabin, L. Duvet, A. Ealet, B. Garilli, P. Gondoin, L. Guzzo, J. Hoar, H. Hoekstra, R. Holmes, T. Kitching, T. Maciaszek, Y. Mellier, F. Pasian, W. Percival, J. Rhodes, G. Saavedra Criado, M. Sauvage, R. Scaramella, L. Valenziano, S. Warren, R. Bender, F. Castander, A. Cimatti, O. Le Fèvre, H. Kurki-Suonio, M. Levi, P. Lilje, G. Meylan, R. Nichol, K. Pedersen, V. Popa, R. Rebolo Lopez, H. W. Rix, H. Rottgering, W. Zeilinger, F. Grupp, P. Hudelot, R. Massey, M. Meneghetti, L. Miller, S. Paltani, S. Paulin-Henriksson, S. Pires, C. Saxton, T. Schrabback, G. Seidel, J. Walsh, N. Aghanim, L. Amendola, J. Bartlett, C. Baccigalupi, J. P. Beaulieu, K. Benabed, J. G. Cuby, D. Elbaz, P. Fosalba, G. Gavazzi, A. Helmi, I. Hook, M. Irwin, J. P. Kneib, M. Kunz, F. Mannucci, L. Moscardini, C. Tao, R. Teyssier, J. Weller, G. Zamorani, M. R. Zapatero Osorio, O. Boulade, J. J. Foumond, A. Di Giorgio, P. Guttridge, A. James, M. Kemp, J. Martignac, A. Spencer, D. Walton, T. Blümchen, C. Bonoli, F. Bortoletto, C. Cerna, L. Corcione,

- C. Fabron, K. Jahnke, S. Ligorì, F. Madrid, L. Martin, G. Morgante, T. Pamplona, E. Prieto, M. Riva, R. Toledo, M. Trifoglio, F. Zerbi, F. Abdalla, M. Douspis, C. Grenet, S. Borgani, R. Bouwens, F. Courbin, J. M. Delouis, P. Dubath, A. Fontana, M. Frailis, A. Grazian, J. Koppenhöfer, O. Mansutti, M. Melchior, M. Mignoli, J. Mohr, C. Neissner, K. Nodde, M. Poncet, M. Scodiggio, S. Serrano, N. Shane, J. L. Starck, C. Surace, A. Taylor, G. Verdoes-Kleijn, C. Vuerli, O. R. Williams, A. Zacchei, B. Altieri, I. Escudero Sanz, R. Kohley, T. Oosterbroek, P. Astier, D. Bacon, S. Bardelli, C. Baugh, F. Bellagamba, C. Benoist, D. Bianchi, A. Biviano, E. Branchini, C. Carbone, V. Cardone, D. Clements, S. Colombi, C. Conselice, G. Cresci, N. Deacon, J. Dunlop, C. Fedeli, F. Fontanot, P. Franzetti, C. Giocoli, J. Garcia-Bellido, J. Gow, A. Heavens, P. Hewett, C. Heymans, A. Holland, Z. Huang, O. Ilbert, B. Joachimi, E. Jennins, E. Kerins, A. Kiessling, D. Kirk, R. Kotak, O. Krause, O. Lahav, F. van Leeuwen, J. Lesgourgues, M. Lombardi, M. Magliocchetti, K. Maguire, E. Majerotto, R. Maoli, F. Marulli, S. Maurogordato, H. McCracken, R. McLure, A. Melchiorri, A. Merson, M. Moresco, M. Nonino, P. Norberg, J. Peacock, R. Pello, M. Penny, V. Pettorino, C. Di Porto, L. Pozzetti, C. Quercellini, M. Radovich, A. Rassat, N. Roche, S. Ronayette, E. Rossetti, B. Sartoris, P. Schneider, E. Semboloni, S. Serjeant, F. Simpson, C. Skordis, G. Smadja, S. Smartt, P. Spano, S. Spiro, M. Sullivan, A. Tilquin, R. Trotta, L. Verde, Y. Wang, G. Williger, G. Zhao, J. Zoubian, and E. Zucca. Euclid Definition Study Report. *arXiv e-prints*, page arXiv:1110.3193, October 2011.
- [26] Pablo Lemos, Anthony Challinor, and George Efstathiou. The effect of Limber and flat-sky approximations on galaxy weak lensing. , 2017(5):014, May 2017.
- [27] Antony Lewis, Anthony Challinor, and Duncan Hanson. The shape of the CMB lensing bispectrum. , 2011(3):018, March 2011.
- [28] Michele Liguori, Emiliano Sefusatti, James 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, January 2010.
- [29] Juan Maldacena. Non-gaussian features of primordial fluctuations in single field inflationary models. *Journal of High Energy Physics*, 2003(5):013, May 2003.
- [30] Azita Mayeli. Asymptotic Uncorrelation for Mexican Needlets. *arXiv e-prints*, page arXiv:0806.3009, June 2008.
- [31] D. Munshi, T. Namikawa, T. D. Kitching, J. D. McEwen, and F. R. Bouchet. Weak lensing skew-spectrum. , 498(4):6057–6068, November 2020.
- [32] D. Munshi, T. Namikawa, J. D. McEwen, T. D. Kitching, and F. R. Bouchet. Morphology of weak lensing convergence maps. , 507(1):1421–1433, October 2021.
- [33] Dipak Munshi and Alan Heavens. A new approach to probing primordial non-Gaussianity. , 401(4):2406–2418, February 2010.
- [34] N. Palanque-Delabrouille, Ch. Yeche, A. D. Myers, P. Petitjean, N. P. Ross, E. Sheldon, E. Aubourg, T. Delubac, J. M. Le Goff, I. Pâris, J. Rich, K. S. Dawson, D. P. Schneider, and B. A. Weaver. Variability selected high-redshift quasars on SDSS Stripe 82. , 530:A122, June 2011.

- [35] Ruth Pearson, Antony Lewis, and Donough Regan. CMB lensing and primordial squeezed non-gaussianity. , 2012(3):011, March 2012.
- [36] Cyril Pitrou, Thiago S. Pereira, and Jean-Philippe Uzan. Weak lensing by the large scale structure in a spatially anisotropic universe: Theory and predictions. , 92(2):023501, July 2015.
- [37] Planck Collaboration. Planck 2018 results. VI. Cosmological parameters. , 641:A6, September 2020.
- [38] Natalia Porqueres, Alan Heavens, Daniel Mortlock, and Guilhem Lavaux. Bayesian forward modelling of cosmic shear data. , 502(2):3035–3044, April 2021.
- [39] Geraint Pratten and Antony Lewis. Impact of post-Born lensing on the CMB. , 2016(8):047, August 2016.
- [40] Susan Pyne and Benjamin Joachimi. Self-calibration of weak lensing systematic effects using combined two- and three-point statistics. , 503(2):2300–2317, May 2021.
- [41] Alexandre Refregier. Weak Gravitational Lensing by Large-Scale Structure. , 41:645–668, January 2003.
- [42] Matteo Rizzato, Karim Benabed, Francis Bernardeau, and Fabien Lacasa. Tomographic weak lensing bispectrum: a thorough analysis towards the next generation of galaxy surveys. , 490(4):4688–4714, December 2019.
- [43] M. J. Rycroft. Cosmology–The Origin and Evolution of Cosmic Structure. *Journal of Atmospheric and Terrestrial Physics*, 58:1818–1819, November 1996.
- [44] Masanori Sato and Takahiro Nishimichi. Impact of the non-Gaussian covariance of the weak lensing power spectrum and bispectrum on cosmological parameter estimation. , 87(12):123538, June 2013.
- [45] Peter Schneider. Weak Gravitational Lensing. *arXiv e-prints*, pages astro-ph/0509252, September 2005.
- [46] S. Scodeller, Ø. Rudjord, F. K. Hansen, D. Marinucci, D. Geller, and A. Mayeli. Introducing Mexican Needlets for CMB Analysis: Issues for Practical Applications and Comparison with Standard Needlets. , 733(2):121, June 2011.
- [47] George F. Smoot. COBE observations and results. In Luciano Maiani, Francesco Melchiorri, and Nicola Vittorio, editors, *3K cosmology*, volume 476 of *American Institute of Physics Conference Series*, pages 1–10, May 1999.
- [48] G. Soucail, Y. Mellier, B. Fort, G. Mathez, and S. D’Odorico. CL 2244-02. , 4456:4, September 1987.
- [49] Masahiro Takada and Bhuvnesh Jain. Cosmological parameters from lensing power spectrum and bispectrum tomography. , 348(3):897–915, March 2004.

- [50] Ryuichi Takahashi, Takashi Hamana, Masato Shirasaki, Toshiya Namikawa, Takahiro Nishimichi, Ken Osato, and Kosei Shiroyama. Full-sky Gravitational Lensing Simulation for Large-area Galaxy Surveys and Cosmic Microwave Background Experiments. , 850(1):24, November 2017.
- [51] J. A. Tyson, F. Valdes, and R. A. Wenk. Detection of Systematic Gravitational Lens Galaxy Image Alignments: Mapping Dark Matter in Galaxy Clusters. , 349:L1, January 1990.
- [52] L. Van Waerbeke, Y. Mellier, T. Erben, J. C. Cuillandre, F. Bernardeau, R. Maoli, E. Bertin, H. J. McCracken, O. Le Fèvre, B. Fort, M. Dantel-Fort, B. Jain, and P. Schneider. Detection of correlated galaxy ellipticities from CFHT data: first evidence for gravitational lensing by large-scale structures. , 358:30–44, June 2000.
- [53] D. Walsh, R. F. Carswell, and R. J. Weymann. 0957+561 A, B: twin quasistellar objects or gravitational lens? , 279:381–384, May 1979.
- [54] David M. Wittman, J. Anthony Tyson, David Kirkman, Ian Dell’Antonio, and Gary Bernstein. Detection of weak gravitational lensing distortions of distant galaxies by cosmic dark matter at large scales. , 405(6783):143–148, May 2000.