Higher-order statistics of weak gravitational lensing

Susan Pyne

Thesis submitted for the degree of Doctor of Philosophy

Supervisors:
Dr Benjamin Joachimi
Prof. Hiranya V. Peiris

Examiners:
Prof. Antony Lewis
Dr Andreu Font-Ribera

Department of Physics and Astronomy
University College London

July 2020

I, *Susan Pyne*, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

This thesis is dedicated to my mother
who sadly did not live to see its completion,
but is always in my thoughts.

Yet what are all such gaieties to me

Whose thoughts are full of indices and surds?

$$x^2 + 7x + 53 = \frac{11}{3}$$

Lewis Carroll: Four Riddles

ABSTRACT

The next decade will see the advent of unprecedentedly large cosmological surveys, optimised to provide data about the gravitational lensing of galaxies. This opens up the possibility of exploring methods and statistics which are out of reach of current surveys. In this spirit, this thesis focuses on exploiting three-point and higher-order weak lensing statistics.

First we consider the deflection of three-point correlation functions by weak lensing, a small, subtle signal which is not accessible with current surveys. We derive a general expression for the lensing deflection but show that its detection must await even larger and deeper surveys.

We next consider the information content of the weak lensing bispectrum. We confirm that using the bispectrum as well the power spectrum can help to reduce statistical errors on cosmological parameters. Moreover we show that the bispectrum can help mitigate two major systematic uncertainties, the intrinsic alignment of galaxies and redshift errors. We find that these affect the power spectrum and bispectrum differently, and that using the bispectrum can facilitate self-calibration. This is a promising finding which could be extended to other systematics.

Future surveys will probe small, non-linear scales so in a Bayesian weak lensing analysis it may not be valid to approximate the likelihood as Gaussian. We discuss theoretical alternatives in Fourier space and show that the real space weak lensing likelihood is also theoretically non-Gaussian. In practice if a Gaussian likelihood is assumed then the covariance matrix should be calculated at a fixed point in parameter space. Working to the accuracy required by future surveys, it is important to choose this point optimally. We develop an emulator for the weak lensing power spectrum covariance and demonstrate an iterative method to determine this fixed cosmology in a principled way.

IMPACT STATEMENT

This thesis is part of a body of work currently being undertaken across the cosmology community to prepare for the next generation of galaxy surveys, including *Euclid*, the Dark Energy Spectroscopic Instrument and the Rubin Observatory Legacy Survey of Space and Time. The science goals of these surveys focus on some profound mysteries about the Universe, in particular the nature and properties of dark matter and dark energy. To this end the surveys will look deeper into the Universe, and hence farther back in time, than current surveys do, gathering an unprecedented wealth of highly accurate and detailed data about tens of billions of galaxies. This exquisitely accurate data will transform weak gravitational lensing, the subject of this thesis, into one of a handful of truly precision cosmological probes.

These next-generation surveys are financed by public investment of hundreds of millions of dollars from countries across the globe; their data is an extremely valuable resource which cosmologists have an obligation to exploit fully. This thesis focuses on ways to get more out of the new data. We may be able to detect signals which cannot be measured in current surveys. Equally it may become practicable to obtain new information from higher-order statistics which at present are not viable. These statistics do not require additional data, just more complex analysis of data which will be collected anyway for conventional analysis. Even more fruitfully, as this thesis shows, higher-order statistics provide a promising new way of handling systematic uncertainties in weak lensing, again without the need for extra data. This is important because controlling systematics will become more pressing as abundant data reduces statistical errors. Finally, new insights can be gained by taking into account fine details which were previously inaccessible or by relaxing simplifying assumptions.

This thesis considers examples of all these possibilities for extracting extra value from future weak lensing surveys and paves the way for three-point statistics to become a standard tool for analysing weak lensing data. Similar methods may also be feasible for other cosmological probes.

ACKNOWLEDGEMENTS

First I would like to thank my supervisor, Benjamin Joachimi, for taking on a rather 'non-standard' student in the first place, but also for his patient support and scientific guidance, for making time for regular meetings despite his busy schedule, for helping me to do things I didn't know I could, for the breadth and depth of his knowledge of weak lensing, and for reliably being right. I couldn't wish for a better supervisor. I also acknowledge use of his comprehensive cosmological code. In many parts of my work I used his code for two-point statistics as a basis for my three-point code.

Next I owe a debt of gratitude to my second supervisor, Hiranya Peiris, an outstanding role model and inspiration for any female cosmologist, who was ready give me time when I needed a listener, and who always provided wise insights and constructive advice. It has been a real privilege to learn from her.

I am particularly grateful to Fabian Köhlinger and Alessio Spurio Mancini for helpful advice about using their KiDS likelihood for the work in Chapter 4, and to Stephen Stopyra for helping me to get to grips with Mathematica.

My thanks also to John Deacon and especially Edd Edmondson for their expertise in sorting out my IT problems, even when they were my own fault, and to Anita Maguire for efficiently dealing with administrative matters and keeping us all happy. Thank you also to all the UCL colleagues who over the course of my PhD have given me advice, tips, bits of code, or just been friendly.

I must also thank my family. Even in her final illness my mother was gently amused by the idea of me doing a PhD. This thesis is dedicated to her. My daughter Ros has been another good listener and source of cheerful, intelligent support, reversing the normal mother-daughter relationship.

Last but certainly not least I thank my husband Geoff who encouraged me to embark on a PhD in the first place, and supported me wholeheartedly throughout. He has helped me enormously by acting as a sounding board for ideas and initiating many stimulating scientific discussions from the perspective of an experimentalist. I'm grateful also for his practical skills: he has often solved technical and IT problems, and gave me one of my best ever birthday presents - a PC which he designed for fast calculations with big datasets. I couldn't have done this without him.

THESIS ROADMAP

The first chapter of this thesis is a three-part overview of necessary background to the rest of the work. The first part outlines fundamental topics in modern cosmology, including some aspects of its historical development. The second introduces the main cosmological probes which are referred to in subsequent chapters, in particular weak gravitational lensing. The third part summarises methods and techniques which are used later in the thesis.

In Chapter 2 we discuss the weak lensing deflection of three-point correlation functions. This is a potential source of information which is not accessible through current surveys. This chapter is based on Pyne, S., Joachimi, B. and Peiris, H.V. (2017), 'Weak lensing deflection of three-point correlation functions', *Journal of Cosmology and Astroparticle Physics* **2017**(12), 043.

Chapter 3 investigates the information content of the weak lensing bispectrum. This is again in three parts. The first is a theoretical account of the composition of the weak lensing bispectrum covariance matrix. In the second part we investigate the information content of the weak lensing bispectrum using Fisher matrix analysis. Finally we explore whether the bispectrum can help to mitigate two major sources of systematic uncertainty in weak lensing, intrinsic alignments and redshift errors.

Chapter 4 covers several slightly eclectic topics related to likelihoods and covariance matrices used in Bayesian estimation of cosmological parameters. First we look theoretically at the Gaussianity of power spectrum likelihoods and review alternative non-Gaussian likelihoods. We also investigate a hierarchical model of weak lensing two-point correlation function likelihoods. Finally we discuss the parameter-dependence of the power spectrum covariance matrix and the cosmology at which the covariance matrix should be calculated.

The final chapter draws together conclusions from the whole thesis and discusses the potential impact of the work for next-generation surveys.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 Introduction

1.1 Cosmology

1.1.1 Historical perspective

Cosmology is the study of the evolution, structure and contents of the Universe on the largest scales. Humans have speculated about the origin and nature of the Universe since the dawn of civilisation, but it is only since the scientific revolution of the past few hundred years that we have been able to propose consistent theories about the Universe as a whole, culminating in Einstein's development of general relativity in the early twentieth century. And it is only in the past 30 years or so that we have had plentiful accurate data with which to test these theories. This has led to spectacular breakthroughs in our understanding of cosmology and the development of a generally agreed 'concordance'model – the Lambda Cold Dark Matter (ΛCDM) model. Remarkably this model contains only a handful of free parameters but can nevertheless explain most of the observed features of the Universe.

Despite these achievements there are many glaring gaps in our understanding. In particular we do not understand either dark matter – which appears to make up 85% of the matter in the Universe but interacts with 'normal' matter only through gravity – or so-called dark energy, which is causing the expansion of the Universe to accelerate. Solving these mysteries requires ever greater ingenuity in analysis and inference. This is the rationale for the next generation of cosmological probes which will yield unprecedented volumes of high-precision data and demand equally sophisticated analytical tools.

1.1.2 General relativity and the Friedmann equations

It is now over a hundred years since Einstein first formulated general relativity (Einstein 1915*b*) but this theory still underpins our understanding of the Universe. Over the years its predictions have been substantiated by numerous observations, from Einstein's own explanation of Mercury's apparently anomalous orbit (Einstein 1915*a*), through Eddington's measurements of the deflection of starlight during the 1919 solar eclipse (Eddington 1919), down to the first observation of extragalactic gravitational lensing in 1979 (Walsh et al. 1979) and the detection of gravitational

waves a century after Einstein's ground-breaking paper (Abbott et al. 2016). The key insight of general relativity is that the gravitational field is not caused by a classical Newtonian gravitational force but by the geometrical properties of spacetime. The theory relates the curvature of spacetime to the matter and energy content of the Universe via Einstein's field equations (Einstein 1915*c*):

$$G_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu} \, . \tag{1.1}$$

G is Newton's gravitational constant and c is the speed of light in a vacuum.

The left-hand side of Equation 1.1 describes the curvature of spacetime through the Einstein tensor $G_{\mu\nu}$. Associated with this is the metric which describes how coordinate distances are related to physical distances, or in other words what the geometry of the Universe is,

$$\mathrm{d}s^2 = g_{\mu\nu} \mathrm{d}x^\mu \mathrm{d}x^\nu \, , \tag{1.2}$$

where $g_{\mu\nu}$ is the metric tensor and we use the Einstein summation convention to sum over repeated indices.

The right-hand side of Equation 1.1 describes the constituents of the Universe through the energy–momentum tensor $T_{\mu\nu}$. Fundamentally the equation as a whole says that the gravitational field is sourced by the energy and matter content of the Universe, and in turn the geometry of spacetime determines the dynamics of matter and energy (Ford and Wheeler 1998).

Equation 1.1 is a very general set of 10 coupled non-linear second-order partial differential equations which could describe many universes. In the most general case it is difficult or impossible to find solutions and Einstein initially did not apply his results to the Universe as a whole. However, soon after the seminal 1915 paper, others realised that the equations could be solved in specific cases which assumed various symmetries. In 1916 Schwarzchild published a solution for the case of a spherically symmetric spacetime surrounding a massive object (Schwarzschild 1916). Then in 1922 Friedmann found a solution for a spatially symmetric system which evolved in time (Friedmann 1922), and in 1927 Lemaître also published a solution for an expanding universe (Lemaître 1927). Similar solutions were also found independently by Robertson (Robertson 1935) and by Walker (Walker 1937).

Einstein embraced these ideas for simplifying his equations and applying them to the observed Universe. In particular he readily accepted the spatial symmetries which Friedmann had explored because observational evidence suggested that on large scales the Universe is homogeneous (there is no preferred observing position) and isotropic (it looks the same in all directions). This is now

known as the Cosmological Principle. (Clearly this principle does not apply on smaller scales, for example of individual galaxies, and on those scales we need to introduce additional assumptions and complexity). However, like many other scientists of the time, Einstein believed (without strong evidence) that the Universe had to be static and eternal. He therefore viewed evolving solutions as of only theoretical interest, and to describe a static universe he included an arbitrary 'cosmological constant', Λ, in his equations, to give

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu} \, . \tag{1.3}$$

Einstein inserted this on the left-hand side although he did not give strong reasons for this choice, apart from floating the idea that it is effectively a constant of integration (O'Raifeartaigh et al. 2017). The term can equivalently be considered to represent vacuum energy, the energy density of empty space, and subtracted from the right-hand side.

Unknown to Einstein, by 1915 Slipher had measured the spectra of a sample of galaxies and observed that in most cases their spectral lines were redshifted so the galaxies were apparently moving away from us (Slipher 1915). This was confirmed by other observers, leading to gradual acceptance that the Universe is in fact expanding (O'Raifeartaigh 2013). Subsequently Lemaître noted that in an expanding universe recession velocities would be proportional to distance from the observer, and estimated the constant of proportionality (Lemaître 1927). A little later Hubble independently produced an improved estimate of this parameter using his own more accurate observations, especially of distances (Hubble 1929; Hubble and Humason 1931). The relationship between distance and recession velocity is now known as the Hubble–Lemaître Law.

Thus by the early 1930s there was increasing evidence that we live in an expanding universe. At this stage Einstein dropped the cosmological constant, which he is said to have described as his 'biggest blunder', although this may be apocryphal (O'Raifeartaigh, Cormac and Mitton, Simon 2018). Since then attention has focused on time-dependent solutions to the field equations.

To parameterise the time evolution of the spatial part of the metric we can introduce the scale factor $a(t)$, and we can also parameterise the (unknown) curvature of spacetime. This leads to the metric which describes a homogeneous, isotropic and evolving universe, the Friedmann–Lemaître–Robertson–Walker (FLRW) metric:

$$\mathrm{d}s^2 = c^2\mathrm{d}t^2 - a^2(t)\left[\frac{\mathrm{d}r^2}{1-Kr^2} + r^2\mathrm{d}\Omega^2\right] , \tag{1.4}$$

where t is coordinate time, r, θ, ϕ are spherical polar coordinates with $\mathrm{d}\Omega^2 = \mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}\phi^2$, and K quantifies the curvature ($K = -1, 0, +1$ according to whether the Universe is open, flat or closed). It is easy to see that this metric is in fact isotropic since the time and space parts are separated, and also homogeneous because it has translational symmetry: it is the same at every point.

It is often convenient to separate the spatial part of the metric into radial and angular parts, using the comoving coordinate χ defined by

$$\mathrm{d}\chi = \frac{\mathrm{d}r}{\sqrt{1 - Kr^2}} \,. \tag{1.5}$$

Integrating this gives

$$r = f_K(\chi) \,, \tag{1.6}$$

where $f_K(\chi)$ is the comoving angular distance which depends on the curvature:

$$f_K(\chi) = \begin{cases} (-K)^{-1/2} \sinh((-K)^{-1/2}\chi), & \text{if } K < 0 \\ \chi, & \text{if } K = 0 \\ K^{-1/2} \sin(K^{-1/2}\chi), & \text{if } K > 0 \,. \end{cases} \tag{1.7}$$

The FLRW metric can then be written as

$$\mathrm{d}s^2 = c^2\mathrm{d}t^2 - a^2(t)[\mathrm{d}\chi^2 + f_K^2(\chi) d\Omega^2] \,. \tag{1.8}$$

Returning to Equation 1.1, the right-hand side can be interpreted by assuming that the individual components of the Universe act like perfect fluids. This is a reasonable assumption given what we know about the constituents of the Universe. With this interpretation, T^0_0 quantifies the density, ρ, T^0_i is the momentum density and T^i_j is the pressure P for $i = j, (i, j = 1, 2, 3)$ and zero if $i \neq j$.

Thus the energy–momentum tensor takes the form

$$T^{\mu}_{\ \nu} = \begin{pmatrix} \rho c^2 & 0 & 0 & 0 \\ 0 & -P & 0 & 0 \\ 0 & 0 & -P & 0 \\ 0 & 0 & 0 & -P \end{pmatrix} . \tag{1.9}$$

If we now substitute Equation 1.9 into Equation 1.1, together with the FLRW metric, we obtain the Friedmann equations. The time-time component produces the first Friedmann equation:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{Kc^2}{a^2} , \tag{1.10}$$

where the overdot denotes differentiation with respect to time. The space-space component of Equation 1.1 gives us the second Friedmann equation :

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3P}{c^2}\right) . \tag{1.11}$$

The quantity $\dot{a}/a$ is the Hubble parameter $H(t)$. This is a crucial parameter for modern cosmology. Its value is known very much more accurately than in Hubble's day, but this greater accuracy has led to tension between measurements of its present-day value H_0 based on local (recent) data and predictions based on distant (past) data. Figure 1.1, taken from Verde et al. (2019), shows recent estimates of H_0 based on multiple, mainly independent, methods. The approximately 5σ tension suggests that there may be a genuine discrepancy between the early and late values which is not due to unknown systematic errors. Notably, inferences from *Planck* data rely on the assumed cosmological model whereas the late-Universe estimates are direct measurements. This could point to the need to revise the model.

The existence of uncertainty in the measurement of H_0 is often expressed through the parameter h with $H_0 = 100h \text{ km s}^{-1}\text{Mpc}^{-1}$.

The second Friedmann equation tells us about the rate of change of the Universe's expansion. Until the late 1990s it was assumed that the expansion would decelerate under the influence of gravity. However two ground-breaking experiments by Riess et al. (1998) and Perlmutter et al. (1999) measured the distances to Type1a supernovae and provided persuasive, but unexpected, evidence that the Universe is not only expanding, but accelerating. For this reason cosmologists

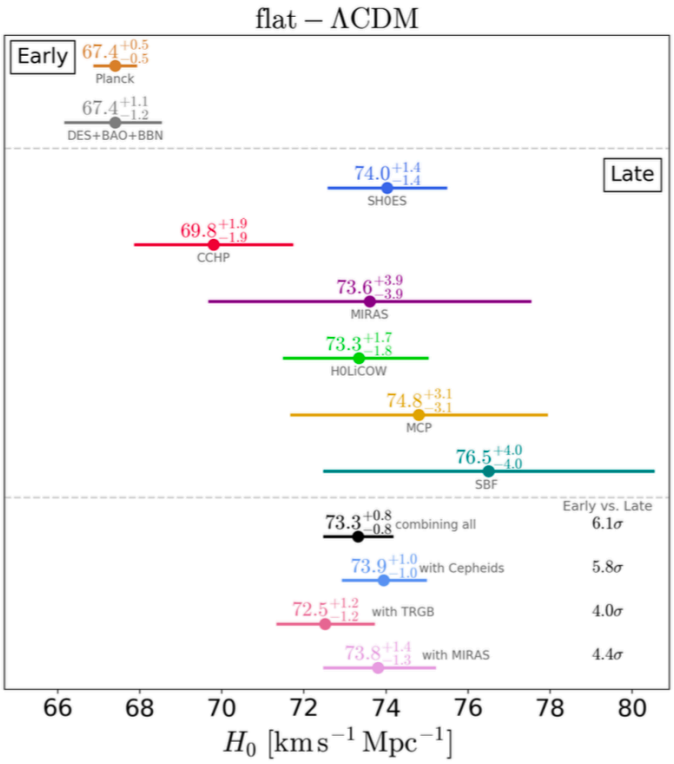

Figure 1.1: Recent estimates of H_0, the present-day value of the Hubble parameter. From top to bottom: two independent early-Universe predictions, multiple late-Universe measurements using a variety of methods, and selected combinations of the late-Universe estimates. From Verde et al. (2019) which provides full details of sources and methods.

re-introduced the cosmological constant, but now by convention added it to the right-hand side of the Einstein equation, consistent with expansion caused by the vacuum energy density. In practice the cause of the expansion is unknown, and we reflect this ignorance by the term 'dark energy'. If dark energy is indeed equivalent to the vacuum energy, it can be described by a cosmological constant and has equation of state

$$w \equiv \frac{P}{\rho c^2} \qquad (1.12)$$
$$= -1 \, .$$

However it is also possible that dark energy evolves with time and it is common to parameterise the dark energy equation of state in terms of a linear model:

$$w = w_0 + (1 - a)w_a \ , \tag{1.13}$$

where w_0 is the value of w at $z = 0$ and a is the scale factor. Determining the dark energy equation of state is a major goal of current cosmological experiments.

With the introduction of a cosmological constant the Friedmann equations become:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{\kappa c^2}{a^2} + \frac{\Lambda c^2}{3} \ , \tag{1.14}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3P}{c^2}\right) + \frac{\Lambda c^2}{3}. \tag{1.15}$$

1.1.3 Cosmological redshift

From Slipher and Hubble onwards it has been observed that features in the spectra of more distant objects are shifted to longer wavelengths compared to their restframe positions. These shifts arise because the wavelength of light is stretched by the expansion of the Universe. So instead of using the scale factor to parameterise the expansion of the Universe it is common to use the redshift z, defined as

$$z = \frac{\lambda_{\rm obs} - \lambda_{\rm em}}{\lambda_{\rm em}} \ , \tag{1.16}$$

which is more directly related to observable quantities. Now consider a photon emitted from the source at time $t_{\rm em}$ and observed at time $t_{\rm obs}$ and construct the coordinate system so the observer is at the origin. The photon travels along a null geodesic so that $\mathrm{d}s^2 = 0$ and Equation 1.7 gives us

$$c^2 dt^2 - a^2(t)\mathrm{d}\chi^2 = 0 \ . \tag{1.17}$$

Thus the comoving distance χ between the source and observer is

$$\chi = \int_{t_{\rm em}}^{t_{\rm obs}} \frac{c\,\mathrm{d}t}{a(t)} \ . \tag{1.18}$$

By definition this is constant, so

$$\frac{\mathrm{d}t_{\mathrm{obs}}}{\mathrm{d}t_{\mathrm{em}}} = \frac{a_{\mathrm{obs}}}{a_{\mathrm{em}}} \,. \tag{1.19}$$

The time intervals are inversely proportional to the frequency of the light and hence proportional to the wavelength. So we have

$$\frac{\lambda_{\mathrm{obs}}}{\lambda_{\mathrm{em}}} = \frac{a_{\mathrm{obs}}}{a_{\mathrm{em}}} \,. \tag{1.20}$$

If we take the scale factor today to be equal to 1 this leads to the relationship between scale factor and redshift

$$a_{\mathrm{em}} = \frac{1}{1 + z_{\mathrm{em}}} \,. \tag{1.21}$$

1.1.4 Components of the Universe

The components of the Universe can be categorised as:

- relativistic species, mainly radiation but also neutrinos;
- non-relativistic matter;
- dark energy;
- spatial curvature component.

To understand how the total density of all components changes as the scale factor changes we can use the energy-momentum tensor T^{μ}_{ν} (Equation 1.9). Energy conservation requires that the covariant time derivative of T^{μ}_{ν} is zero. Thus we obtain the continuity equation

$$\frac{\partial \rho}{\partial t} + 3\frac{\dot{a}}{a}\left(\rho + \frac{P}{c^2}\right) = 0 \,. \tag{1.22}$$

This equation can be decomposed into separate equations for each component. For pressureless matter $P = 0$ by definition and we get $\rho_{\mathrm{m}} \propto a^{-3}$. Radiation has $P = \rho c^2/3$ which leads to $\rho_{\mathrm{r}} \propto a^{-4}$. Finally, the density of dark energy does not evolve with time and so is independent of the scale factor.

We can thus see that the composition of the total energy density has changed significantly over the history of the Universe as in Figure 1.2. At early times the Universe was radiation-dominated

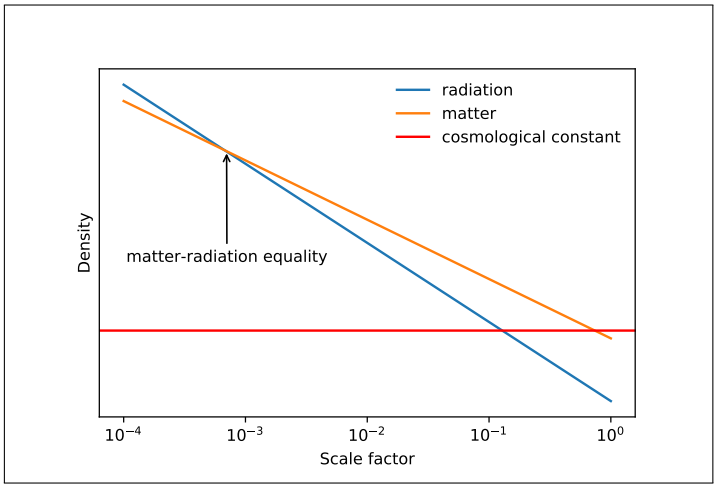

Figure 1.2: Schematic diagram of the evolution of radiation, matter and dark energy densities.

Epoch	Density $\rho(a)$	Scale factor $a(t)$	Hubble parameter $H(t) \equiv \dot{a}/a$
Radiation domination	$\propto a^{-4}$	$\propto t^{1/2}$	$= \frac{1}{2t}$
Matter domination	$\propto a^{-3}$	$\propto t^{2/3}$	$= \frac{2}{3t}$
Dark energy domination	constant	$\propto \exp(H_0 t)$	$H_0 = \sqrt{\frac{\Lambda}{3}}$

Table 1.1: The evolution of the density, scale factor and Hubble parameter at different epochs.

but because the energy density of radiation fell away faster than that of matter, the matter and radiation densities subsequently equalised. As the Universe expanded even more, the energy density of non-relativistic matter fell further. Eventually it will be less than the density of dark energy. Intriguingly, at the present epoch the energy densities of matter and of dark energy are similar. This is known as the coincidence problem – why should we happen to exist at a time when the Universe is changing from matter domination to dark energy domination?

It is also useful to see how the scale factor evolves in a spatially flat Universe. For $w \neq -1$ the scale factor satisfies

$$a(t) = a_0 t^{2/(3(w+1))} \,, \tag{1.23}$$

where w is given by the relevant equation of state (Equation 1.13) which relates pressure and density. For radiation $w = 1/3$; for pressureless matter $w = 0$; and for vacuum energy $w = -1$. Thus during the radiation-dominated era $a(t) \propto t^{1/2}$, when matter dominates $a(t) \propto t^{2/3}$, and when dark energy dominates $a(t)$ grows exponentially.

Table 1.1 summarises the evolution of the Universe at each epoch.

The total density of the Universe can alternatively be quantified as

$$\Omega = \frac{8\pi G}{3H^2}\rho \ = \frac{\rho}{\rho_{\rm c}} \ . \tag{1.24}$$

The quantity $\rho_{\rm c} = 3H^2/8\pi G$ is called the critical density. This density was originally defined to help investigate the curvature of a universe with no dark energy. If the total density is smaller than $\rho_{\rm c}$ such a universe is open, whereas if the density is greater than $\rho_{\rm c}$ the universe will be closed. If the total density exactly equals the critical density then the universe is flat. Observational evidence from the cosmic microwave background (CMB) (Akrami et al. 2018) suggests that the Universe today is very nearly flat (see Section 1.2.1). This in turn implies that at early times the total density must have been even closer to the critical density. This is known as the 'flatness' problem. We do not understand how or why the initial density could be so finely tuned to the critical value.

Density parameters for individual components satisfy:

$$\Omega_{\rm m} = \frac{8\pi G}{3H^2}\rho_{\rm m} \ , \tag{1.25}$$

$$\Omega_{\rm r} = \frac{8\pi G}{3H^2}\rho_{\rm r} \ , \tag{1.26}$$

$$\Omega_{\Lambda} = \frac{\Lambda c^2}{3H^2} \ , \tag{1.27}$$

$$\Omega_{\rm K} = \frac{-Kc^2}{H^2} \ . \tag{1.28}$$

Determining the current values of these parameters is another important concern of modern cosmology.

The first Friedmann equation, Equation 1.10, can be re-written in terms of the density parameters as

$$\frac{H^2}{H_0^2} = \Omega_{\rm r,0}(1+z)^4 + \Omega_{\rm m,0}(1+z)^3 + \Omega_{\rm K,0}(1+z)^2 + \Omega_{\Lambda} \ , \tag{1.29}$$

where the subscript 0 indicates the value of a parameter at the present time.

1.1.5 Evolution of the Universe

The expansion of the Universe implies that it was originally smaller and also hotter. So the evolution of the Universe is broadly a story of expansion and cooling. Important events which contributed to the growth of large-scale structure are:

- There was an early period of exponential expansion known as inflation.
- As the Universe cooled, protons and neutrons formed, together with electrons and neutrinos. Following this nucleosynthesis occurred, creating heavier nuclei – mainly helium but also some lithium – from protons and neutrons. This process ended when the Universe became too cool for fusion to take place. The Universe then existed as an opaque plasma of photons, nuclei and electrons, plus neutrinos.
- The energy density of radiation fell faster than the energy density of matter as the universe cooled, and at $z \approx 3400$ the two densities became equal. After this the Universe was matter-dominated, which allowed the eventual development of large-scale structure.
- The Universe cooled further so that atoms were able to form. This is known as recombination (even though no previous combination had taken place).
- Hydrogen atoms generally formed in excited states but quickly fell to the ground state, emitting photons. This is known as decoupling and occurred at a redshift $z \approx 1100$. The decoupled photons were able to stream away from the neutral hydrogen.
- Once photons no longer scattered off electrons they could travel freely through the Universe. The expansion of the Universe has caused the wavelength of photons from the last scattering surface to increase so we now observe them at microwave frequencies: the cosmic microwave background.
- Throughout this time dark matter, which feels only gravitational effects, had gathered into halos and filaments. After recombination baryonic matter fell into the dark matter potential wells and eventually the first stars and galaxies began to form.
- As the Universe expanded and cooled further galaxies, clusters and superclusters formed, leading to the local Universe we see today.
- The dark energy density is becoming increasingly dominant, so expansion will continue exponentially and structures which are not gravitationally bound will be increasingly isolated from each other.

1.1.6 Structure formation and perturbation theory

Study of the CMB tells us that after recombination the Universe was very homogeneous and isotropic, and the dark matter density field had the statistical properties of a Gaussian random

field. On scales of several hundred megaparsecs the Universe is still statistically homogeneous and isotropic. However on smaller scales galaxy surveys reveal a rich nonlinear structure of galaxies, clusters and superclusters, filaments and voids. To understand how these nonlinear structures arose we need to explore how gravitational instability acts in an expanding universe. There are many excellent expositions of the development of nonlinear structure and its description by perturbation theory; this section draws in particular on Mukhanov et al. (1992), Jain and Bertschinger (1993) and Bernardeau et al. (2002).

After recombination the Universe was dominated by dark matter which interacted gravitationally with baryons. The scales of cosmological interest are much larger than the Jeans length so that gravitational collapse dominates over pressure. Thus in a simplified model the combined dark matter-baryon fluid can be treated as a non-relativistic, pressureless ideal fluid with zero vorticity. Moreover we are interested only in scales much smaller than the Hubble distance ct where t is the age of the Universe. In these circumstances Newtonian theory is an adequate description of gravity.

In a static universe the equations of motion for the fluid can be expressed in terms of physical coordinates $\mathbf{r}$ as

$$\frac{\partial \rho}{\partial t} + \nabla_{\mathbf{r}} \cdot (\rho \mathbf{v}) = 0 \,, \tag{1.30}$$

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla_{\mathbf{r}})\mathbf{v} + \nabla_{\mathbf{r}} \Phi = -\frac{\nabla_{\mathbf{r}} P}{\rho c^2}, \tag{1.31}$$

$$\nabla^2_{\mathbf{r}} \Phi = 4\pi G \rho \,, \tag{1.32}$$

where ρ is the density, $\mathbf{v} = \mathrm{d}\mathbf{r}/\mathrm{d}t$ is the velocity of the fluid, P is the pressure, and Φ is the gravitational potential. These three equations are respectively the continuity equation which expresses conservation of mass, the Euler equation which encapsulates momentum conservation, and the Poisson equation which describes the relationship between the gravitational potential and the matter density. We can assume that on large scales $P \ll \rho c^2$ for the dominant dark matter, so the right-hand side of Equation 1.31 can be set to zero.

If we now consider an expanding universe, it is helpful to change to comoving coordinates $\mathbf{x} = \mathbf{r}/a(t)$. The velocity of the fluid is now the sum of its peculiar velocity $\mathbf{u}$ and the velocity due to the Hubble flow

$$\mathbf{v}(\mathbf{x}, t) = a(t) H(t) \mathbf{x} + \mathbf{u}(\mathbf{x}, t) \,. \tag{1.33}$$

The spatial gradients with respect to $\mathbf{r}$ and $\mathbf{x}$ are related by

$$\nabla_{\mathbf{r}} = \frac{1}{a}\nabla_{\mathbf{x}} \, . \tag{1.34}$$

We also change to conformal time $\tau = t/a$. Finally, we are interested in the fate of small fluctuations in the density field so we define the density contrast $\delta(\mathbf{x}, \tau)$ by $\rho(\mathbf{x}, \tau) = \bar{\rho}(\tau)(1 + \delta(\mathbf{x}, \tau))$ where $\bar{\rho}$ is the mean (background) density.

With these changes, and keeping only first order terms in δ, Equations 1.30 to 1.32 become

$$\frac{\partial \delta(\mathbf{x}, \tau)}{\partial \tau} + \nabla_{\mathbf{x}} \cdot \mathbf{u}(\mathbf{x}, \tau) = 0 \, , \tag{1.35}$$

$$\frac{\partial \mathbf{u}(\mathbf{x}, \tau)}{\partial \tau} + \mathcal{H}(\tau)\mathbf{u}(\mathbf{x}, \tau) + \nabla_{\mathbf{x}}\Phi(\mathbf{x}, \tau) = 0 \, , \tag{1.36}$$

$$\nabla^2_{\mathbf{x}}\Phi(\mathbf{x}, \tau) \ = 4\pi a^2(\tau) G\bar{\rho}\delta(\mathbf{x}, \tau) \, , \tag{1.37}$$

where $\mathcal{H} = aH$ is the conformal expansion rate.

We now change to work in Fourier space. While fluctuations remain small different Fourier modes evolve independently but eventually perturbations grow sufficiently for coupling to occur between modes. Following Bernardeau et al. (2002) we take the divergence of Equation 1.36, arriving at

$$\frac{\partial \tilde{\delta}(\mathbf{k}, \tau)}{\partial \tau} + \tilde{\theta}(\mathbf{k}, \tau) = -\int \mathrm{d}^3\mathbf{k}_1 \mathrm{d}^3\mathbf{k}_2 \delta_D(\mathbf{k} - \mathbf{k}_{12}) \alpha(\mathbf{k}_1, \mathbf{k}_2) \tilde{\theta}(\mathbf{k}_1, \tau) \tilde{\delta}(\mathbf{k}_2, \tau) \, , \tag{1.38}$$

$$\frac{\partial \tilde{\theta}(\mathbf{k}, \tau)}{\partial \tau} + H(\tau)\tilde{\theta}(\mathbf{k}, \tau) + \frac{3}{2}H^2(\tau)\tilde{\delta}(\mathbf{k}, \tau) = -\int \mathrm{d}^3\mathbf{k}_1 \mathrm{d}^3\mathbf{k}_2 \delta_D(\mathbf{k} - \mathbf{k}_{12}) \times \beta(\mathbf{k}_1, \mathbf{k}_2) \tilde{\theta}(\mathbf{k}_1, \tau) \tilde{\theta}(\mathbf{k}_2, \tau) \, , \tag{1.39}$$

where $\theta \equiv \nabla_{\mathbf{x}} \cdot \mathbf{u}$, $\mathbf{k}_{12} = \mathbf{k}_1 + \mathbf{k}_2$ and δ_D is the Dirac delta function. At large scales we can linearise these equations, retaining only terms of order δ. However this is not valid at small scales, and here the functions

$$\alpha(\mathbf{k}_1, \mathbf{k}_2) = \frac{\mathbf{k}_{12} \cdot \mathbf{k}_1}{k_1^2} \, , \tag{1.40}$$

$$\beta(\mathbf{k}_1, \mathbf{k}_2) = \frac{k_{12}^2(\mathbf{k}_1 \cdot \mathbf{k}_2)}{2k_1^2 k_2^2} \tag{1.41}$$

encode the coupling between separate Fourier modes. Thus the nonlinear evolution of $\tilde{\delta}(\mathbf{k},\tau)$ and $\tilde{\theta}(\mathbf{k},\tau)$ depends on all pairs of wavevectors $\mathbf{k}_1$ and $\mathbf{k}_2$ whose sum is $\mathbf{k}$, consistent with translational invariance in a spatially homogeneous Universe (Goroff et al. 1986; Fry 1984).

For an Einstein-de Sitter Universe in which $\Omega_{\rm m} = 1$ and $\Omega_\Lambda = 0$ it is possible to find perturbative solutions to Equations 1.38 and 1.39 which separate the scale- and time-dependencies (Bernardeau et al. 2002):

$$\tilde{\delta}(\mathbf{k},\tau) = \sum_{n=1}^{\infty} a^n(\tau)\delta_n(\mathbf{k}) \; , \tag{1.42}$$

$$\tilde{\theta}(\mathbf{k},\tau) = -H(\tau)\sum_{n=1}^{\infty} a^n(\tau)\theta_n(\mathbf{k}) \; . \tag{1.43}$$

These perturbative expansions are not valid for a general universe in which $\Omega_{\rm m} \neq 1$ or $\Omega_\Lambda \neq 0$. However Scoccimarro et al. (1998) showed that, even in the general case, with reasonable assumptions it is possible to find a separable solution to any order which is valid at the percent level.

When $a(\tau)$ is small the first terms of Equations 1.42 and 1.43 are dominant. Also from the continuity equation $\theta_1(\mathbf{k}) = \delta_1(\mathbf{k})$ so in the linear regime the growth of fluctuations depends only on $\delta_1(\mathbf{k})$.

Inserting Equations 1.42 and 1.43 into Equations 1.38 and 1.39 gives:

$$\delta_n(\mathbf{k}) = \int \mathrm{d}^3\mathbf{q}_1 \ldots \int \mathrm{d}^3\mathbf{q}_n \delta_D(\mathbf{k}-\mathbf{q}_{1\ldots n}) F_n(\mathbf{q}_1,\ldots,\mathbf{q}_n)\delta_1(\mathbf{q}_1)\ldots\delta_1(\mathbf{q}_n) \; , \tag{1.44}$$

$$\theta_n(\mathbf{k}) = \int \mathrm{d}^3\mathbf{q}_1 \ldots \int \mathrm{d}^3\mathbf{q}_n \delta_D(\mathbf{k}-\mathbf{q}_{1\ldots n}) G_n(\mathbf{q}_1,\ldots,\mathbf{q}_n)\delta_1(\mathbf{q}_1)\ldots\delta_1(\mathbf{q}_n) \; , \tag{1.45}$$

where F_n and G_n are constructed from the mode-coupling functions $\alpha(\mathbf{k}_1,\mathbf{k}_2)$ and $\beta(\mathbf{k}_1,\mathbf{k}_2)$. For $n \geq 2$ these functions obey the recursion relations (Goroff et al. 1986; Jain and Bertschinger 1993):

$$F_n(\mathbf{q}_1 \dots \mathbf{q}_n) = \sum_{m=1}^{n-1} \frac{G_m(\mathbf{q}_1 \dots \mathbf{q}_n)}{(2n+3)(n-1)} \Big[(2n+1)\alpha(\mathbf{k}_1, \mathbf{k}_2) F_{n-m}(\mathbf{q}_{m+1}, \dots, \mathbf{q}_n) + 2\beta(\mathbf{k}_1, \mathbf{k}_2) G_{n-m}(\mathbf{q}_{m+1}, \dots, \mathbf{q}_n) \Big] , \tag{1.46}$$

$$G_n(\mathbf{q}_1 \dots \mathbf{q}_n) = \sum_{m=1}^{n-1} \frac{G_m(\mathbf{q}_1, \dots, \mathbf{q}_n)}{(2n+3)(n-1)} \Big[3\alpha(\mathbf{k}_1, \mathbf{k}_2) F_{n-m}(\mathbf{q}_{m+1}, \dots, \mathbf{q}_n) + 2n\beta(\mathbf{k}_1, \mathbf{k}_2) G_{n-m}(\mathbf{q}_{m+1}, \dots, \mathbf{q}_n) \Big] , \tag{1.47}$$

where $\mathbf{k}_1 \equiv \mathbf{q}_1 + \ldots + \mathbf{q}_m$ and $\mathbf{k}_2 \equiv \mathbf{q}_{m+1} + \ldots + \mathbf{q}_n$.

To explore nonlinear structure formation we need at least $n = 2$. It is often convenient to use versions of the kernels F_n and G_n which are symmetric in $\mathbf{k}_1$ and $\mathbf{k}_2$. The symmetrised kernels are obtained by summing over all possible permutations of the set $\pi = \{1 \dots n\}$ to give (Pielorz et al. 2010)

$$F_n^{(\mathrm{s})}(\mathbf{q}_1, \dots, \mathbf{q}_n) = \frac{1}{n!} \sum_{\pi} F_n(\mathbf{q}_{\pi(1)}, \dots, \mathbf{q}_{\pi(n)}) , \tag{1.48}$$

$$G_n^{(\mathrm{s})}(\mathbf{q}_1, \dots, \mathbf{q}_n) = \frac{1}{n!} \sum_{\pi} G_n(\mathbf{q}_{\pi(1)}, \dots, \mathbf{q}_{\pi(n)}) , \tag{1.49}$$

The symmetrised kernels for $n = 2$ are:

$$F_2^{(\mathrm{s})}(\mathbf{q}_1, \mathbf{q}_2) = \frac{5}{7} + \frac{1}{2} \frac{\mathbf{q}_1 \cdot \mathbf{q}_2 (q_1^2 + q_2^2)}{q_1^2 q_2^2} + \frac{2}{7} \frac{(\mathbf{q}_1 \cdot \mathbf{q}_2)^2}{q_1^2 q_2^2} , \tag{1.50}$$

$$G_2^{(\mathrm{s})}(\mathbf{q}_1, \mathbf{q}_2) = \frac{3}{7} + \frac{1}{2} \frac{\mathbf{q}_1 \cdot \mathbf{q}_2 (q_1^2 + q_2^2)}{q_1^2 q_2^2} + \frac{4}{7} \frac{(\mathbf{q}_1 \cdot \mathbf{q}_2)^2}{q_1^2 q_2^2} , \tag{1.51}$$

and for $n = 3$:

$$\begin{aligned} F_3^{(\mathrm{s})}&(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3) \\ &= \frac{7}{54} [\alpha(\mathbf{q}_1, \mathbf{q}_{23}) F_2^{(\mathrm{s})}(\mathbf{q}_2, \mathbf{q}_3) + \alpha(\mathbf{q}_2, \mathbf{q}_{13}) F_2^{(\mathrm{s})}(\mathbf{q}_1, \mathbf{q}_3) + \alpha(\mathbf{q}_3, \mathbf{q}_{12}) F_2^{(\mathrm{s})}(\mathbf{q}_1, \mathbf{q}_2)] \\ &\quad + \frac{4}{54} [\beta(\mathbf{q}_1, \mathbf{q}_{23}) G_2^{(\mathrm{s})}(\mathbf{q}_2, \mathbf{q}_3) + \beta(\mathbf{q}_2, \mathbf{q}_{13}) G_2^{(\mathrm{s})}(\mathbf{q}_1, \mathbf{q}_3) + \beta(\mathbf{q}_3, \mathbf{q}_{12}) G_2^{(\mathrm{s})}(\mathbf{q}_1, \mathbf{q}_2)] \\ &\quad + \frac{7}{54} [\alpha(\mathbf{q}_{12}, \mathbf{q}_3) G_2^{(\mathrm{s})}(\mathbf{q}_1, \mathbf{q}_2) + \alpha(\mathbf{q}_{13}, \mathbf{q}_2) G_2^{(\mathrm{s})}(\mathbf{q}_1, \mathbf{q}_3) + \alpha(\mathbf{q}_{23}, \mathbf{q}_1) G_2^{(\mathrm{s})}(\mathbf{q}_2, \mathbf{q}_3)] . \end{aligned} \tag{1.52}$$

1.1.7 The matter power spectrum

The matter power spectrum is a key theoretical and observational quantity which describes the distribution of matter as a function of scale. Figure 1.3 shows the inferred linear matter power spectrum at $z = 0$ based on several independent probes together with the predictions of the ΛCDM model. The observation-based estimates are very consistent especially at intermediate scales and the fit between model and observations is extremely good.

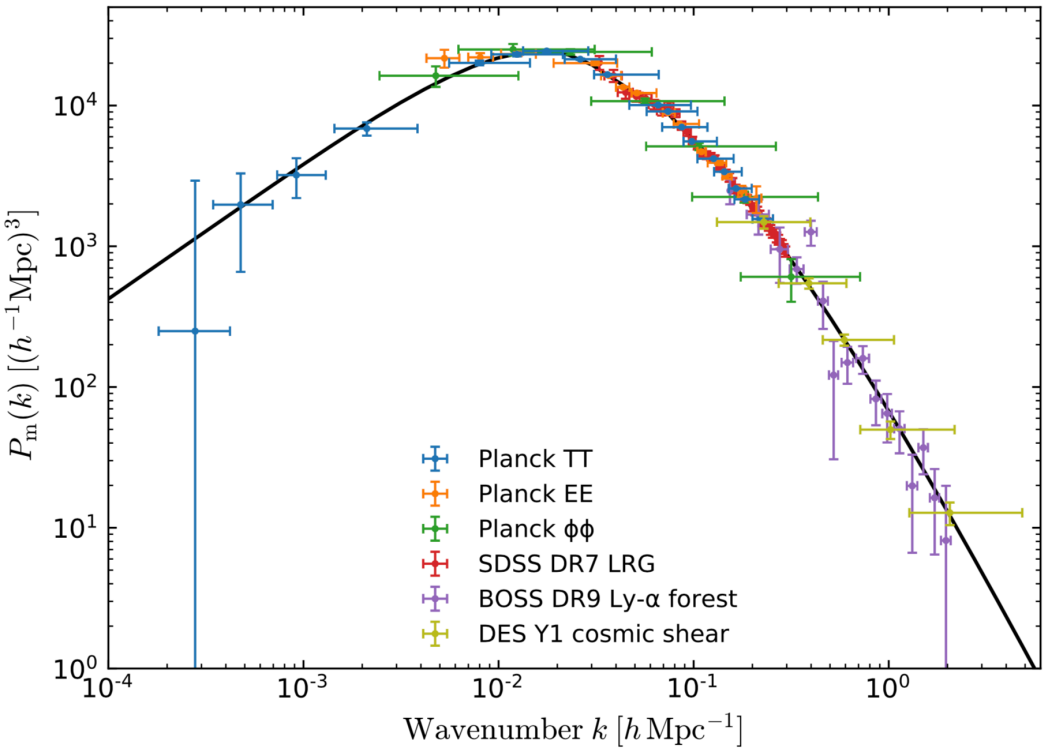

Figure 1.3: The linear matter power spectrum at $z = 0$ inferred from different cosmological probes. The black line is the prediction from the ΛCDM model. From Figure 19 of Akrami et al. (2018).

The broad shape of the linear power spectrum shown in Figure 1.3 can be explained by considering the evolution of density fluctuations in the early Universe. Inflationary theory predicts that the primordial power spectrum is given by a power law $P(k) \propto k^{n_{\rm s}}$. (Observation tells us that $n_{\rm s}$, the scalar spectral index, is very close to one.) Subsequent evolution depends on the fate of individual perturbations which grow independently in the linear regime.

At all stages fluctuations on scales greater than the comoving Hubble horizon, $c/(a(t)H(t))$, grow by gravitational collapse. However, as the Universe expands and this horizon grows, fluctuations may enter the horizon. This affects their growth. During the radiation-dominated era, radiation pressure dominates over gravitational effects. Fluctuations which enter the horizon

during this epoch undergo acoustic oscillations and their growth is inhibited. Once matter-radiation equality is reached, overdensities on all scales grow by gravitational collapse. Thus the power spectrum turns over at the wavenumber k_h equivalent to the horizon size at matter-radiation equality. Very approximately $P(k) \propto k$ if $k \leq k_h$ and $P(k) \propto k^{-3}$ if $k > k_h$. In practice the power spectrum has a smooth shape rather than a sharp peak because the transition from radiation- to matter-dominance is gradual.

The discussion above is a simplification. There are also other processes which affect the growth of perturbations, for example dissipative effects. Also baryonic perturbations, which are coupled to photons, evolve differently to dark matter. It is normal to consolidate all the transitions which convert the primordial power spectrum to the late-time power spectrum into a single function, the transfer function $T(k)$. Then the evolved power spectrum can be expressed as

$$P(k) \propto T^2(k) D^2(a) k^{n_\mathrm{s}} \,, \tag{1.53}$$

where $D(a)$ is the growth function which describes the growth of structure as a function of the scale factor. Various fitting functions for the transfer function have been developed, for example Bardeen et al. (1985) and Eisenstein and Hu (1998).

The amplitude of the matter power spectrum is not predicted by theory and must be measured observationally. It is generally expressed in terms of σ_8, the root mean square amplitude of fluctuations on a length scale $8\ h^{-1}$Mpc, evolved to the present day.

In the linear regime fluctuations remain small, the density field is Gaussian, and the growth of structure can be described by perturbation theory. However as perturbations continue to grow, eventually in some regions the overdensity becomes greater than the mean density of the Universe and nonlinearities arise. Individual Fourier modes become coupled and linear theory no longer applies. This happens first at small scales where gravitational potentials are large. The end result is that the nonlinear power spectrum is enhanced at high wavenumbers compared with the linear case. The density field on these scales ceases to be Gaussian and cannot be described by the power spectrum alone: higher-order statistics are also needed. The nonlinear power spectrum cannot be calculated analytically and must be estimated using simulations. However several fitting functions have been developed, including early work by Peacock and Dodds (1996), the Halofit method described in Smith et al. (2003) and its later development by Takahashi et al. (2012). Another possibility is to use emulation to interpolate the power spectrum between simulated points, for example Heitmann et al. (2010).

1.2 Cosmological probes

Many different probes are used to investigate cosmology but this section discusses only those which are relevant to the work in Chapters 2 to 4. The section begins with an overview of the cosmic microwave background since this provides much of the evidence for the ΛCDM model and the tightest constraints on cosmological parameters. Next comes an account of the evolution of baryon acoustic oscillations which are relevant to Chapter 2. Finally the bulk of this section discusses the theory behind weak gravitational lensing which is relevant to all the later chapters.

1.2.1 Cosmic microwave background

As explained in Section 1.1.5 our best understanding of the history of the Universe predicts that at a redshift $z \approx 1100$ photons were for the first time able to free-stream across the Universe. Furthermore the expansion of the Universe has redshifted these 'last-scattering' photons into microwave frequencies, so the radiation is known as the cosmic microwave background or CMB. Good general reviews of what we know about the CMB are Samtleben et al. (2007) and Staggs et al. (2018).

The existence of this microwave radiation was first predicted by early cosmological models designed to explain Big Bang nucleosynthesis, for example Alpher and Herman (1948); Gamow (1948). At that stage no deliberate attempts were made to detect the CMB, although in fact several measurements of the 'background' temperature were made in the 1940s and 1950s including by McKellar (1941) who estimated a temperature of 2.3K, remarkably close to the current best estimate of 2.73K. However the cosmological implications of these measurements were not understood or followed up. The CMB was eventually rediscovered by Penzias and Wilson (1965) and correctly identified by Dicke et al. (1965).

It was immediately realised that the CMB was an unrivalled source of information about the evolution of the Universe, because it carries evidence both of the history of the photons up to the time of last scattering and of their subsequent travel through the inhomogeneous Universe. The theoretical understanding of the statistical properties of CMB temperature and polarisation anistropies was worked out in the 1970s and 1980s (Sachs and Wolfe 1967; Rees and Sciama 1968; Silk 1968; Peebles and Yu 1970; Sunyaev and Zel'dovich 1972; Bond and Efstathiou 1987). This body of work predicted the form of the CMB temperature power spectrum and elucidated the information which could be extracted from CMB data once suitable probes were available.

There have been three major space-based missions to study the CMB: the Cosmic Background Explorer[1] (COBE), the Wilkinson Microwave Anisotropy Probe[2] (WMAP) and *Planck*[3] (as well as numerous smaller mainly ground-based experiments). As a result the CMB has yielded a wealth of high-precision cosmological data and provides the best current constraints on cosmological parameters, culminating in the most recent definitive measurements of the angular power spectrum of temperature anisotropies shown in Figure 1.4. This illustrates both the high precision of the latest *Planck* measurements across all scales, and also impressive agreement with the ΛCDM model.

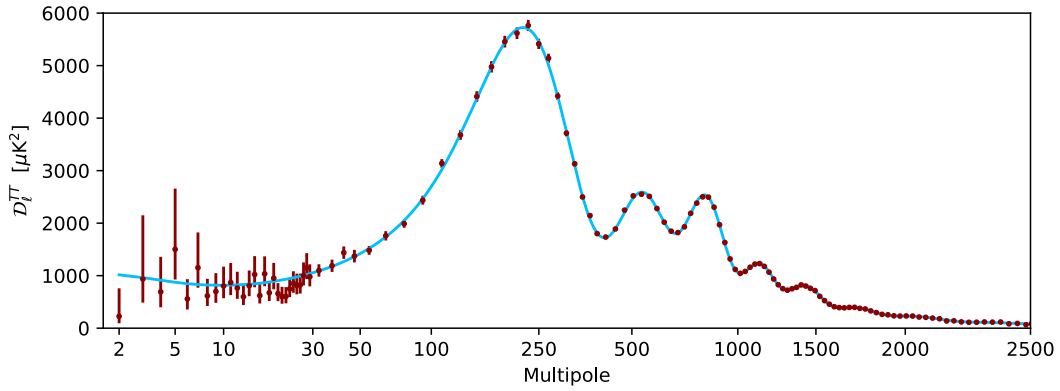

Figure 1.4: Angular power spectrum of CMB temperature anisotropies. Red dots and error bars are the latest *Planck* foreground-subtracted, frequency-averaged measurements. Blue lines show the best-fitting ΛCDM model. From Figure 9 of Akrami et al. (2018).

[1] http://science.nasa.gov/missions/cobe
[2] https://map.gsfc.nasa.gov
[3] sci.esa.int/planck/

Scientific results which are supported by CMB data include

- confirmation of the Big Bang theory of the origin of the Universe;
- confirmation of the theoretical prediction that the CMB has an almost perfect blackbody spectrum;
- confirmation that to the limits of current measurement the CMB is homogeneous, isotropic and Gaussian, which supports inflationary theory;
- evidence that primordial curvature fluctuations are adiabatic (which rules out certain forms of inflation), Gaussian and nearly scale-invariant;
- evidence that the Universe is very close to being flat;
- support for the existence of dark matter and dark energy;
- tight constraints on the values of cosmological parameters. The most recent results from the Planck satellite are shown in Table 1.2.

Parameter	**68% confidence limits from *Planck* 2018 results**
Baryon density $\Omega_{\rm b}h^2$	0.02237 ± 0.00015
Dark matter density $\Omega_{\rm c}h^2$	0.1200 ± 0.0012
Angular acoustic scale $100\theta_{\rm MC}$	1.04092 ± 0.00031
Optical depth τ	0.0544 ± 0.0073
Perturbation amplitude $\ln(10^{10}A_{\rm s})$	3.044 ± 0.014
Scalar spectral index $n_{\rm s}$	0.9649 ± 0.0042

Table 1.2: 68% parameter confidence limits from *Planck* CMB temperature and polarisation power spectra (including lensing reconstruction). These are the parameters which are most readily measured from the CMB. Other commonly-used parameters can be derived from them. From Table 7 in Akrami et al. (2018).

Because there are degeneracies in parameter estimates from the CMB only (for example between $\Omega_{\rm K}$ and $\Omega_{\rm m}$), even better estimates can be obtained by combining CMB data with other tracers such as baryon acoustic oscillations or weak lensing. Future CMB research will focus on synergies between probes and on exploiting additional information from CMB polarisation (Staggs et al. 2018).

1.2.2 Baryon acoustic oscillations

Baryon acoustic oscillations are features discernible in both the CMB and large-scale structure which can be used as standard rulers to investigate dark energy and the growth of structure.

After inflation, but before recombination, the Universe contained relativistic species (photons and neutrinos), non-relativistic baryons (mainly electrons and ionised hydrogen and helium nuclei), dark matter, and dark energy. The evolution of these components, in particular the matter components, forms the basis of later structure formation. Each component can be treated as a separate ideal fluid, but it is possible to simplify the model further, first by ignoring dark energy which simply provides a background to the evolution of the other species, and secondly by invoking the tight-coupling approximation which treats the electrons and photons (and nuclei) as a single fluid. This is justified because the electrons and photons are coupled by Thomson scattering and the electrons are coupled to the nuclei. Thus we can consider a universe with three fluid components: a baryon-photon plasma, dark matter, and neutrinos. In the following discussion we neglect the neutrinos for simplicity. In reality neutrinos do affect the background expansion rate in the radiation-dominated era and would be included in a more complete explanation.

We can then investigate the evolution of a perturbation to the baryon-photon density field with characteristic scale λ. This was first explored by Peebles and Yu (1970). They considered four regimes:

- While $\lambda \gg ct$ perturbations are adiabatic and grow with time.
- Once $ct \gg \lambda$, but before recombination, each perturbation oscillates like an acoustic wave.
- After recombination starts photons are much less likely to collide with baryonic matter particles so the wave becomes attenuated.
- Once recombination is sufficiently complete the photons can free-stream away. Baryonic matter can now collapse under gravity to form bound systems, the seeds of galaxies and clusters. Crucially the baryonic matter field retains residual features created by the acoustic oscillations.

The interesting part of this sequence is the second regime where perturbations in the baryonic matter density field feel the influence of outward radiation pressure from the relativistic species and are also affected by the gravitational potential of the non-relativistic species including dark matter. We can consider a perturbation δ_b in the baryonic matter density which forms a plane wave with wavenumber k. The evolution of a single Fourier mode of the perturbation has been derived as (Peebles and Yu 1970; Hu and White 1996; Eisenstein et al. 2007)

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\left[(1+R)\dot{\delta_\mathrm{b}}\right] + \frac{c^2k^2}{3}\delta_\mathrm{b} = -k^2(1+R)\Psi - \frac{\mathrm{d}}{\mathrm{d}\tau}\left[\frac{3(1+R)}{c^2}\dot{\Phi}\right], \tag{1.54}$$

where $\tau = \int \mathrm{d}t/a$ is the conformal time, $R = 3\rho_\mathrm{b}/4\rho_\mathrm{r}$ is the ratio of the baryon and photon densities, Ψ is the perturbation to the spatial curvature and Φ is the Newtonian potential. Overdots denote derivatives with respect τ. If there is no anisotropic stress $\Psi = \Phi$.

This is the equation of a driven oscillator with frequency $c_s k$ where c_s is the speed of sound which in this case is $c/\sqrt{3(1+R)}$. Thus the baryon-photon perturbation undergoes oscillations, known as baryon acoustic oscillations or BAOs. Meanwhile the dark matter simply collapses under gravity to form a higher density region around the site of the original perturbation. At first the oscillations are dominated by the photon pressure but as the Universe expands R increases, decreasing the sound speed. Eventually the photons decouple from the baryons and are able to free-stream away. Slightly later baryons cease to feel any effect of photons and the oscillations cease. The baryon density perturbation stalls in a shell at a distance from the central dark matter peak equal to the sound horizon at this time which is $r_s(\tau) = \int_o^\tau \mathrm{d}\tau' c_s(\tau')$.

This is most easily pictured in configuration space. Figure 1.5, taken from Eisenstein et al. (2007), shows snapshots of the sequence of events. The top row shows an initial delta-function pulse of photons and baryons travelling outwards, which also creates a wake in the central dark matter overdensity. The second row shows the situation at recombination. The photons start to leak away, leaving a shell of baryons at the sound horizon distance from the central peak. Finally in the bottom row the photons play no further part and both the dark matter and baryons fall into the pre-existing overdensities. There is an increased probability that at later times galaxies will form within these residual overdensities.

The radius of the baryon shell depends on the sound speed and the propagation time of the wave, which in turn depend on cosmological parameters. The sound speed depends on the ratio of baryons to photons and hence on $\Omega_\mathrm{b}h^2$. The propagation time depends on the expansion rate at the relevant epoch. This in turn depends partly on the redshift of matter-radiation equality which is a function of $\Omega_\mathrm{m}h^2$ only (Eisenstein et al. 2007). Using CMB constraints on $\Omega_\mathrm{b}h^2$ and $\Omega_\mathrm{m}h^2$ at decoupling (Bassett and Hlozek 2010), together with the CMB temperature, a value $r_s \approx 150\ \mathrm{Mpc}$ has been obtained for the present day (comoving) value of the sound horizon.

As an aside, in the absence of dark matter the oscillations would still occur but the excess density would be seen only in the spherical shell. So we would expect to see an excess of galaxies at all distances up to $2r_s$ apart, with the correlation function falling away at larger distances. We do not in fact observe this, which substantiates the existence of dark matter.

The upshot is that we can expect the galaxy power spectrum or two-point correlation function to reveal an excess of galaxies a distance r_s apart. In practice statistical methods must be used to

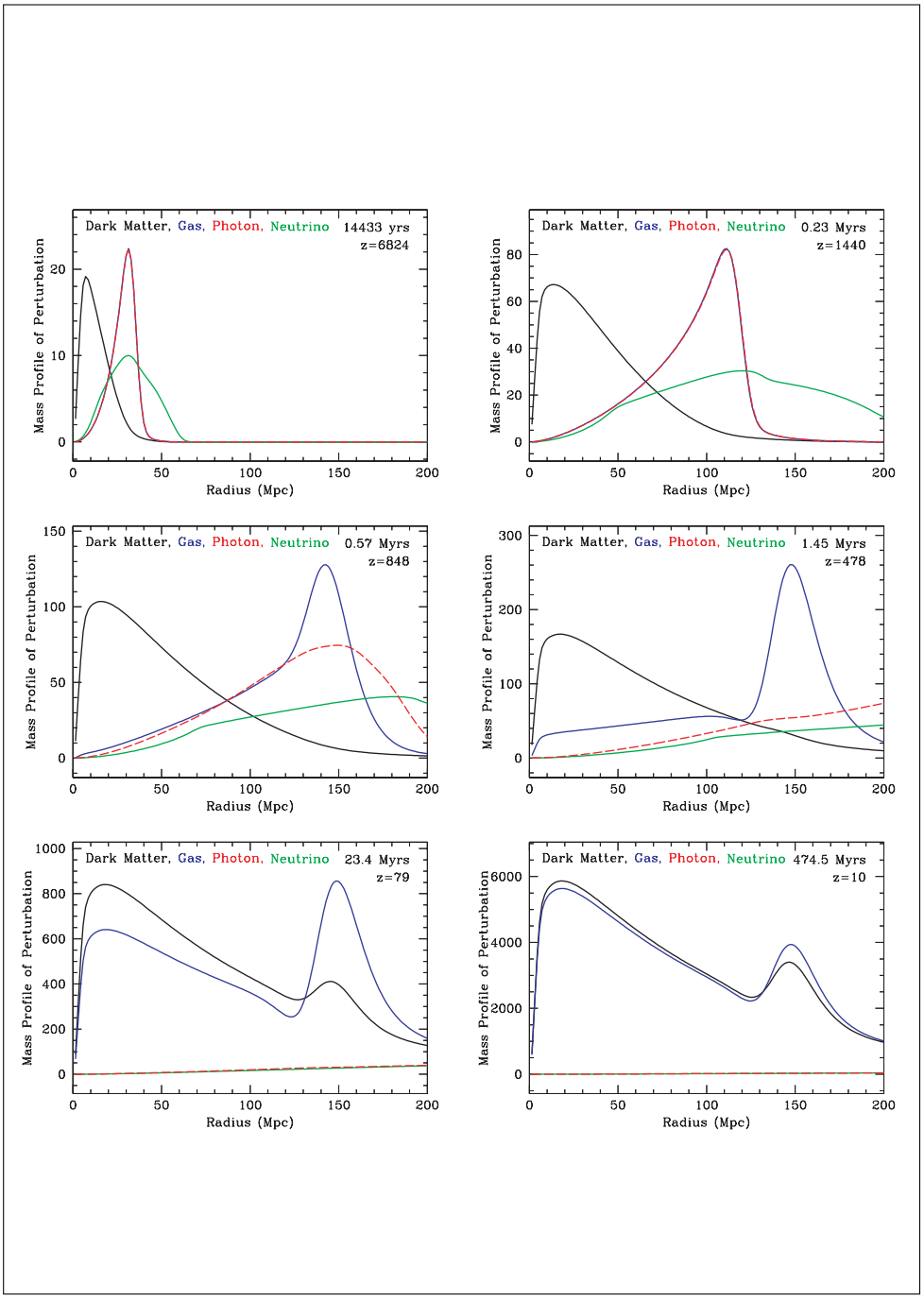

Figure 1.5: Snapshots of the radial mass profiles of dark matter, baryons, photons and neutrinos as an initially point-like overdensity evolves. The plots show fractional perturbations for each species, and the perturbations for relativistic species have been rescaled to match those of non-relativistic species. At first the photons and baryons travel outwards in unison, leaving a wake of dark matter. At recombination the photons leak away from the baryonic pulse. Two overdensities remain: dark matter close to the origin of the perturbation, and baryons in an outer shell. Subsequently both dark matter and baryons fall into the overdensities. From Figure 1.2 of Eisenstein et al. (2007)

.

disentangle the signal because we do not have just a single perturbation but many perturbations emanating from all points. Despite this complication the BAO signal forms a very robust standard ruler, which can be used to infer the expansion history of the Universe and to constrain Ω_Λ and the Hubble parameter (Eisenstein et al. 1998; Seo and Eisenstein 2003). It is independent of and complementary to other distance measurement methods such as those using supernovae (Blake and Glazebrook 2003).

The first detections of the BAO signal were made by in 2005 using data from the Sloan Digital Sky Survey[4] (SDSS) (Eisenstein et al. 2005) and from the 2dF Galaxy Redshift Survey (Cole et al. 2005). The Baryon Oscillation Spectroscopic Survey (BOSS), which is part of the SDSS, produced the first BAO measurements with more than 1% precision (Anderson et al. 2012) and has continued to yield increasingly precise data (Alam et al. 2017). BAO measurements are now a standard part of the cosmological toolkit. Figure 1.6 shows illustrative recent results from BOSS, indicating the shape and scale of the BAO feature in the galaxy two-point correlation function. The key result is the monopole measurement, shown in blue.

[4] sdss.org

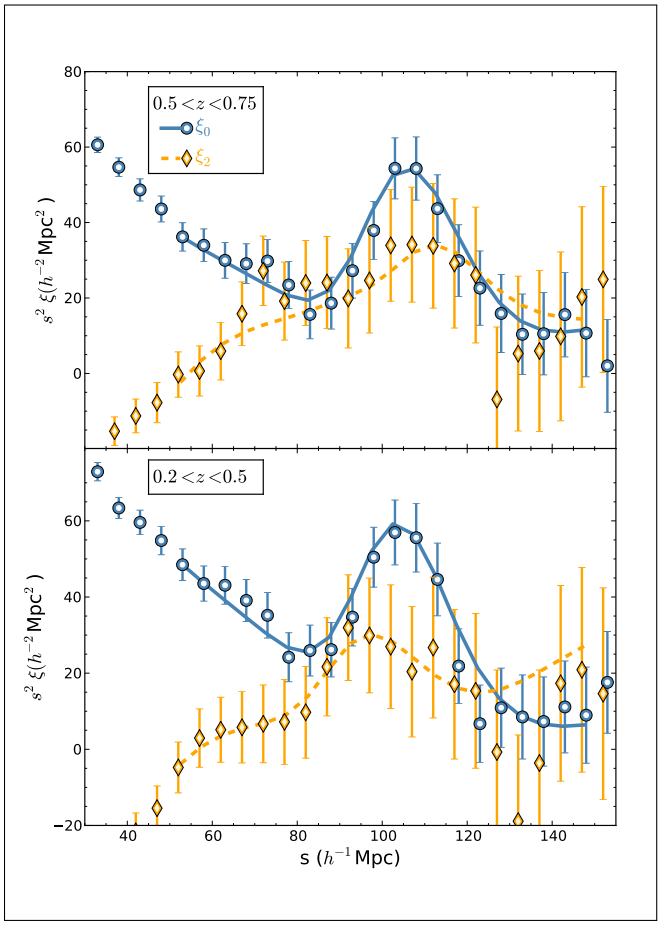

Figure 1.6: The measured BAO signal in the monopole (blue) and quadrupole (orange) of the two-point correlation function of BOSS galaxies plotted against the redshift-space separation, s of pairs of galaxies. Points and error bars are from measurements. Lines indicate bestfit models. The upper and lower panels show results for two different redshift ranges. From Figure 11 in Ross et al. (2016).

1.2.3 Weak gravitational lensing

One prediction of general relativity is that massive objects deflect the paths of light rays. As a result the images we observe of sources such as distant galaxies are distorted. By analogy with optical lensing this effect is known as gravitational lensing. In this section we discuss mainly *weak* gravitational lensing and in particular cosmic shear – the lensing which results from the passage of light through the large-scale inhomogeneous matter distribution. This can be contrasted with strong lensing which is caused by very large gravitational potentials along the line of sight, resulting in arcs and multiple images.

Weak lensing is a very subtle effect which cannot be detected for any individual source. However it can be measured statistically by observing many sources. It is invaluable for cosmology because it is a direct probe of the matter distribution and can also be used to study dark energy.

There are many comprehensive reviews of weak lensing. This section draws in particular on Bartelmann and Schneider (2001), Kilbinger (2015) and Bartelmann and Maturi (2016). We derive the equations governing the deflection of light by large-scale structure, introduce the key concepts of convergence and shear and their statistical properties, and discuss some issues related to measurement.

To understand lensing we need to revisit the metric (Equation 1.4) in the presence of a gravitational field which alters the spacetime curvature and hence the paths of photons. We assume isotropy and introduce the Newtonian gravitational potential Φ with $|\Phi|/c^2 \ll 1$. Locally (that is, for a small region) we can use the perturbed Minkowski metric

$$\mathrm{d}s^2 = c^2\left(1+\frac{2\Phi}{c^2}\right)\mathrm{d}t^2 - \left(1-\frac{2\Phi}{c^2}\right)\mathrm{d}\mathbf{r}^2 \,. \tag{1.55}$$

Light travels along null geodesics ($\mathrm{d}s^2 = 0$) so, using a Taylor expansion in $|\Phi|/c^2$, the effective light speed is

$$\left|\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\right| = c\left(1+\frac{2\Phi}{c^2}\right) \,, \tag{1.56}$$

and the refractive index, which quantifies the change in direction of the light path, is $1-2\Phi/c^2$. We can now apply Fermat's principle that light takes the path of least time between any two points and obtain an equation for the (vector) deflection angle $\hat{\boldsymbol{\alpha}}$, the difference in angle between the emitted and observed light rays. To do this we take the gradient of the potential perpendicular to the ray

and integrate along the ray, so that

$$\hat{\boldsymbol{\alpha}} = -\frac{2}{c^2}\int \nabla_\perp \Phi \, \mathrm{d}r \, . \tag{1.57}$$

This is twice the prediction obtained from Newtonian physics (assuming photons have mass) (von Soldner 1804), because the metric has equal perturbations in the time and the space parts.

Figure 1.7 shows the basic geometry of gravitational lensing. This simplified depiction is based on the thin lens approximation which assumes that the lens is thin compared with the overall extent of the lensing system. This is clearly not valid for lensing by large-scale structure but serves to illustrate the basic set-up. In particular we can see that

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \boldsymbol{\alpha} \, , \tag{1.58}$$

where $\boldsymbol{\beta}$ is the angle between the source and the optical axis and $\boldsymbol{\theta}$ is the angle between the image and the optical axis, $\boldsymbol{\alpha}$ is the reduced deflection angle measured by the observer.

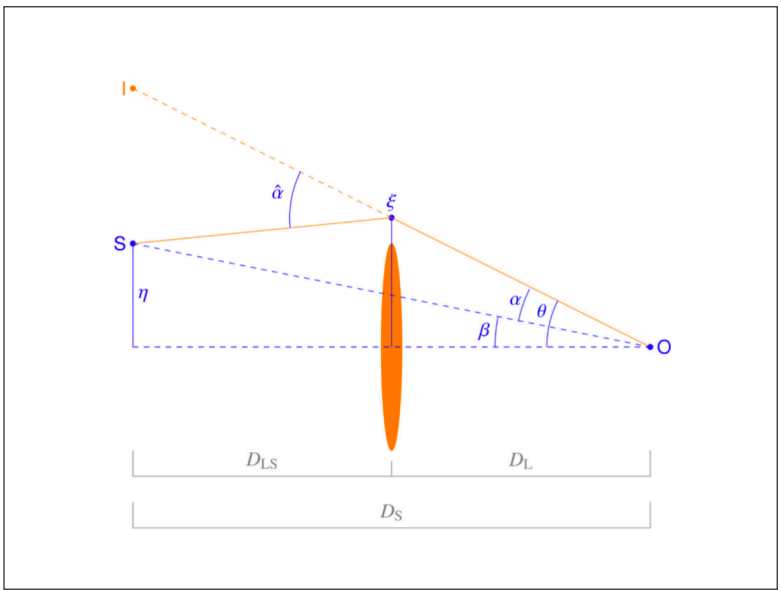

Figure 1.7: Sketch of a gravitational lens system. The observer, source and image are denoted by O, S and I respectively. D_L, D_LS and D_S are respectively the distances from the observer to the lens, from the lens to the source, and from the observer to the source. The lens is at position $\boldsymbol{\xi}$ and the source is at position $\boldsymbol{\eta}$. A light ray from the source is deflected by the angle $\hat{\boldsymbol{\alpha}}$, measured at the lens. The observer measures the reduced deflection angle $\boldsymbol{\alpha}$. $\boldsymbol{\beta}$ is the angle between the source and the optical axis and $\boldsymbol{\theta}$ is the angle between the image and the optical axis. From Figure 1 in Bartelmann and Maturi (2016).

In the weak lensing regime we are interested in very small deflections of light rays as they travel through the large-scale structure. Consider two light rays on neighbouring geodesics with transverse comoving separation $\boldsymbol{x}(\chi)$ travelling towards an observer. If there were no deflectors along the light paths the observer would see an angular separation $\boldsymbol{\beta}$ which depends on the geometry of the Universe and satisfies

$$\boldsymbol{x}(\chi) = f_K(\chi)\boldsymbol{\beta}\,, \tag{1.59}$$

where $f_K(\chi)$ is given by Equation 1.7.

Suppose now that there are deflectors along the light paths and consider a single deflector on one light path at a distance χ' from the observer ($\chi' < \chi$). From Equation 1.57, this deflects the ray by $\mathrm{d}\boldsymbol{\alpha} = -(2/c^2)\nabla_\perp\Phi(\boldsymbol{x},\chi')\mathrm{d}\chi'$. To obtain the total separation induced by all deflectors we integrate along the path of the ray. The same applies to the second ray. If the observed angular separation caused by the combined effect of all deflections along the paths of both light rays is $\boldsymbol{\theta}$, we have

$$\boldsymbol{x}(\chi) = f_K(\chi)\boldsymbol{\theta} - \frac{2}{c^2}\int_0^\chi \mathrm{d}\chi' f_K(\chi-\chi')(\nabla_\perp\Phi(\boldsymbol{x},\chi') - \nabla_\perp\Phi^{(0)}(\chi'))\,, \tag{1.60}$$

where $\nabla_\perp\Phi^{(0)}$ is the gradient of the potential along the second light ray. Comparing with the no-deflection situation, we again obtain the lens equation

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \boldsymbol{\alpha}\,, \tag{1.61}$$

where

$$\boldsymbol{\alpha} = \frac{2}{c^2}\int_0^\chi \mathrm{d}\chi' \frac{f_K(\chi-\chi')}{f_K(\chi)}[\nabla_\perp\Phi(\boldsymbol{x},\chi') - \nabla_\perp\Phi^{(0)}(\chi')]\,. \tag{1.62}$$

The reduced deflection angle can be expressed as the gradient of a two-dimensional potential, the lensing potential ψ, so that

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \nabla\psi\,. \tag{1.63}$$

We observe the angle $\boldsymbol{\theta}$ and want to infer $\boldsymbol{\beta}$. In weak lensing the deflections and angles are small which means there are several valid approximations which make this inference easier.

First, because the deflections are small we can take the light paths to be unperturbed, which simplifies the integration. This is the Born approximation. We can also ignore any coupling between deflections at different redshifts. Finally, the separation of the rays is small so we can treat both rays as coincident. With these approximations we can find the total deflection simply by integrating along a single ray. This leads to

$$\psi(\boldsymbol{\theta}, \chi) = \frac{2}{c^2} \int_0^\chi \mathrm{d}\chi' \frac{f_K(\chi - \chi')}{f_K(\chi) f_K(\chi')} \Phi(f_K(\chi')\boldsymbol{\theta}, \chi') \; . \tag{1.64}$$

We can linearise this to define a matrix $\mathbf{A}$, the Jacobian matrix of the lens mapping

$$\frac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\theta}} = \mathbf{A} \; . \tag{1.65}$$

The Jacobian describes how the lensed image is mapped onto the source. Its components are

$$A_{ij} = \frac{\partial \beta_i}{\partial \theta_j} \tag{1.66}$$

$$= \delta_{ij} - \frac{\partial^2 \psi}{\partial \theta_i \partial \theta_j} \; . \tag{1.67}$$

Another way of writing this is

$$\mathbf{A} = \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{pmatrix} \; . \tag{1.68}$$

This introduces the convergence κ and shear $\gamma = \gamma_1 + \mathrm{i}\gamma_2$ which are second derivatives of the lensing potential:

$$\kappa = \frac{1}{2}\left(\frac{\partial^2 \psi}{\partial \theta_1^2} + \frac{\partial^2 \psi}{\partial \theta_2^2} \right) \tag{1.69}$$

$$\gamma_1 = \frac{1}{2}\left(\frac{\partial^2 \psi}{\partial \theta_1^2} - \frac{\partial^2 \psi}{\partial \theta_2^2} \right) \tag{1.70}$$

$$\gamma_2 = \frac{\partial^2 \psi}{\partial \theta_1 \partial \theta_2} \; . \tag{1.71}$$

The convergence quantifies the isotropic increase or decrease in the image size. In contrast the shear quantifies anisotropic stretching which changes circular sources into ellipses.

In practice we need to know how sources are mapped onto images (rather than the other way round) which is given by the inverse of the Jacobian matrix. For weak lensing we can assume the

Jacobian is always invertible because both κ and γ are very small so that $\det\mathbf{A} \equiv (1-\kappa^2) - \gamma^2$, where $\gamma^2 = \gamma_1^2 + \gamma_2^2$, is close to one. The inverse Jacobian is

$$\mathbf{A}^{-1} = \frac{1}{\det\mathbf{A}} \begin{pmatrix} 1-\kappa+\gamma_1 & \gamma_2 \\ \gamma_2 & 1-\kappa-\gamma_1 \end{pmatrix} . \tag{1.72}$$

The prefactor is referred to as the magnification μ. It is equal to the ratio of the image flux to the source flux and can be approximated as

$$\begin{aligned} \mu &= \frac{1}{(1-\kappa^2)-\gamma^2} \\ &\approx 1 + 2\kappa \, . \end{aligned} \tag{1.73}$$

The change in flux is a consequence of the fact that lensing conserves surface brightness. Since the apparent size of a source changes it follows that its flux must also change. This means that magnification affects the number density of sources. In a flux-limited galaxy survey lensing may push sources above the flux limit, increasing the number density. However lensing also stretches the region of sky behind the lens, reducing the number density. The net effect depends on the slope of the cumulative flux distribution: the steeper the slope, the greater the increase in number density.

The convergence is related to the underlying density contrast δ. To see this we invoke the Poisson equation

$$\nabla^2\Phi = 4\pi G a^2 \bar{\rho}\delta \, , \tag{1.74}$$

where $\bar{\rho} = a^{-3}\rho_\mathrm{c}$ is the mean density of the Universe and $\rho_\mathrm{c} = 3H_0^2/8\pi G$ is the critical density. We can then relate the two-dimensional Laplacian of ψ to the three-dimensional Laplacian of Φ, relate the mean matter density to the critical density, and arrive at (Kilbinger 2015; Bartelmann and Maturi 2016)

$$\kappa(\boldsymbol{\theta}, \chi) = \frac{3H_0^2\Omega_\mathrm{m}}{2c^2} \int_0^\chi \frac{\mathrm{d}\chi'}{a(\chi')} \frac{f_K(\chi-\chi')}{f_K(\chi)} f_K(\chi')\delta(f_K(\chi')\boldsymbol{\theta}, \chi') \, . \tag{1.75}$$

When applied to a galaxy survey the mean convergence $\kappa(\boldsymbol{\theta})$ is obtained by weighting by the probability distribution of galaxies $p(\chi)$ and integrating to the maximum comoving distance of the

survey $\chi_{\rm lim}$:

$$\kappa(\boldsymbol{\theta}) = \int_0^{\chi_{\rm lim}} \mathrm{d}\chi\, p(\chi)\kappa(\boldsymbol{\theta},\chi) \tag{1.76}$$

$$= \frac{3H_0^2\Omega_{\rm m}}{2c^2}\int_0^{\chi_{\rm lim}} \frac{\mathrm{d}\chi}{a(\chi)} g(\chi)\, f_K(\chi)\delta(f_K(\chi)\boldsymbol{\theta},\chi) \tag{1.77}$$

$$= \int_0^{\chi_{\rm lim}} \mathrm{d}\chi\, q(\chi)\delta(f_K(\chi)\boldsymbol{\theta},\chi)\ , \tag{1.78}$$

where we have introduced the lensing efficiency

$$g(\chi) = \int_\chi^{\chi_{\rm lim}} \mathrm{d}\chi'\, p(\chi')\frac{f_K(\chi'-\chi)}{f_K(\chi')}\ , \tag{1.79}$$

and, for convenience later, the weight function (Joachimi and Bridle 2010)

$$q(\chi) = \frac{3H_0^2\Omega_{\rm m}}{2c^2}\frac{f_K(\chi)}{a(\chi)} g(\chi)\ . \tag{1.80}$$

This shows that the convergence is a projection of the matter density along the line of sight which depends on the geometry of the Universe and the source galaxy distribution. This is why it is such a useful probe of both dark energy and dark matter. In practice the convergence is only accessible through the magnification, which is difficult to measure. The shear is more readily observable and is the quantity which is normally measured in weak lensing surveys. (Equations 1.69–1.71 can be used to derive the convergence from the shear).

For a flat Universe $f_K(\chi) = \chi$ and the lensing efficiency is

$$g(\chi) = \int_\chi^{\chi_{\rm lim}} \mathrm{d}\chi'\, p(\chi')\frac{\chi'-\chi}{\chi'}\ . \tag{1.81}$$

Moreover if the Universe is flat the expression in Equation 1.75 has a maximum when $\chi' \approx \chi/2$ so that structures about half the distance to the source have the greatest lensing effect.

Weak lensing power spectrum

By definition the convergence and shear both have zero mean, $\langle\kappa\rangle = \langle\gamma\rangle = 0$, so to describe the lensing field we need at least second-order statistics – either the two-point correlation function or the power spectrum. There are several different ways to represent the weak lensing power spectrum (Kitching et al. 2017). Firstly, since we observe on the celestial sphere, we can expand the shear in terms of spherical harmonics in the angular dimension and spherical Bessel functions in the radial

dimension. It is also possible to make use of the redshift of each galaxy separately by expanding in spherical harmonics on the sky and also expanding along the line of sight, an approach known as three-dimensional cosmic shear (Heavens 2003). These formalisms yield multi-dimensional oscillating integrals and it is common to make simplifying assumptions. In particular most cosmic shear studies make use of two specific simplifications, the Limber and flat-sky approximations, and also employ tomography which divides the three-dimensional field into two-dimensional slices, as originally suggested by Hu (1999).

The Limber approximation (Limber 1953; Kaiser 1992; LoVerde and Afshordi 2008) relates the power spectrum of a three-dimensional isotropic random field to its two-dimensional projection. It states that if a quantity $x(\boldsymbol{\theta})$ is a two-dimensional projection of a three-dimensional quantity $y(f_K(\chi)\boldsymbol{\theta}, \chi)$ with weight function $w(\chi)$, so that

$$x(\boldsymbol{\theta}) = \int_0^{\chi_\mathrm{H}} \mathrm{d}\chi \; w(\chi) y(f_K(\chi)\boldsymbol{\theta}, \chi) \; , \tag{1.82}$$

then the power spectrum can be approximated as

$$P_x(\ell) = \int_0^{\chi_\mathrm{H}} \mathrm{d}\chi \; \frac{w^2(\chi)}{f_K^2(\chi)} P_y\left(k = \frac{\ell}{f_K(\chi)}, \chi\right) . \tag{1.83}$$

This approximation is justified when the weight function w varies on length scales much greater than those of the quantity y.

The expression for the convergence is just such a weighted projection with weight function $q(\chi)$ defined by Equation 1.80.

Thus so long as the condition on $q(\chi)$ is satisfied, we can approximate the power spectrum as

$$P_\kappa(\ell) = \int_0^{\chi_\mathrm{H}} \mathrm{d}\chi \; q^2(\chi') P_\delta\left(\frac{\ell}{f_K(\chi)}, \chi\right) . \tag{1.84}$$

The flat-sky approximation treats a small angular region on the sky as a plane perpendicular to the line of sight. This is often a valid assumption for current lensing studies. It means that instead of using spherical harmonics we can expand the convergence (or shear) in Fourier modes to give

$$\kappa(\theta) = \int \frac{\mathrm{d}^2\boldsymbol{\ell}}{(2\pi)^2} \tilde{\kappa}(\boldsymbol{\ell}) \mathrm{e}^{\mathrm{i}\boldsymbol{\ell}\cdot\boldsymbol{\theta}} \; . \tag{1.85}$$

Both these approximations have been widely used in lensing studies. However their validity cannot be taken for granted, especially in the context of the next generation of wide-field surveys. For

example Kitching et al. (2017) and Kilbinger et al. (2017) both report careful investigations of the validity of these and other common approximations.

Weak lensing tomography

Although weak lensing observations are essentially two-dimensional because they are projections onto the sky, it is possible to obtain information about the third dimension. This is highly desirable because in two dimensions effects such as the growth of structure and the time evolution of dark energy are integrated over. In principle the method of three-dimensional cosmic shear could be used, exploiting redshifts for individual galaxies, but it is more common to use tomography in which source galaxies are binned into redshift ranges – essentially a form of data compression. Each redshift bin has a separate lensing efficiency and a separate shear. This approach gives us a practical way to investigate the third dimension and is used in most weak lensing studies, for example Köhlinger et al. (2015); Hildebrandt et al. (2016); Abbott et al. (2018); Hildebrandt et al. (2020).

However tomography has some drawbacks. In particular the tomographic bins are generally correlated, mainly because components of the large-scale structure contribute to more than one bin and also because generally only photometric redshifts are available, introducing extra uncertainty into the binning.

Nevertheless tomographic weak lensing can now produce tight constraints on cosmological parameters which are complementary to other probes. This is illustrated in Figure 1.8 which shows constraints on $\Omega_{\rm m}$ and σ_8 from four tomographic weak lensing studies together with the latest CMB analysis. (This figure in fact shows some tension between the lensing and CMB results. It is not yet clear whether this due to unrecognised systematic errors or has a deeper explanation.)

Weak lensing two-point correlation functions

In practice the most readily observable quantity is not the power spectrum but the two-point correlation function. Commonly the shear two point correlation functions are expressed in terms of tangential and cross-components of the shear, γ_t and $\gamma_\times$. These are transformations of the shear components in Equations 1.70 and 1.71. In these equations the real and imaginary parts of the shear are defined relative to a Cartesian coordinate frame. The tangential and cross components are

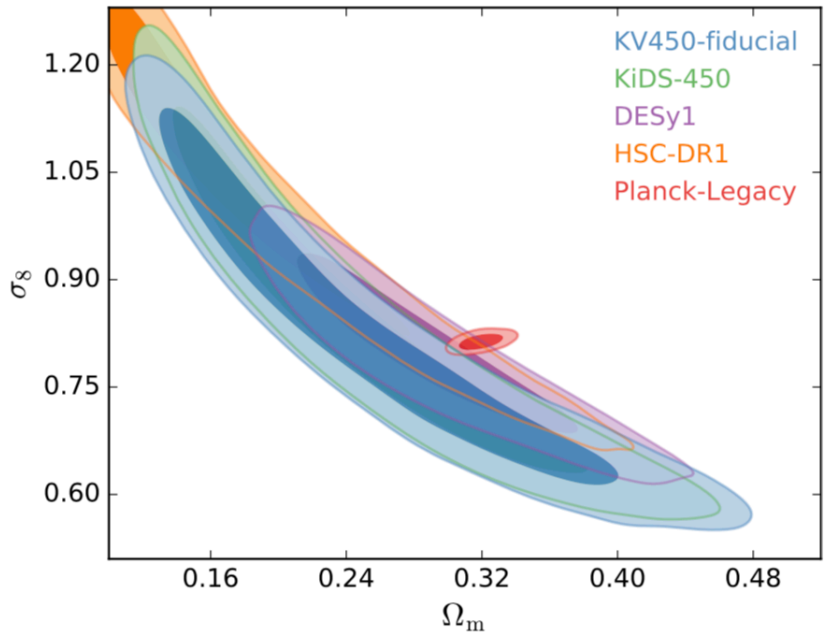

Figure 1.8: Parameter constraints from tomographic weak lensing surveys. Marginalised posterior contours (68% and 95% confidence levels) in the Ω_m–σ_8 plane for four tomographic weak lensing surveys (blue, green, purple and orange contours) and the most recent *Planck* CMB analysis (red contours). From Figure 4 in Hildebrandt et al. (2020) which gives full details of all the underlying surveys.

defined in a reference frame rotated by an angle ϕ relative to the original frame (Schneider 2006):

$$\gamma_t = -\mathrm{Re}[\gamma \mathrm{e}^{-2\mathrm{i}\phi}] \,, \tag{1.86}$$

$$\gamma_\times = -\mathrm{Im}[\gamma \mathrm{e}^{-2\mathrm{i}\phi}] \,. \tag{1.87}$$

The factor of 2 in the exponentials arises from the fact that the shear is a polar quantity. From these components two non-vanishing two-point correlators can be formed, $\langle \gamma_t \gamma_t \rangle$ and $\langle \gamma_\times \gamma_\times \rangle$ (the cross-correlator is zero for parity reasons). It is normal to combine these into two components of the two-point correlation function:

$$\xi_+(\theta) = \langle \gamma_t \gamma_t \rangle + \langle \gamma_\times \gamma_\times \rangle \tag{1.88}$$

$$\xi_-(\theta) = \langle \gamma_t \gamma_t \rangle - \langle \gamma_\times \gamma_\times \rangle \,. \tag{1.89}$$

Observationally these functions are often more convenient to use than the power spectrum partly because they are not affected by the survey mask. They can be expressed as Hankel transforms of

the convergence (or shear) power spectrum:

$$\xi_+(\theta) = \frac{1}{2\pi} \int \mathrm{d}\ell\ell J_0(\ell\theta) P_\kappa(\ell) \,, \tag{1.90}$$

$$\xi_-(\theta) = \frac{1}{2\pi} \int \mathrm{d}\ell\ell J_4(\ell\theta) P_\kappa(\ell) \,. \tag{1.91}$$

These functions mix the gradient (E) and curl (B) modes of the lensing signal. Filtered versions of the correlation functions have been developed to separate the E-and B-modes (Schneider 1996; Schneider et al. 2010). Since the pure lensing signal is curl-free, detection of B-modes may point to the presence of uncorrected systematics.

1.2.3.1 Weak lensing surveys

Weak lensing analysis requires extremely high image quality with a small point spread function, supplemented by accurate redshift measurements to aid tomographic analysis. In practice it is generally not feasible to obtain spectroscopic redshifts for each of the tens of millions of observed galaxies but ideally surveys should be designed so that photometric measurements can be calibrated against spectroscopic counterparts. Obviously a wide survey footprint and large redshift range are also advantageous. Past surveys were not necessarily optimised for weak lensing but current and planned surveys have been developed specifically to exploit the technique.

Later chapters refer in particular to one current survey, the Kilo-Degree Survey (KiDS) (de Jong et al. 2013; Kuijken et al. 2015), and one planned survey, the European Space Agency *Euclid* mission (Laureijs et al. 2011).

KiDS[5] is a 1500 square degrees optical survey using the Very Large Telescope (VLT) Survey Telescope. It observes the southern sky in four filters (*u*, *g*, *r*, *i*), and is optimised to produce high quality cosmic shear measurements. Observations began in 2011 and the first science results, based on the initial survey of 450 square degrees, were published in 2015 (Viola et al. 2015). Eventually the survey area will extend over two 750 square degree areas. KiDS fully overlaps with the VISTA Kilo-Degree Infrared Galaxy Survey[6] (VIKING) (Edge et al. 2013). Using infrared data from VIKING improves photometric redshifts, especially at high redshift (Hildebrandt et al. 2020).

Euclid[7] is a space-based mission due to be launched in 2021 and to continue observations until 2027. It will survey 15,000 square degrees of the sky at optical and near-infrared wavelengths. Like KiDS, the *Euclid* survey is optimised for weak lensing and will produce around 10^7 spectroscopic

[5] http://kids.strw.leidenuniv.nl/index.php
[6] https://www.eso.org/sci/observing/PublicSurveys/sciencePublicSurveys.html
[7] http://sci.esa.int/euclid/

redshifts as well as measuring galaxy shapes and photometric redshifts of over 10^9 galaxies. *Euclid* will be highly complementary to other next-generation surveys such as the ground-based Dark Energy Spectroscopic Instrument[8] and Rubin Observatory Legacy Survey of Space and Time (LSST) [9], and NASA's space-based Wide Field Infrared Survey Telescope[10].

1.2.3.2 Systematics in weak lensing

In any weak lensing analysis numerous systematic uncertainties need to be controlled. Some are caused by observational limitations, for example measured galaxy shapes can be distorted by the atmosphere, by detector or camera effects or by inaccuracies in the modelling of the point spread function (Massey et al. 2012; Cropper et al. 2013). Another practical obstacle is the unavoidable use of photometric redshift measurements which introduces errors and biases into the redshift distributions of source galaxies. Other systematics stem from lack of knowledge. For example we only partially understand how galaxies form, especially the impact of baryonic physics on the shape and orientation of observed galaxies at non-linear scales. However perhaps the most important systematic uncertainty is the intrinsic alignment of galaxies which affects all lensing measurements made from a galaxy survey. During their formation galaxies become aligned with the local tidal gravitational field which is also the source of the lensing signal. This distorts the measured ellipticities of galaxies. We discuss this further in Chapter 3.

1.3 Methods and techniques

In this section we present some techniques and models which are referred to in later chapters. First we discuss summary statistics which quantify clustering and which are used throughout this thesis, then we outline the halo model which is particularly relevant to Chapter 3, and finally we explain methods of inference and the importance of covariance matrices which underpin the research described in Chapters 3 and 4.

1.3.1 Clustering statistics

In cosmology we are particularly interested in clustering statistics which can describe, for example, the distribution of large-scale over- and underdensities, or of CMB anistropies.

[8] www.desi.lbl.gov
[9] lsst.org
[10] www.nasa.gov/wfirst

A standard way to measure clustering is through correlation functions in real space. For example the two-point correlation function measures the excess clustering between pairs of points compared to a random distribution. However it is often simpler to work in Fourier space, for example because convolutions in real space become multiplications. For a statistically isotropic field the power spectrum is defined as the Fourier transform of the two-point correlation function, and similarly the bispectrum and three-point correlation function are a Fourier-transform pair (Bernardeau et al. 2002). So for observables $A, B, C \ldots$ in n dimensions we have

$$\langle A(\mathbf{k}_1)B(\mathbf{k}_2)\rangle = (2\pi)^n \delta(\mathbf{k}_1 + \mathbf{k}_2) P_{AB}(\mathbf{k}_1) \tag{1.92}$$

$$\langle A(\mathbf{k}_1)B(\mathbf{k}_2)C(\mathbf{k}_3)\rangle = (2\pi)^n \delta(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) B_{ABC}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) \tag{1.93}$$

$$\langle A(\mathbf{k}_1)B(\mathbf{k}_2)C(\mathbf{k}_3)D(\mathbf{k}_4)\rangle_{\mathrm{c}} = (2\pi)^n \delta(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4) T_{ABCD}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) \tag{1.94}$$

where P_{AB}, B_{ABC}, and T_{ABCD} are the power spectrum, bispectrum and trispectrum. The subscript 'c' indicates the connected part of the four-point correlation function: the part represented by a connected graph which cannot be decomposed into the product of lower-order correlation functions.

In two dimensions the two-point correlation function $\xi(r)$ and power spectrum $P(k)$ are related by

$$\xi^{2D}(r) = \frac{1}{(2\pi)^2} \int \mathrm{d}^2\mathbf{k}\, \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}} P(k) \tag{1.95}$$

$$= \frac{1}{2\pi} \int k\, \mathrm{d}k\, J_0(kr) P(k)\,, \tag{1.96}$$

where J_0 is a Bessel function of the first kind (resulting from cylindrical symmetry).

Similarly in three dimensions

$$\xi^{3D}(r) = \frac{1}{(2\pi)^3} \int \mathrm{d}^3\mathbf{k}\, \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}} P(k) \tag{1.97}$$

$$= \frac{1}{2\pi^2} \int k^2\, \mathrm{d}k\, j_0(kr) P(k)\,, \tag{1.98}$$

where j_0 is a spherical Bessel function (resulting from spherical symmetry).

Since surveys measure angles on the sky it is often most natural to consider angular correlation functions. The 2D angular correlation function is

$$\omega(\theta) = \frac{1}{(2\pi)^2} \int \mathrm{d}^2\boldsymbol{\ell}\, \mathrm{e}^{\mathrm{i}\boldsymbol{\ell}\cdot\theta} C_\ell\,, \tag{1.99}$$

where C_ℓ is the angular power spectrum.

In the case of a field which is projected onto the celestial sphere (such as the CMB) we can use the angular all-sky correlation function

$$\xi^{\mathrm{all-sky}}(\theta) = \sum_{\ell=0}^{\infty} \frac{2\ell+1}{4\pi} P_\ell(\cos(\theta)) C_\ell \ , \tag{1.100}$$

where P_ℓ are Legendre polynomials.

If a field is Gaussian (for example CMB temperature fluctuations) then two-point statistics encapsulate all the information in the field. So in this case there is nothing to be gained from studying higher-order statistics and the two-point correlation function or power spectrum provides an effective means of compressing all the information in the underlying data. However, many fields of interest to cosmology are highly non-Gaussian, including the matter density contrast at small scales and the weak lensing field. In these cases higher-order statistics can be expected to provide additional information, in particular extra constraints on cosmological parameters.

While two-point statistics can easily be visualised in two or three dimensions, even three-point statistics involve many more configurations and are difficult to handle theoretically or observationally. Often three-point results are presented for only a subset of possible triangle configurations, for example only equilateral triangles or only those triangles which make the greatest contribution to the bispectrum in question. Alternatively, computation can be reduced by efficient data compression (Tegmark et al. 1997).

When considering higher-order statistics of a Gaussian field it is often useful to employ Wick's theorem (Wick 1950), also known as Isserlis' theorem. This states that if a field δ is Gaussian then (in Fourier space)

$$\begin{aligned} \langle \delta(\mathbf{k}_1) \dots \delta(\mathbf{k}_{2p+1}) \rangle &= 0 \\ \langle \delta(\mathbf{k}_1) \dots \delta(\mathbf{k}_{2p}) \rangle &= \sum_{\text{all pairs}} \prod_{\substack{p \text{ pairs} \\ i,j, i \neq j}} \langle \delta(\mathbf{k}_i) \delta(\mathbf{k}_j) \rangle \ . \end{aligned} \tag{1.101}$$

Weak lensing clustering statistics

In the remainder of this thesis we make repeated use of weak lensing clustering statistics. These can all be expressed as integrals over their matter counterparts. Using the Limber and flat-sky

approximations and assuming a flat Universe, the convergence power spectrum can be written as

$$P_\kappa(\ell) = \int \mathrm{d}\chi \; q^2(\chi)\chi^{-2} P_\delta(k;\chi) \; , \tag{1.102}$$

where $k = \ell/\chi(z)$ from the Limber approximation and $q(\chi)$ is the lensing weight function defined by Equation 1.80.

The convergence bispectrum and trispectrum for a flat Universe can be expressed in a similar way as (Kayo and Takada 2013)

$$B_\kappa(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \boldsymbol{\ell}_3) = \int \mathrm{d}\chi \; q^3(\chi)\chi^{-4} B_\delta(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3; \chi) \tag{1.103}$$

$$T_\kappa(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \boldsymbol{\ell}_3, \boldsymbol{\ell}_4) = \int \mathrm{d}\chi \; q^4(\chi)\chi^{-6} T_\delta(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4; \chi) \; , \tag{1.104}$$

where again $\mathbf{k}_i = \boldsymbol{\ell}_i/\chi(z)$.

This can be generalised to spectra of any order (Kayo and Takada 2013):

$$P_{n,\kappa}(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2 \ldots \boldsymbol{\ell}_n) = \int \mathrm{d}\chi \; q^n(\chi)\chi^{-2(n-1)} P_{n,\delta}(\mathbf{k}_1, \mathbf{k}_2 \ldots \mathbf{k}_n) \; . \tag{1.105}$$

1.3.2 Halo model

The halo model (Cooray and Sheth 2002) is an analytical model which can be used to estimate power spectra and higher-order clustering statistics. It is widely used to avoid running large numbers of costly simulations and also as an alternative, or adjunct, to other analytical methods such as perturbation theory.

The model starts from the assumption that all mass in the Universe is in discrete virialised halos and that the density field can be described as a superposition of spherically symmetric halos (Mead et al. 2015). With these assumptions the distribution of matter can be modelled through the spatial distribution of halos and the distribution of matter within each halo. This means that we need expressions for the number density of halos of different mass, and for the distribution of matter within a halo (of a given mass at a given redshift). To estimate clustering statistics we also need assumptions about how points which are being correlated are related. The halo model assumes that points in two different halos are mediated through the linear matter power spectrum and perturbation theory. Points within a single halo are mediated by the distribution of matter in the halo. For points in different halos (but not those in a single halo) we need to take account of halo bias which arises because the distribution of halos does not exactly match the underlying matter

distribution: halos are more likely to form, or to form early, at the peaks of the matter density field. A comprehensive review of all aspects of bias is given in Desjacques et al. (2018).

Because this model involves convolutions of distributions it is defined in Fourier space so that the convolutions become multiplications. So we start with the Fourier transform of the matter density contrast at wavenumber $\mathbf{k}$. This can be expressed as an integral of the matter distributions of all contributing halos (Cooray and Hu 2001)

$$I_{\mu}^{\beta}(k_1, k_2, ..., k_{\mu}) \equiv \int \mathrm{d}M \frac{\mathrm{d}n}{\mathrm{d}M} \left(\frac{M}{\bar{\rho}}\right)^{\mu} b_{\beta} \tilde{u}_M(k_1) \tilde{u}_M(k_2) \dots \tilde{u}_M(k_{\mu}) \, . \tag{1.106}$$

Here M is halo mass, $n(M, z)$ is the number density of halos of mass M at redshift z, $\tilde{u}_M(k)$ is the Fourier transform of the halo density profile, μ is the number of points being correlated (two for the power spectrum, three for the bispectrum and so on), and b_{β} is the halo bias. $\mathrm{d}n/\mathrm{d}M$ is known as the halo mass function. Usually we assume linear bias so that $b_0 = 1$, $b_1 = b(M)$, and $b_{\beta} = 0$ for $\beta > 1$. The bias can then be normalised so that it satisfies

$$\int \mathrm{d}M \frac{\mathrm{d}n}{\mathrm{d}M} \left(\frac{M}{\bar{\rho}}\right) b(M) = 1 \, . \tag{1.107}$$

The components of the halo model have been modelled in a variety of ways to suit different purposes:

- *Average halo mass* This is generally defined as the mass of a sphere at a fixed radius, often the virial radius r_{vir}.

- *Halo mass function* The halo mass function describes the number density of halos per unit mass. Most formulations build on the seminal paper by Press and Schechter (1974) who considered the number density of fluctuations which were dense enough to collapse at a given time, in other words those above a defined cut-off density. This can be understood from Figure 1.9 which exemplifies the so-called peak–background split in which the density field is regarded as the sum of a long-wavelength, low amplitude mode and a high amplitude, short-wavelength, noisy component (Kaiser 1984; Bardeen et al. 1985; Mo and White 1996). Halos are most likely to form at the peaks above the horizontal line in Figure 1.9. The Press–Schechter model defines the cutoff for halo formation through the peak height $\nu = \delta_{\mathrm{c}}^2(z)/\sigma^2(M)$ where $\delta_{\mathrm{c}}(z)$ is the critical overdensity for collapse at redshift z and $\sigma^2(M)$ is the variance of the initial density fluctuations smoothed by a suitable top-hat filter. The value of the critical density has conventionally been estimated from the ‘top-hat

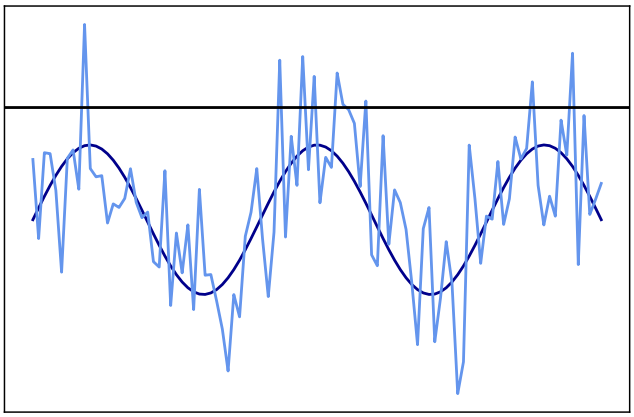

Figure 1.9: The concept of peak–background split. In this view, background long-wavelength modes of the density field are overlaid by noisy short-wavelength modes. Halos are most likely to form at peaks which are above a defined threshold (represented by the horizontal line) which depends on the critical density.

spherical collapse' model which considers the collapse of an idealised spherical overdensity. For an Einstein–de Sitter Universe $\delta_{\rm c}(z) \approx 1.686$. This value is only weakly dependent on cosmology and so is often adopted.

The mass function is then

$$\nu f(\nu)\frac{{\rm d}\nu}{\nu} = \sqrt{\frac{\nu}{2\pi}}{\rm e}^{-\nu/2}\frac{{\rm d}\nu}{\nu} \,. \tag{1.108}$$

This relatively simple formula was subsequently refined by Sheth and Tormen (1999) who used a more sophisticated definition of the peak–background split, and introduced additional free parameters to obtain a better fit with simulations. More recently Tinker et al. (2008) proposed:

$$\frac{{\rm d}n}{{\rm d}M} = f(\sigma)\frac{\bar{\rho}_{\rm m}}{M}\frac{{\rm d}\ln\sigma^{-1}}{{\rm d}M} \,, \tag{1.109}$$

where $f(\sigma)$ has the parameterised form

$$f(\sigma) = A\left[\left(\frac{\sigma}{b}\right)^{-a} + 1\right]{\rm e}^{-c/\sigma^2} \,, \tag{1.110}$$

$$\sigma = \int P(k)\tilde{W}(kR)k^2{\rm d}k \,, \tag{1.111}$$

and $\tilde{W}$ is the Fourier transform of a top-hat window function of radius R.

- *Halo density profile* The profile of the halo density is typically taken to be the Navarro–Frenk–White (NFW) profile which was first described in the very influential paper by Navarro et al. (1997) which first studied halo profiles in simulations. The NFW profile is

$$\rho(r) = \frac{\rho_0}{\frac{r}{r_s}\left(1+\frac{r}{r_s}\right)^2} \,, \tag{1.112}$$

  where ρ_0 is a characteristic density and r_s is a characteristic scale radius. The NFW profile has been found to be almost universal for different halo masses, redshifts and cosmologies.

- *Halo concentration* The NFW profile can alternatively be expressed in terms of the halo mass and the concentration which is defined as $c = r_\Delta / r_s$ where r_Δ is some defined radius, often taken to be $r_{\rm vir}$. Since the seminal early work it has been found that the relationships between mass, concentration, redshift and cosmology are more complex and less universal than the original models suggested. For example it appears that concentration is a decreasing function of halo mass, but also the mass-concentration relationship becomes shallower at higher redshift (Okoli 2017). A comprehensive summary of approaches to modelling concentration is given in the introduction to Diemer and Joyce (2019) who identify two prevalent categories of model: those linked to the assembly history of the halo, and those which model the mass–concentration relationship for a given redshift and cosmology as a power law or other simple function. The original work by Navarro et al. (1996) falls into the former category. An example of the fitting function approach is given in Duffy et al. (2008) who suggest

$$c = A\left(\frac{M}{M_{\rm pivot}}\right)^B (1+z)^C \,, \tag{1.113}$$

  where A, B and C are free parameters and $M_{\rm pivot}$ is the median halo mass. The two categories of model both have advantages, but neither approach has been entirely successful in explaining concentration at all masses and redshifts, and this continues to be an active area of research.

Clustering statistics are obtained by summing all possible configurations of points in different halos which contribute to the statistic. The power spectrum correlates two points which may both be in the same halo or in two different halos. Thus two terms contribute, a one-halo term and a two-halo term:

$$P(k) = P^{\rm 1h} + P^{\rm 2h} \,. \tag{1.114}$$

The integral for the first term involves only one halo but two points ($\mu = 2$) related by the matter distribution within the halo. Bias is not relevant, so

$$P^{1\mathrm{h}} = I_2^0(k, k) \; . \tag{1.115}$$

In contrast the second term of Equation 1.114 involves two separate halos, each containing one point. Bias needs to be taken into account and the points (halos) are related through the linear power spectrum. So this part of the power spectrum is

$$P^{2\mathrm{h}} = [I_1^1(k)]^2 P_{\mathrm{L}}(k) \; . \tag{1.116}$$

Similarly the bispectrum is the sum of terms involving all possible configurations of three points: three points in one halo, two points in one halo and one in another, and three points in separate halos. A pair of halos is related through the linear power spectrum P_{L} and three separate halos are related through B_{PT}, the bispectrum given by tree-level perturbation theory (Bernardeau et al. 2002).

$$\begin{aligned}
B(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) &= B^{1\mathrm{h}} + B^{2\mathrm{h}} + B^{3\mathrm{h}} \\
B^{1\mathrm{h}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) &= I_3^0(k_1, k_2, k_3) \\
B^{2\mathrm{h}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) &= I_1^1(k_1) I_2^1(k_2, k_3) P_{\mathrm{L}}(k_1) + \text{ 2 perms.} \\
B^{3\mathrm{h}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) &= I_1^1(k_1) I_1^1(k_2) I_1^1(k_3) B_{\mathrm{PT}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) \; .
\end{aligned} \tag{1.117}$$

The trispectrum has contributions from four points in a single halo, two points in each of two halos and so on (Cooray 2001; Cooray and Sheth 2002):

$$\begin{aligned}
T(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) &= T^{1\mathrm{h}} + T_{22}^{2\mathrm{h}} + T_{13}^{2\mathrm{h}} + T^{3\mathrm{h}} + T^{4\mathrm{h}} \\
T^{1\mathrm{h}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) &= I_4^0(k_1, k_2, k_3, k_4) \\
T_{22}^{2\mathrm{h}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) &= I_2^1(k_1, k_2) I_2^1(k_3, k_4) P_{\mathrm{L}}(k_{12}) + \text{ 2 perms.} \\
T_{13}^{2\mathrm{h}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) &= I_1^1(k_1) I_3^1(k_2, k_3, k_4) P_{\mathrm{L}}(k_1) + \text{ 3 perms.} \\
T^{3\mathrm{h}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) &= I_1^1(k_1) I_1^1(k_2) I_2^1(k_3, k_4) B_{\mathrm{PT}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_{34}) + \text{ 5 perms.} \\
T^{4\mathrm{h}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) &= I_1^1(k_1) I_1^1(k_2) I_1^1(k_3) I_1^1(k_4) T_{\mathrm{PT}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) \; .
\end{aligned} \tag{1.118}$$

where T_{PT} is the tree-level perturbation theory trispectrum.

1.3.3 Bayesian inference and the importance of covariance matrices

Much of current cosmological research involves making inferences from data. This data may come from surveys, simulations or other analytical calculations and is almost inevitably subject to statistical and systematic uncertainties, and (for surveys) is limited by cosmic variance: the fact that we have only one Universe to sample from.

Common reasons for making statistical inferences in cosmology are

- estimating constraints on cosmological parameters from survey data, in other words determining the best fit values of the parameters and their errors, given the data and a cosmological model;
- choosing between alternative models;
- forecasting the precision with which a given data set (for example from a future survey) can estimate parameters, and determining the best statistics to use for this.

Modern cosmology generally uses a Bayesian framework for these tasks. In this framework probability represents degree of belief in a proposition, in contrast to the frequentist concept of probability as the long-run expected occurrence of an event. Central to Bayesian inference is Bayes' theorem, often attributed to Bayes (1763), but in fact first explicitly derived by Laplace (1812). This states

$$P(\boldsymbol{\theta}|\mathbf{d}) = \frac{P(\mathbf{d}|\boldsymbol{\theta})P(\boldsymbol{\theta})}{P(\mathbf{d})} \,, \tag{1.119}$$

where $\mathbf{d} = (d_1, d_2, \cdots, d_n)$ is the data vector and $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_m)$ is the vector of parameters. $P(\boldsymbol{\theta}|\mathbf{d})$ is the posterior probability – the probability of the parameters given the data – which is what we want to know. $P(\mathbf{d}|\boldsymbol{\theta})$ is the likelihood, the probability distribution of the observed data given the parameters, and is often written as $L(\boldsymbol{\theta})$ with the dependence on the data understood. $P(\boldsymbol{\theta})$ is the prior probability which encapsulates what we already know or believe about the parameters (for example upper and lower limits). Priors are often assumed to be constant or 'flat' so that the posterior is proportional to the likelihood. The denominator of Equation 1.119, $P(\mathbf{d})$, is called the evidence. This is important for model selection but, since it is independent of the parameters, for parameter estimation the evidence can be considered simply as a normalising factor. In what follows we discuss only parameter estimation, mainly in the context of weak lensing.

To calculate the posterior we need to define and compute the likelihood. Often we can safely assume that the data has a Gaussian distribution which simplifies the likelihood. However the

assumption of Gaussianity should not be made without thought. This is discussed further in Chapter 4.

As the simplest example of a Gaussian likelihood, consider a model which is Gaussian with two parameters, the mean μ and standard deviation σ, so that $\boldsymbol{\theta} = (\mu, \sigma)$. The likelihood is then

$$L(\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{(\mathbf{d} - \mu)^2}{2\sigma^2} \right]. \tag{1.120}$$

We can generalise this to a model with a multivariate Gaussian distribution, n data values and m parameters:

$$L(\boldsymbol{\theta}) = \frac{1}{\sqrt{(2\pi)^n \det(\mathbf{C})}} \exp\left[-\frac{1}{2}(\mathbf{d} - \boldsymbol{\mu})^{\mathrm{T}} \mathbf{C}^{-1} (\mathbf{d} - \boldsymbol{\mu}) \right], \tag{1.121}$$

where $\mathbf{C}$ is the data covariance matrix.

The covariance matrix is thus crucial to calculation of the likelihood and posterior. Computing the covariance is not always straightforward and errors in the matrix can propagate into errors in parameter estimation or model selection (Taylor et al. 2013).

Several methods of estimating the cosmic shear covariance matrix have been suggested and used. These can be broadly classified into analytical models, estimation from the data itself, or estimation from mock catalogues based on simulations. In their recent analysis of KiDS data Hildebrandt et al. (2016) compared all three approaches. They concluded that in this case an analytical method based on the halo model was most suitable, being reliable, quick, noise-free and accurate at all scales. In other recent work Hikage et al. (2019) used a similar analytical model for their analysis of lensing data from the Hyper Suprime-Cam Subaru Strategic Program[11]. In contrast previous studies, such as the analysis of data from the Canada-France Hawaii Telescope Lensing Survey[12] by Kilbinger et al. (2013), used ray-tracing simulations (with some subtleties). The main disadvantage of this is the cost of running many simulations.

Another possibility is to compress the data vector in order to reduce the size of the covariance matrix. This allows the covariance to be estimated with fewer analytical calculations or simulations. Data compression may be particularly desirable for three-point and higher-order statistics where the dimension of the covariance matrix is large. A classic data compression method is the Karhunen-Loève algorithm which is discussed in Tegmark et al. (1997). Alternative approaches using proxy

[11] hsc.mtk.nao.ac.jp
[12] http://www.cfhtlens.org

measures to summarise the bispectrum (Byun et al. 2017) and cosmic shear power spectrum (Asgari and Schneider 2015) have also been proposed.

Of course, reducing the covariance matrix is not justifiable if this throws away a lot of information, but it has been shown that this is not necessarily the case. A practical demonstration of the potential for data compression is given by Gualdi et al. (2018) who considered the redshift space galaxy power spectrum and bispectrum. Their analysis showed that the dimension of the data vector could be reduced from around 1000 down to the number of cosmological parameters with negligible loss of information.

Evaluating the likelihood

Once the covariance matrix has been estimated the next step is to find the peak (maximum likelihood) and spread of the posterior distribution. This may not be straightforward if there are many parameters or the likelihood has a complicated shape. Several methods have been developed to sample the likelihood efficiently and rapidly determine the maximum value and variance. Here we restrict discussion to two methods which are commonly used in cosmology: the very simple Metropolis-Hastings algorithm (Metropolis et al. 1953) which is a Monte Carlo Markov Chain (MCMC) method, and nested sampling (Skilling et al. 2006; Mukherjee et al. 2006; Feroz and Hobson 2008) which is more suitable for complex multi-modal likelihoods or if the Bayesian evidence is required as well as the posterior.

The Metropolis-Hastings algorithm

The aim of this algorithm is to generate chains of points which sample parameter space and converge on the maximum posterior and its surrounding credible regions (Heavens 2009).

The algorithm relies on two pre-defined functions

- a target density $f(\boldsymbol{\theta})$ which is proportional to the likelihood at any point;
- a proposal distribution $g(\boldsymbol{\theta})$ which is used to generate trial points (which may or may not be accepted into the chain).

A random starting point $\boldsymbol{\theta}$ in parameter space is chosen and the target density is calculated at this point. A step in parameter space is then generated by sampling from the proposal distribution $g(\boldsymbol{\theta})$, producing a new candidate point $\boldsymbol{\theta}'$. This point is accepted or rejected with a probability p

which depends on the ratio of the target densities at the last and proposed points.

$$p(\text{acceptance}) = \min\left(1, \frac{f(\boldsymbol{\theta}')g(\boldsymbol{\theta}'|\boldsymbol{\theta})}{f(\boldsymbol{\theta})g(\boldsymbol{\theta}|\boldsymbol{\theta}')}\right) . \tag{1.122}$$

The sequence of proposing and selecting steps is repeated until the chain has converged to a stationary distribution. The proposal distribution is crucial. If the typical steps which it generates are small the algorithm will be slow to explore the whole parameter space. If steps are too large then the chain may have difficulty homing in on regions where the likelihood is high. It is important to test that the chain has indeed converged, and not rely on the appearance of convergence. This is most commonly done with the Gelman–Rubin convergence criterion (Gelman et al. 1992) which compares the between-chain variance of several chains with the within-chain variance. In practice the Metropolis-Hastings method is quite inefficient and is only suitable for very basic problems.

Nested sampling

An alternative approach is nested sampling which is mainly designed to estimate the Bayesian evidence but as a by-product also estimates the posterior (Skilling et al. 2006). This algorithm is more computationally efficient than the Metropolis-Hastings method and has been tuned to deal effectively with complex multi-modal posteriors (Feroz and Hobson 2008).

The algorithm starts from the definition of the Bayesian evidence, Z,

$$Z = \int L(\boldsymbol{\theta})\pi(\boldsymbol{\theta})\, \mathrm{d}^m\boldsymbol{\theta} , \tag{1.123}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_m)$ is the vector of parameters, $L(\boldsymbol{\theta})$ is the likelihood and $\pi(\boldsymbol{\theta})$ is the prior. The integral extends over the entire volume defined by the prior.

Next the normalised prior volume X is defined as

$$X = \int \pi(\boldsymbol{\theta})\, \mathrm{d}^m\boldsymbol{\theta} \equiv 1 . \tag{1.124}$$

More specifically the algorithm uses the prior volume within a likelihood contour λ, in other words the volume in parameter space which encompasses all points at which the likelihood is greater than λ:

$$X(\lambda) = \int_{L(\boldsymbol{\theta})>\lambda} \pi(\boldsymbol{\theta})\, \mathrm{d}^m\boldsymbol{\theta} . \tag{1.125}$$

By a change of variables Equation 1.123 can be transformed to

$$Z = \int_0^1 L(X)\mathrm{d}X \ , \tag{1.126}$$

where $L(X(\lambda)) \equiv \lambda$. Thus the multi-dimensional integral is changed to a one-dimensional one.

The nested sampling algorithm starts by sampling n points within an initial volume X_0 defined by the priors and an initial (arbitrary) likelihood limit λ_0. The likelihoods at the n points are evaluated and the point with the lowest likelihood, L_0, is identified. The position of this point in parameter space and its associated likelihood are recorded and it is then removed from the set of sample points. The remaining points all have likelihoods greater than L_0. MCMC or another suitable method is used to generate a new trial point which is accepted into the sample if its likelihood is greater than L_0. Otherwise it is rejected and further trial points are generated until an acceptable one is found. The point in the updated sample with the smallest likelihood, L_1, is then identified and removed from the sample as before. The 'live' prior volume, $X_1 = X(\lambda_1)$ now consists of the volume where the likelihood is greater than L_1. The process is continued, at each stage drawing from a smaller volume, until the evidence has been determined to a pre-specified precision. The set of discarded points maps out the full likelihood. The prior volumes X_i necessarily decrease monotonically while the likelihoods L_i increase.

Figure 1.10 shows the method schematically. The left panel shows a two-dimensional problem, with contours demarcating increasing likelihoods and decreasing prior volumes. The right panel plots the relationship between the likelihood and the prior volume.

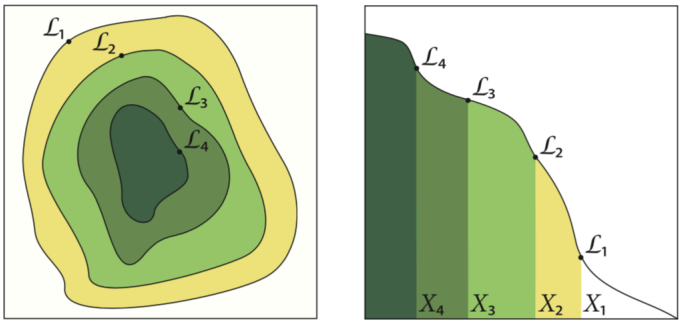

Figure 1.10: *Left*: Decreasing likelihood contours. *Right*: Prior volumes X_i associated with likelihoods L_i. From Figure 1 in Feroz et al. (2013).

1.3.4 Fisher matrix methods

Another common question is how well future experiments will be able to constrain cosmological parameters. This might then inform survey design or suggest whether novel estimators will provide additional information and thus reduce the error bars of the parameter estimates. The way to forecast how informative an experiment will be was first worked out by Fisher (1935). Clear derivations relevant to cosmology are given in Tegmark et al. (1997) and Heavens (2009).

Assume that $\mathbf{d} = (d_1, d_2, \cdots, d_n)$ is a set of future data which we plan to collect, for example cosmic shear power spectra. We want to know how well this data can constrain the parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_m)$.

We use the data to produce estimates $\boldsymbol{\theta}$. For these to be 'good' estimates we would like them to be unbiased so that

$$\langle \boldsymbol{\theta} \rangle = \boldsymbol{\theta}_0 \ , \tag{1.127}$$

where $\boldsymbol{\theta}_0$ is the true (unknown) parameter vector.

We would also like the error bars on the estimates to be as small as possible. Thus we need to minimise the standard deviations of the individual parameter estimates, $(\langle \langle (\theta_i^2 \rangle - \langle \theta_i \rangle^2)^{1/2}$. The smaller the standard deviations, the more certain we are about the estimated parameter values, or equivalently the more information the data provides about the parameters. The unbiased estimate which has the minimum variance is (unimaginatively) called the Best Unbiased Estimator (BUE).

Another way of looking at this is that if the likelihood is sharply peaked around the maximum then the data carries a lot of information about the parameters. If the likelihood is flat around the maximum then the data is uninformative – we are uncertain what the true parameter values are. So we need a formalism to explore the shape of the likelihood in parameter space.

We first define the negative log likelihood

$$\mathcal{L}(\boldsymbol{\theta}) \equiv -\ln L(\boldsymbol{\theta}) \ . \tag{1.128}$$

Then the Fisher information matrix is defined by

$$F_{ij} \equiv \left\langle \frac{\partial^2 \mathcal{L}}{\partial \theta_i \partial \theta_j} \right\rangle \Bigg|_{\boldsymbol{\theta}_{\rm ML}} \ , \tag{1.129}$$

where the expectation value is over all possible realisations of the data evaluated at $\boldsymbol{\theta}_{\rm ML}$, the parameter vector which maximises the likelihood. The Fisher matrix describes the way the likelihood varies with the parameters and is key to understanding how well the data constrains the parameters. It has several very useful properties (Kenney and Keeping 1951; Kendall and Stuart 1966):

- For any unbiased estimator θ_i of an individual parameter the standard deviation of the estimate is greater than $\mathbf{F}_{ii}^{-1/2}$. This is the Cramér–Rao inequality. It places a lower limit on the error on θ_i if all the other parameters are known. If the other parameters are also estimated from the data then the minimum standard deviation is $(\mathbf{F}^{-1})_{ii}^{1/2}$.

- If an unbiased estimator exists which attains the Cramér–Rao bound, then it is equal to $\boldsymbol{\theta}_{\rm ML}$.

- As the size of the data vector increases, $\boldsymbol{\theta}_{\rm ML}$ is asymptotically the BUE with the lowest variance – which is what we want to find.

Thus the Fisher matrix encodes the amount of information which the data can provide about the parameters. Since the Fisher matrix depends on the covariance, this once again underlines the importance of having an accurate covariance matrix.

To derive the Fisher matrix for multivariate Gaussian data with mean $\boldsymbol{\mu}$ and standard deviation $\mathbf{C}$, we start from Equations 1.121 and 1.128. Using the identity $\ln\det\mathbf{C} = \mathrm{Tr}\ln\mathbf{C}$, where Tr denotes the trace of the matrix, and defining the data matrix $\mathbf{D} \equiv (\mathbf{d}-\boldsymbol{\mu})(\mathbf{d}-\boldsymbol{\mu})^{\rm T}$ leads to

$$2\mathcal{L} = \mathrm{Tr}[\ln\mathbf{C} + \mathbf{C}^{-1}\mathbf{D}]\,, \tag{1.130}$$

ignoring the additive normalisation constant. Making use of the identity $(\mathbf{C}^{-1})_{,i} = -\mathbf{C}^{-1}\mathbf{C}_{,i}\mathbf{C}^{-1}$, the first derivative with respect to the parameter θ_i is

$$2\mathcal{L}_{,i} = \mathrm{Tr}[\mathbf{C}^{-1}\mathbf{C}_{,i} - \mathbf{C}^{-1}\mathbf{C}_{,i}\mathbf{C}^{-1}\mathbf{D} + \mathbf{C}^{-1}\mathbf{D}_{,i}]\,, \tag{1.131}$$

and differentiating again gives

$$\begin{aligned} 2\mathcal{L}_{,ij} = \mathrm{Tr}[&-\mathbf{C}^{-1}\mathbf{C}_{,j}\mathbf{C}^{-1}\mathbf{C}_{,i} + \mathbf{C}^{-1}\mathbf{C}_{,ij} + \mathbf{C}^{-1}\mathbf{C}_{,j}\mathbf{C}^{-1}\mathbf{C}_{,i}\mathbf{C}^{-1}\mathbf{D} \\ &- \mathbf{C}^{-1}\mathbf{C}_{,ij}\mathbf{C}^{-1}\mathbf{D} + \mathbf{C}^{-1}\mathbf{C}_{,i}\mathbf{C}^{-1}\mathbf{C}_{,j}\mathbf{C}^{-1}\mathbf{D} \\ &- \mathbf{C}^{-1}\mathbf{C}_{,i}\mathbf{C}^{-1}\mathbf{D}_{,j} - \mathbf{C}^{-1}\mathbf{C}_{,j}\mathbf{C}^{-1}\mathbf{D}_{,i} + \mathbf{C}^{-1}\mathbf{D}_{,ij}]\,. \end{aligned} \tag{1.132}$$

The Fisher matrix is the expectation value of $\mathcal{L}_{,ij}$ at the true parameter values. At these values we have

$$\langle \mathbf{D} \rangle = \mathbf{C} \,, \tag{1.133}$$

$$\langle \mathbf{D}_{,i} \rangle = 0 \,, \tag{1.134}$$

$$\langle \mathbf{D}_{,ij} \rangle = \boldsymbol{\mu}_{,i} \boldsymbol{\mu}_{,j}^{\mathrm{T}} + \boldsymbol{\mu}_{,j} \boldsymbol{\mu}_{,i}^{\mathrm{T}} \,. \tag{1.135}$$

Thus when we take the expectation value some terms are zero and others cancel, leaving

$$\begin{aligned} F_{ij} &\equiv \langle \mathcal{L}_{,ij} \rangle \\ &= \frac{1}{2} \mathrm{Tr}[\mathbf{C}^{-1} \mathbf{C}_{,i} \mathbf{C}^{-1} \mathbf{C}_{,j} + \mathbf{C}^{-1} \langle \mathbf{D}_{,ij} \rangle] \end{aligned} \tag{1.136}$$

The two terms depend respectively on how the covariance and the mean vary with the parameters. Often the first term is neglected on the grounds that the covariance matrix depends only weakly on cosmology. This assumption is explored in Chapter 4.

The Fisher matrix can be used to plot confidence ellipses for any pair of parameters and thus to visualise forecast errors (Coe 2009). The inverse Fisher matrix is the covariance matrix between the parameters. If we consider two parameters θ_1 and θ_2 with uncertainties σ_1 and σ_2,

$$\mathbf{F}^{-1} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \,,$$

where ρ is the correlation between the parameters. The ellipse representing a credible region has axes a and b satisfying

$$a^2 = \alpha \left(\frac{\sigma_1^2 + \sigma_2^2}{2} + \sqrt{\frac{(\sigma_1^2 - \sigma_2^2)^2}{4} + \rho^2 \sigma_1^2 \sigma_2^2} \right) \tag{1.137}$$

$$b^2 = \alpha \left(\frac{\sigma_1^2 + \sigma_2^2}{2} - \sqrt{\frac{(\sigma_1^2 - \sigma_2^2)^2}{4} + \rho^2 \sigma_1^2 \sigma_2^2} \right) \,, \tag{1.138}$$

where α is a coefficient which depends on the confidence level being plotted (for example $\alpha \approx 2.48$ for a 95% credible region). The ellipse is at an angle ϕ to the θ_1 axis, with

$$\tan\, 2\phi = \frac{\rho\theta_1\theta_2}{\theta_1^2 - \theta_2^2} \,. \tag{1.139}$$

Removing the row and column of the Fisher matrix corresponding to a particular parameter is equivalent to treating this parameter as fixed and certain. Alternatively a prior can be placed on parameter θ, say $\Delta\theta = \sigma$, by adding $1/\sigma^2$ to the on-diagonal element corresponding to θ (Coe 2009).

Fisher matrices can also be used to assess the additional constraining power of a second independent experiment. To do this we simply add Fisher matrices together and compare constraints with and without the second experiment.

A Fisher matrix example

Figure 1.11, from Sato and Nishimichi (2013), is an example which uses Fisher matrix methods to explore the impact of non-Gaussian terms of the weak lensing power spectrum covariance. The weak lensing covariance matrix is not diagonal because of mode-coupling caused by non-linear structure formation. In fact the covariance can be written as the sum of three terms: a Gaussian term (which may include shape/shot noise), a non-Gaussian term which arises from mode-coupling within the survey area, and a so-called supersample term which is due to coupling between in-survey modes and long-wavelength modes larger than the survey area. These terms are derived in Takada and Hu (2013) for the matter power spectrum, and extended to weak lensing in Sato and Nishimichi (2013) (see also Barreira et al. (2018*b*)). Chapter 3 discusses these terms in more detail. Figure 1.11 uses Fisher matrix analysis to show that in the absence of shape noise the non-Gaussian terms from the weak lensing power spectrum covariance are crucial for constraining cosmological parameters.

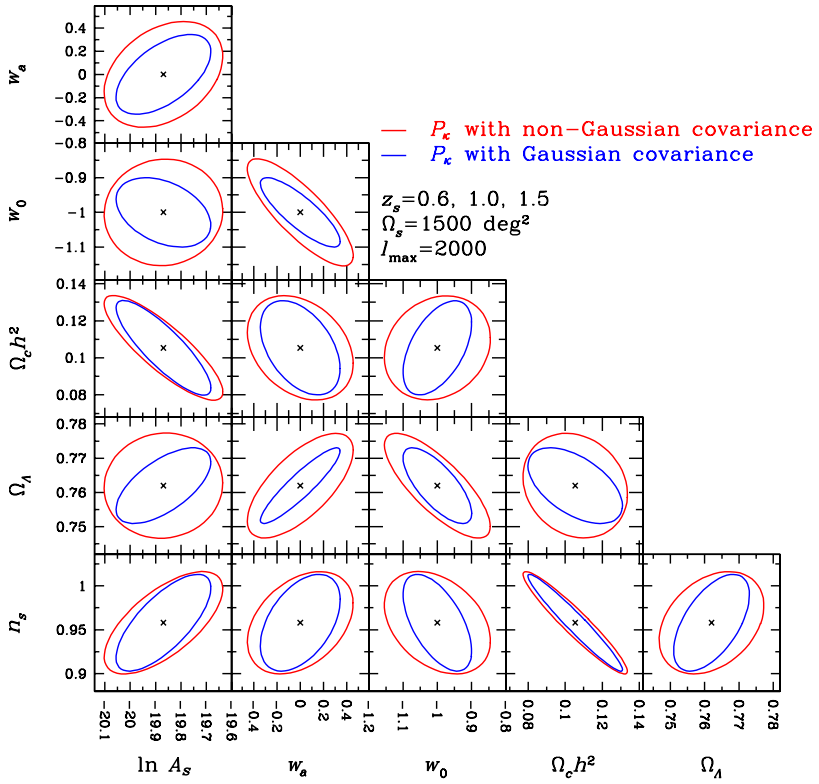

Figure 1.11: Illustrative Fisher matrix analysis of the weak lensing power spectrum covariance. The ellipses show forecast constraints using only the Gaussian term of the weak lensing power spectrum covariance (blue) and using Gaussian and non-Gaussian terms (red). Excluding the non-Gaussian terms would under-estimate errors. From Figure 9 in Sato and Nishimichi (2013).

2 Weak lensing deflection of three-point correlation functions

2.1 Introduction

On scales larger than galaxies the structure of the Universe depends only on properties of primordial inhomogeneities and their subsequent evolution under gravity. The resulting matter distribution is commonly quantified through clustering statistics such as the two-point correlation function (2PCF) or, in Fourier space, the power spectrum. The primordial density distribution is determined by inflation and is expected to be nearly Gaussian and therefore fully described by these two-point statistics. However, later gravitational collapse of overdense regions causes coupling between different Fourier modes and makes the matter distribution non-Gaussian. As a result information is transferred into higher-order statistics, which are complementary to two-point statistics. For example, the three-point correlation function (3PCF) or bispectrum has been used to investigate primordial non-Gaussianity (Tasinato et al. 2014; Lazanu et al. 2016), to estimate galaxy bias (the differential clustering of galaxies compared with dark matter) (Sefusatti and Scoccimarro 2005), and to constrain cosmological parameters, in particular $\Omega_{\rm m}$ and σ_8 (Sefusatti et al. 2006). Since the 1970s two- and three-point correlation functions have been measured with increasing accuracy in galaxy surveys such as the Two-degree-Field Galaxy Redshift Survey[1] and the Sloan Digital Sky Survey (SDSS).[2] A recent aim has been the detection of baryon acoustic oscillations (BAOs) in clustering statistics. The position of BAO peaks acts as a standard ruler which can provide information about distances and hence about the expansion of the Universe. The first evidence for BAO peaks in the 3PCF was reported in 2009 by Gaztanaga et al. (2009). Subsequently evidence at the 2.8σ level was reported by Slepian et al. (2017*b*) using data from the SDSS Baryon Oscillation Spectroscopic Survey (BOSS), and more recently the first high confidence detection (4.5σ) was reported, also using SDSS data (Slepian et al. 2017*a*). Future surveys such as the Dark Energy Spectroscopic Instrument[3] (DESI) (Aghamousa et al. 2016), *Euclid*[4] (Laureijs et al. 2011) and the Rubin Observatory Legacy Survey of Space and Time[5] (LSST) (Abell et al. 2009) will improve upon these measurements, for example by halving the uncertainties in distance measurements at the

[1] http://www.2dfgrs.net/
[2] http://www.sdss.org/
[3] https://www.desi.lbl.gov
[4] http://sci.esa.int/euclid/
[5] https://www.lsst.org/about

BAO scale (Aghamousa et al. 2016). Consequently there will be a need for increasingly accurate modelling of the 2PCF and 3PCF.

Theoretical expressions for the matter 2PCF and 3PCF up to second order in the density contrast can be derived using Newtonian perturbation theory (Bernardeau et al. 2002) and are sufficient for comparison with current surveys. However other contributions to the correlation functions may also be important for future surveys. For example in a detailed analysis of the effect of weak lensing magnification on the galaxy 3PCF Schmidt et al. (2008) concluded that this effect was potentially detectable in galaxy and quasar samples at $z = 3$ in future 'ideal' surveys. More recent work has emphasised that future wide and deep surveys will require the inclusion of additional relativistic terms in the power spectrum and bispectrum of galaxy number counts (Bonvin and Durrer 2011; Bertacca et al. 2014; Yoo and Zaldarriaga 2014; Di Dio et al. 2016; Tansella et al. 2018; Bertacca et al. 2018). These terms arise because at higher redshifts the quantities which we observe, such as positions, volumes and densities, differ from their true source values due to the propagation of light through the inhomogeneous matter distribution along the line of sight. Some effects which are first order in the density and volume in perturbation theory are detectable in current surveys, for example redshift-space distortions, weak lensing and the integrated Sachs-Wolfe effect. Many second order effects have also been estimated but have not been shown to be detectable in current or planned surveys (Di Dio et al. 2016).

One second order relativistic effect due to lensing is the alteration of the apparent distance between two or more patches of the sky so that sources are not observed at their true positions (Bonvin and Durrer 2011). For galaxy samples this is a much smaller effect than the more commonly studied weak lensing magnification and shear. However it is potentially larger than other relativistic contributions and its implications for two-point and higher-order statistics could be relevant for future surveys. In this chapter we refer to this effect as lensing deflection; it should not be confused with the deflection angle given by Equation 1.57.

Early studies of this deflection due to weak lensing mainly considered the effect on cosmic microwave background (CMB) anisotropies (Blanchard and Schneider 1987; Seljak 1996). Since then the effect of lensing deflection on the power spectrum of CMB temperature anisotropies has been explored in detail in harmonic space (Bohm et al. 2016; Hanson et al. 2010; Hu 2000; Lewis and Challinor 2006), demonstrating that lensing is a non-trivial contaminant of CMB temperature and polarisation observations but also introduces valuable additional information. In a wider context Dodelson et al. (2008) estimated the impact of deflection on the matter 2PCF, assuming two sources at the same redshift. They concluded that the effect is small except where the correlation function

is rapidly changing, for example near the BAO feature in galaxy surveys (Vallinotto et al. 2007), where they estimated the effect to be around the percent level. Here lensing deflection tends to wash out the details of the peaks and troughs. Importantly, this deflection does not affect the position of the BAO peaks although it sets a limit on the accuracy with which the amplitudes of the peaks can be measured (Vallinotto et al. 2007). Thus it has no implications for the use of the BAO scale as a standard ruler. More recently Di Dio (2017) repeated this analysis in Fourier space and confirmed the size of the smoothing effect on the matter power spectrum at the BAO scale, also showing that at $z = 4$ this effect has approximately the same magnitude as smoothing due to non-linear structure. Even if the deflection effect is not important for two-point statistics, it could potentially make a measurable, and interesting, contribution to three-point statistics. In harmonic space expressions have been developed for the bispectrum induced by CMB lensing (Hu 2000; Lewis and Challinor 2006). In this work we instead extend the general real-space analysis of Dodelson et al. (2008) to the 3PCF, now with three sources at the same redshift, and consider whether the effect could be detected in forthcoming surveys.

This chapter is organised as follows: Section 2.2 derives general expressions for the lensed 3PCF and the associated lensing deflection; Section 2.3 applies our new derivations to the 3PCF of the matter density field; Section 2.4 discusses the observability of the deflection effect; Section 2.5 contains our conclusions. Detailed derivations are given in appendices. Throughout we assume a flat ΛCDM universe.

2.2 Lensed three-point correlation function

Suppose that a physical observable $A(\mathbf{x}_a)$ is observed at position $\mathbf{x}_a$. The true position is not as observed because photons are deflected as they travel to the observer. Thus the (lensed) quantity $\tilde{A}(\mathbf{x}_a)$ which is observed at $\mathbf{x}_a$ is actually at a different position $\mathbf{x}_a + \boldsymbol{\lambda}_a$, where $\boldsymbol{\lambda}_a$ is a deflection vector:

$$\tilde{A}(\mathbf{x}_a) = A(\mathbf{x}_a + \boldsymbol{\lambda}_a) \,. \tag{2.1}$$

If we measure a correlation function of the observable we necessarily measure the correlation between lensed variables, which is not the same as the true correlation function. Taking the 2PCF

as an example,

$$\langle \tilde{A}(\mathbf{x}_a)\tilde{B}(\mathbf{x}_a)\rangle = \langle A(\mathbf{x}_a + \boldsymbol{\lambda}_a)B(\mathbf{x}_b + \boldsymbol{\lambda}_b)\rangle \tag{2.2}$$

$$\neq \langle A(\mathbf{x}_a)B(\mathbf{x}_a)\rangle \, . \tag{2.3}$$

Our approach to deriving an expression for the lensed 3PCF is motivated by results showing that in the two-point case the lensed 2PCF can be expressed as the sum of the unlensed 2PCF and a lensing displacement term, $\langle AB\rangle_2$ (Dodelson et al. 2008). (In the remainder of this chapter we refer to this as lensing deflection, but this should not be confused with the deflection angle $\hat{\boldsymbol{\alpha}}$ given by Equation 1.57). Importantly, Dodelson et al. (2008) showed that the transverse deflection is much greater than the deflection along the line of sight (which they describe as a lensing-induced time delay) so in this work we consider only transverse displacements.

Assuming that the comoving distances to A and B are approximately the same and that the deflection is small so a perturbative approach can be used, the lensed 2PCF is given by Dodelson et al. (2008) as

$$\langle \tilde{A}\tilde{B}\rangle = \langle AB\rangle + \langle AB\rangle_2 \tag{2.4}$$

$$\approx \langle AB\rangle + \frac{1}{r}\left(T - \frac{D}{2}\right)\frac{\mathrm{d}\langle AB\rangle}{\mathrm{d}r} + \left(T + \frac{D}{2}\right)\frac{\mathrm{d}^2\langle AB\rangle}{\mathrm{d}r^2} \, , \tag{2.5}$$

where r is the distance between A and B. The functions T and $D/2$ both depend on the distance between the background sources but T is similar along both transverse directions and $D/2$ differs along the two directions. They define a distortion tensor with components Z_{ij}, where i and j denote two orthogonal directions in the plane of the sky:

$$Z = \begin{pmatrix} T + \frac{D}{2} & 0 \\ 0 & T - \frac{D}{2} \end{pmatrix} . \tag{2.6}$$

The components of the deflection vector can be expressed in terms of integrals of the gravitational potential over the line of sight which arise as solutions of the geodesic equations (see for example Lewis and Challinor (2006) for a derivation in conformal Newtonian gauge). In a flat ΛCDM universe the two orthogonal transverse components of $\boldsymbol{\lambda}_a$ (denoted by the index $l = 1, 2$) are given by (Bartelmann and Schneider 2001)

$$\lambda^l_{a,\perp} = \frac{2}{c^2}\int_0^{\chi_a} \mathrm{d}\chi(\chi_a - \chi)\nabla^l\Phi(\chi) \, , \tag{2.7}$$

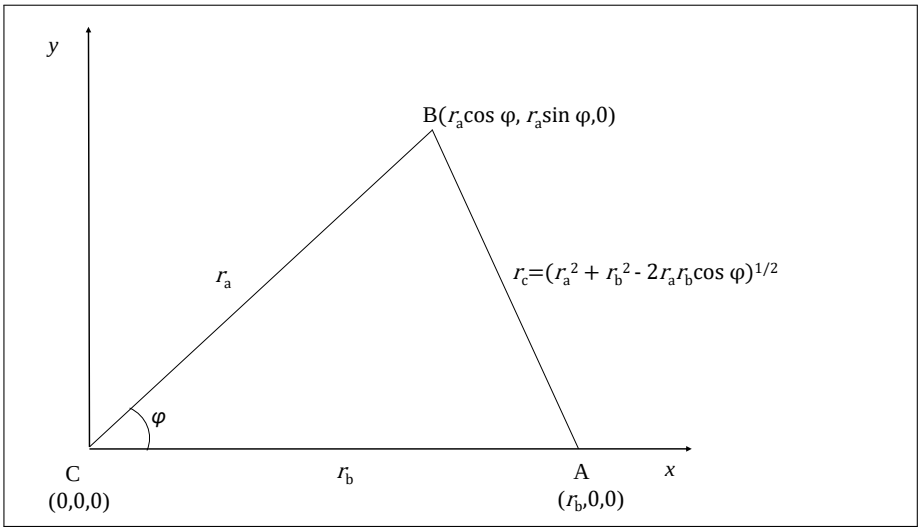

Figure 2.1: Choice of coordinates for the three points. Point A is at $(r_{b1}, r_{b2}, 0) = (r_b, 0, 0)$, point B is at $(r_{a1}, r_{a2}, 0) = (r_a \cos\varphi, r_a \sin\varphi, 0)$ and point C is at the origin.

where χ_a is the comoving distance to the source and $\Phi(\chi)$ is the gravitational potential.

To obtain an expression for the lensed 3PCF, firstly we assume that all three sources are at the same redshift. This is justified by the 2PCF finding that the dominant lensing effect is in the transverse direction, which means that triangles oriented closer to the line of sight will be less affected by lensing. Because of isotropy it is possible to choose coordinates with the x-axis along $\mathbf{x}_a - \mathbf{x}_c$, the y-axis along the line of sight to $\mathbf{x}_c$, and all three points in the $x - y$ plane. The triangle formed from the three points can then be defined in terms of two sides, $\mathbf{r}_a = (r_{a1}, r_{a2}, 0)$ and $\mathbf{r}_b = (r_{b1}, r_{b2}, 0)$, and the angle φ between them, as shown in Figure 2.1. Thus $\mathbf{r}_a$ is the observed distance between $\mathbf{x}_b$ and $\mathbf{x}_c$, and $\mathbf{r}_b$ is the observed distance between $\mathbf{x}_a$ and $\mathbf{x}_c$.

As explained in Appendix 2A, using Equation 2.1 we can write the lensed 3PCF in the same way as for the 2PCF as the sum of the unlensed correlation function and a lensing deflection term:

$$\langle \tilde{A}\tilde{B}\tilde{C} \rangle = \langle ABC \rangle + \langle ABC \rangle_2 \, . \tag{2.8}$$

Following Dodelson et al. (2008), we define three distortion tensors, Z_{ac}, Z_{ab} and Z_{bc}, each quantifying the deflection along one side of the triangle, with elements

$$Z^{ij}_{\alpha\beta} \equiv \frac{(\langle \lambda^i_\alpha \lambda^j_\alpha \rangle + \langle \lambda^i_\beta \lambda^j_\beta \rangle)}{2} - \langle \lambda^i_\alpha \lambda^j_\beta \rangle \, , \tag{2.9}$$

where $\alpha\beta$ is ab, bc or ca. We assume that the lensing deflection is small so that terms above second order can be neglected, and that the observables are not correlated with the lensing deflection field.

Then by Taylor-expanding the expression for the lensed correlation function, Equation 2.8, we can write the lensing deflection in terms of the deflection tensors. Appendix 2A gives fuller details. The final result is

$$\begin{aligned}\langle ABC\rangle_2 = &\frac{\partial^2\langle ABC\rangle}{\partial r_{bi}\partial r_{bj}} Z^{ij}_{ac} \\ &+ \frac{\partial^2\langle ABC\rangle}{\partial r_{ai}\partial r_{aj}} Z^{ij}_{bc} \\ &+ \frac{\partial^2\langle ABC\rangle}{\partial r_{ai}\partial r_{bj}} [Z^{ij}_{ac} + Z^{ij}_{bc} - Z^{ij}_{ab}] \,, \end{aligned} \tag{2.10}$$

where the indices i, j take the values 1 and 2, denoting the two transverse directions, and repeated indices are summed over. In the chosen coordinates $Z^{12}_{ab} = Z^{21}_{ab} = 0$. Thus we only need to consider the $i = j = 1$ and $i = j = 2$ terms and Equation 2.10 simplifies to

$$\begin{aligned}\langle ABC\rangle_2 = &\frac{\partial^2\langle ABC\rangle}{\partial r^2_{b1}} Z^{11}_{ac} + \frac{\partial^2\langle ABC\rangle}{\partial r^2_{a1}} Z^{11}_{bc} + \frac{\partial^2\langle ABC\rangle}{\partial r^2_{a2}} Z^{22}_{bc} \\ &+ \frac{\partial^2\langle ABC\rangle}{\partial r_{a1}\partial r_{b1}} [Z^{11}_{ac} + Z^{11}_{bc} - Z^{11}_{ab}] \,. \end{aligned} \tag{2.11}$$

As shown in Appendix 2A, this can be expressed in terms of derivatives of r_a, r_b and φ only:

$$\begin{aligned}\langle ABC\rangle_2 = &\frac{\partial^2\langle ABC\rangle}{\partial r^2_a} \left[Z^{11}_{bc}\cos^2\varphi + Z^{22}_{bc}\sin^2\varphi\right] \\ &+ \frac{\partial^2\langle ABC\rangle}{\partial r^2_b} Z^{11}_{ac} \\ &+ \frac{\partial^2\langle ABC\rangle}{\partial \varphi^2} \left[\frac{Z^{11}_{bc}\sin^2\varphi + Z^{22}_{bc}\cos^2\varphi}{r^2_a}\right] \\ &+ \frac{\partial^2\langle ABC\rangle}{\partial r_a \partial r_b} \cos\varphi [Z^{11}_{ac} + Z^{11}_{bc} - Z^{11}_{ab}] \\ &- \frac{\partial^2\langle ABC\rangle}{\partial r_a \partial \varphi} \frac{2\sin\varphi\cos\varphi}{r_a} \left[Z^{11}_{bc} - Z^{22}_{bc}\right] \\ &- \frac{\partial^2\langle ABC\rangle}{\partial r_b \partial \varphi} \frac{\sin\varphi}{r_a} [Z^{11}_{ac} + Z^{11}_{bc} - Z^{11}_{ab}] \\ &+ \frac{\partial\langle ABC\rangle}{\partial r_a} \left[\frac{Z^{11}_{bc}\sin^2\varphi + Z^{22}_{bc}\cos^2\varphi}{r_a}\right] \\ &+ \frac{\partial\langle ABC\rangle}{\partial \varphi} \frac{2\sin\varphi\cos\varphi}{r^2_a} \left[Z^{11}_{bc} - Z^{22}_{bc}\right] . \end{aligned} \tag{2.12}$$

To determine Z^{ij}_{ac}, Z^{ij}_{ab} and Z^{ij}_{bc} we follow the arguments of Dodelson et al. (2008) for the 2PCF. Using Equation 2.6 each tensor $Z_{\alpha\beta}$ can be written in terms of its trace $T'_{\alpha\beta}$ plus an off-diagonal traceless part $D'_{\alpha\beta}/2$. The functions T' and D' are related to the functions T and D in Equation 2.5.

From Equations A17 and A18 of Dodelson et al. (2008) we have

$$\langle \lambda^i_\alpha \lambda^j_\beta \rangle = T_{\alpha\beta}\delta_{ij} - \frac{D_{\alpha\beta}}{r^2}\left[r_i r_j - \frac{r^2}{2}\delta_{ij} \right] , \tag{2.13}$$

where δ_{ij} is the Kronecker delta. Now consider Z^{11}_{ac} as an example. From Equation 2.9 this is defined as

$$Z^{11}_{ac} = \frac{\langle \lambda^1_a \lambda^1_a \rangle + \langle \lambda^1_c \lambda^1_c \rangle}{2} - \langle \lambda^1_a \lambda^1_c \rangle . \tag{2.14}$$

Using Equation 2.13 and noting that $D_{aa} = D_{cc} = 0$ (Dodelson et al. 2008), this can be written as

$$Z^{11}_{ac} = \frac{1}{2}\left(T_{aa} + T_{cc}\right) - \left(T_{ac} - \frac{D'_{ac}}{r_b^2}\left(r^2_{b1} - \frac{r_b^2}{2} \right) \right) \tag{2.15}$$

$$= T'_{ac} + \frac{D'_{ac}}{2} . \tag{2.16}$$

The $i = j = 1$ and $i = j = 2$ elements of the other tensors can be derived in a similar way. They are:

$$\begin{aligned} Z^{11}_{bc} &= T'_{bc} + \frac{D'_{bc}}{r_a^2}\left[r^2_{a1} - \frac{r_a^2}{2} \right] \\ &= T'_{bc} + D'_{bc}\left[\cos^2\varphi - \frac{1}{2} \right] , \end{aligned} \tag{2.17}$$

$$\begin{aligned} Z^{22}_{bc} &= T'_{bc} + \frac{D'_{bc}}{r_a^2}\left[r^2_{a2} - \frac{r_a^2}{2} \right] \\ &= T'_{bc} + D'_{bc}\left[\sin^2\varphi - \frac{1}{2} \right] , \end{aligned} \tag{2.18}$$

$$\begin{aligned} Z^{11}_{ab} &= T'_{ab} + D'_{ab}\frac{[2r^2_{c1} - r_c^2]}{2r_c^2} \\ &= T'_{ab} + D'_{ab}\frac{[2(r_b - r_a\cos\varphi)^2 - (r_a^2 + r_b^2 - 2r_a r_b\cos\varphi)]}{2(r_a^2 + r_b^2 - 2r_a r_b\cos\varphi)} . \end{aligned} \tag{2.19}$$

The functions $T'_{\alpha\beta}$ and $D'_{\alpha\beta}$ are derived in Dodelson et al. (2008) using Equation 2.13 together with the Limber approximation, which is valid since the integration kernels are broad.

$$T'_{\alpha\beta}(\chi_0, r) = \frac{1}{c^2}\int_0^{\chi_0} \mathrm{d}\chi(\chi_0 - \chi)^2 \int_0^{\infty} \frac{k^3 \mathrm{d}k}{\pi} P_\Phi(k, \chi)\big[1 - J_0(kr\chi/\chi_0)\big] \,, \tag{2.20}$$

$$D'_{\alpha\beta}(\chi_0, r) = \frac{2}{c^2}\int_0^{\chi_0} \mathrm{d}\chi(\chi_0 - \chi)^2 \int_0^{\infty} \frac{k^3 \mathrm{d}k}{\pi} P_\Phi(k, \chi) J_2(kr\chi/\chi_0) \,, \tag{2.21}$$

where χ_0 is the comoving distance to the plane containing the three points, $P_\Phi(k, \chi)$ is the power spectrum of the gravitational potential, and $\alpha\beta = ac$, bc or ba. $P_\Phi(k, \chi)$ is related to the matter power spectrum, $P_\delta(k, \chi)$, by

$$P_\Phi(\chi, k) = \frac{9}{4}\frac{H_0^4 \Omega_{\mathrm{m}}^2}{k^4 a^2} P_\delta(\chi, k) \,, \tag{2.22}$$

where a is the scale factor.

Equation 2.12 shows that the deflection effect depends on derivatives of the unlensed 3PCF. This means it will be most significant if the correlation function is rapidly varying, for example near the BAO feature in the matter 3PCF. The dependence on derivatives of the 3PCF also means that the result is independent of galaxy bias so long as we can assume that bias is linear. Linear bias would not be a valid assumption for precision modelling of the BAO since the BAO scale is well within the weakly nonlinear regime. However it is a justifiable assumption for our more broadbrush estimates.

2.3 Results

2.3.1 Preliminaries

In this section we apply the results of Section 2.2 to the three-point correlation function, $\zeta(\mathbf{p}, \mathbf{q}, \mathbf{s})$, of the matter density contrast $\delta(\mathbf{x})$, defined as

$$\zeta(\mathbf{p}, \mathbf{q}, \mathbf{s}) = \langle \delta(\mathbf{x}) \delta(\mathbf{x}+\mathbf{p}) \delta(\mathbf{x}+\mathbf{q}) \rangle_{\mathbf{x}} \,, \tag{2.23}$$

where the subscript $\mathbf{x}$ indicates the average over all spatial positions, and $\mathbf{p}$, $\mathbf{q}$ and $\mathbf{s}$ form the sides of a triangle. We compute the unlensed 3PCF using second-order Eulerian perturbation theory (Jing and Borner 1997). Details are given in Appendix 2B. All results are based on a flat ΛCDM cosmology with $\Omega_{\rm m} = 0.3$, $\Omega_{\rm b} = 0.05$, $h = 0.7$ and $\sigma_8 = 0.9$, and use the nonlinear matter power spectrum from Smith et al. (2003).

In Section 2.3.2 we compare the unlensed 3PCF, ζ, with the lensed 3PCF, $\tilde{\zeta}$, defined by Equation 2.8, across a range of scales and at different redshifts. We present the lensing deflection $|\zeta - \tilde{\zeta}|$, and also the relative deflection $|\zeta - \tilde{\zeta}|/\zeta$. Again we stress that by deflection we mean the difference in the 3PCF due to lensing. To exemplify the properties of the 3PCF and the lensing deflection we give results only for triangles with two equal sides ($r_1 = r_2$) and focus on two illustrative triangle shapes: equilateral triangles ($r_1 = r_2 = r_3$) and a specific 'squeezed' shape with two equal sides with angle $\varphi = 5$ degrees between them. Section 2.3.3 discusses the redshift dependence of the lensing deflection, and in Section 2.3.4 we present results at the BAO scale $80 \leq r \leq 120\ h^{-1}$ Mpc.

2.3.2 Comparison between lensed and unlensed 3PCF

Figure 2.2 compares the magnitudes of the lensed 3PCF and the lensing deflection at $z = 1$ for equilateral and squeezed triangles. At this redshift, typical of current and planned galaxy surveys, the absolute value of the lensing deflection is around 10^{-8}. In general, the lensing effect is larger at small scales because the photon paths are more highly correlated. However the relative contribution is strongest near the BAO feature where the partial derivatives in Equation 2.12 are large.

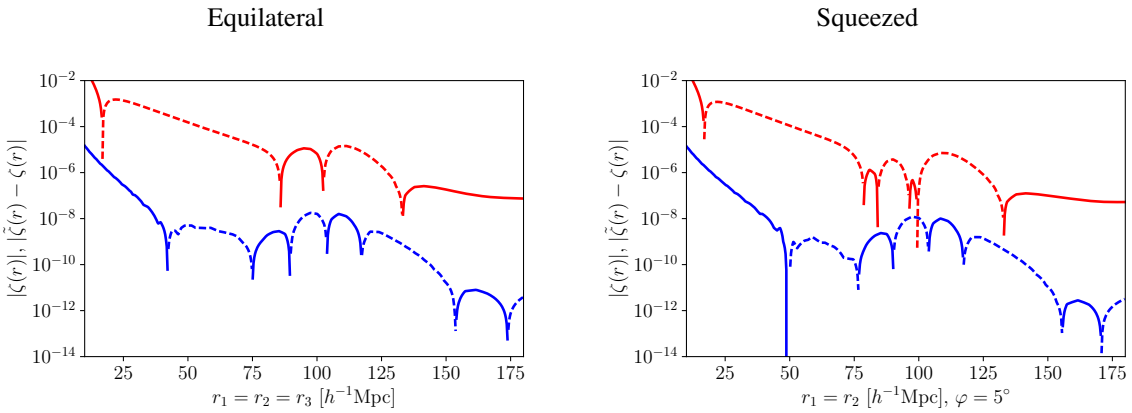

Figure 2.2: Unlensed three-point correlation function (red) and lensing deflection (blue) at $z = 1.0$. *Left*: Equilateral triangles. *Right*: Squeezed triangles with $r_1 = r_2$ and $\varphi = 5$ degrees. Dashed lines indicate negative values.

To explore the shape, rather than size, of the 3PCF it is convenient to define the reduced (or normalised) 3PCF, Q (Hu 2000), which is essentially independent of redshift. It is defined as

$$Q = \frac{\zeta(\mathbf{p}, \mathbf{q}, \mathbf{s})}{\xi(\mathbf{p}, \mathbf{q}) + \xi(\mathbf{q}, \mathbf{s}) + \xi(\mathbf{s}, \mathbf{p})} , \tag{2.24}$$

where ξ is the two-point correlation function. The reduced lensed 3PCF can be defined similarly, with the lensed 3PCF in the numerator and lensed 2PCFs in the denominator. Figure 2.3 shows how the reduced lensed 3PCF at $z = 1$ varies with angle φ between two equal sides $r_1 = r_2 = 20\ h^{-1}$ Mpc. The scale chosen is illustrative of a small scale away from the BAO feature; similar results apply at other scales. The 3PCF attains a minimum for approximately equilateral triangles and increases as the length of the third side decreases (squeezed triangles) or increases (flattened triangles). Lensing has least effect on the 3PCF of equilateral triangles because the correlation function is relatively smooth and lacking in detail; lensing has most effect when the 3PCF is rapidly changing.

2.3.3 Lensing deflection as a function of redshift

Figure 2.4 shows the lensing deflection and the relative deflection at different redshifts for equilateral and squeezed triangles. This demonstrates the significant effect of redshift on the lensing deflection, which has implications for the observability of the effect. The left panel of Figure 2.5 shows how the absolute lensing deflection varies with z for equilateral triangles with sides of 10 h^{-1} Mpc (chosen as illustrative of small scales, although in fact the relationship is similar at all scales). The absolute size of the deflection initially increases as z increases up to about $z = 2$, then decreases.

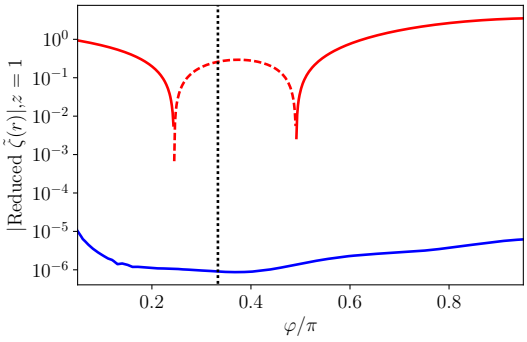

Figure 2.3: Reduced lensed three-point correlation function (red) and lensing deflection (blue) at $z = 1$ for triangles with $r_1 = r_2 = 20\ h^{-1}$ Mpc as a function of angle φ between these sides. The vertical line marks the position of equilateral triangles. Dashed lines indicate negative values.

This also occurs for the lensed 2PCF, also shown in Figure 2.5 for comparison, but in this case the highest deflection is at $z \sim 3$. The shapes of these curves are due to the interplay between the lensing factors in Equations 2.5 and 2.12 and the shapes of the correlation functions, through their derivatives. By contrast, as the right panel of Figure 2.5 shows, the lensing deflection as a proportion of the unlensed 3PCF and 2PCF increases monotonically with redshift. Deflection decreases at higher redshifts but the correlation functions fall more rapidly. This can also be seen in Figure 2.4.

2.3.4 BAO scale

Figure 2.6 'zooms in' on the unlensed 3PCF at the BAO scale at $z = 1$. At this scale the 3PCF oscillates rapidly and vanishes at several points. This is particularly evident for squeezed triangles which display more structure.

Figure 2.7 shows the lensing deflection and the relative deflection near the BAO feature at different redshifts.

Lensing deflection is more prominent at the BAO scale because the partial derivatives in Equation 2.12 can be large. Figures 2.6 and 2.7 show that the lensing deflection smooths out oscillations. At extrema of the 3PCF its first derivatives vanish and the lensing deflection depends on second derivatives. These are positive at local minima, which means that the lensing deflection increases the 3PCF, and negative at local maxima, decreasing the 3PCF. At $z = 1$ the peak-to-trough difference is smoothed by about 0.1 percent. This rises to around 2.3 percent at $z = 10$.

Since the unlensed 3PCF is zero at several values of r, the relative deflection becomes very large in some regions. However, we caution that the observability of the modification due to lensing

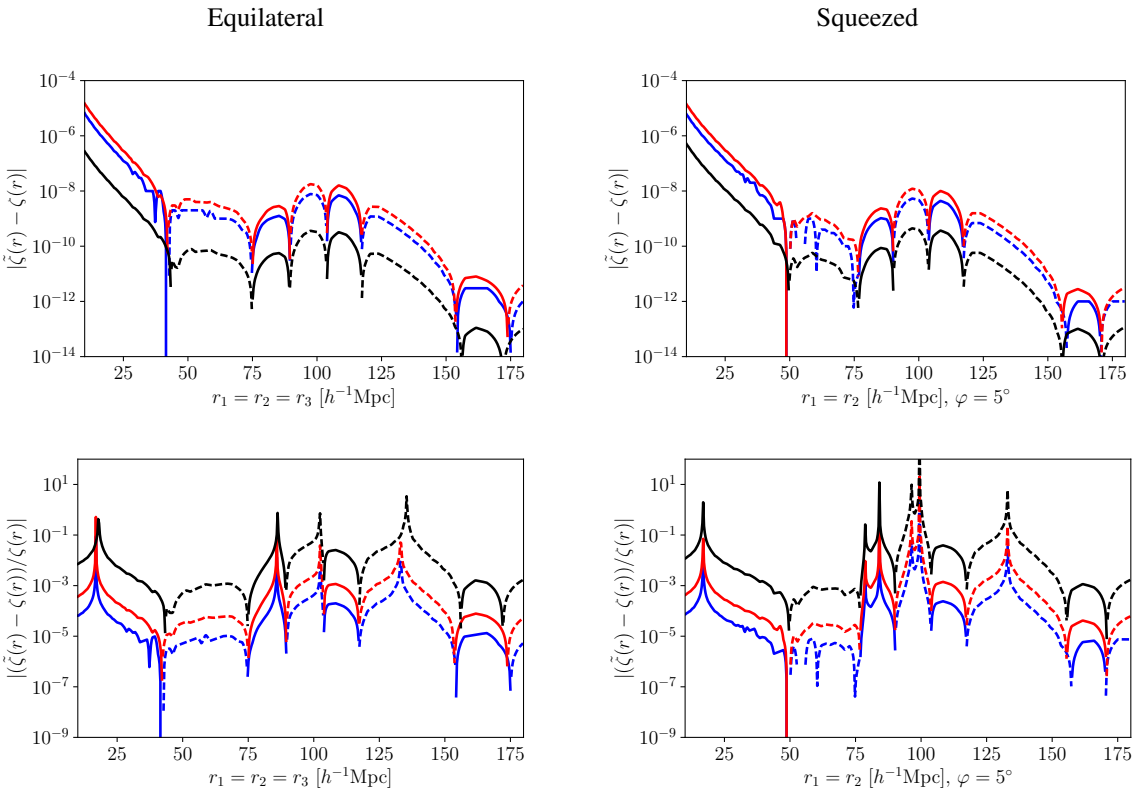

Figure 2.4: Lensing deflection of 3PCF (blue: $z = 0.5$, red: $z = 1.0$, black: $z = 10.0$). *Top*: Lensed - unlensed 3PCF. *Bottom*: As proportion of unlensed 3PCF. *Left*: Equilateral triangles. *Right*: Squeezed triangles. Dashed lines indicate negative values.

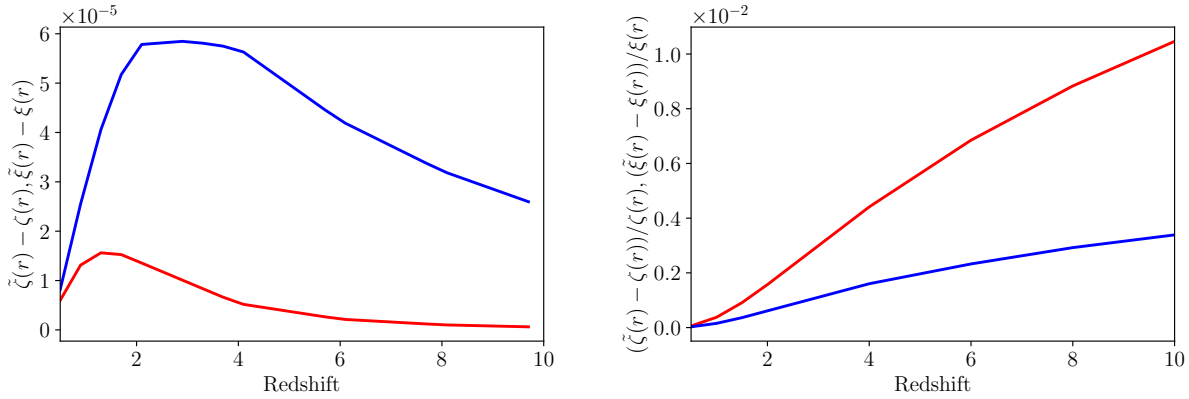

Figure 2.5: Lensing deflection of the 2PCF (blue) and 3PCF (red) for equilateral triangles with sides $r = 10\ h^{-1}$ Mpc as a function of redshift. *Left*: Absolute deflection. *Right*: Relative deflection.

depends on comparisons between the absolute (not relative) deflection and statistical uncertainty on the 3PCF. We discuss this in the next section.

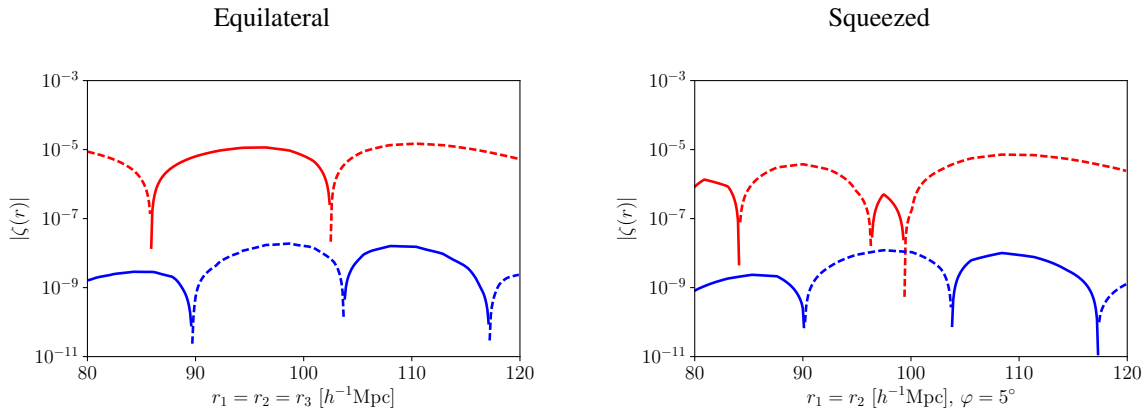

Figure 2.6: Unlensed three-point correlation function (red) and lensing deflection (blue) near the BAO feature at $z = 1$. *Left*: Equilateral triangles. *Right*: Squeezed triangles. Dashed lines indicate negative values.

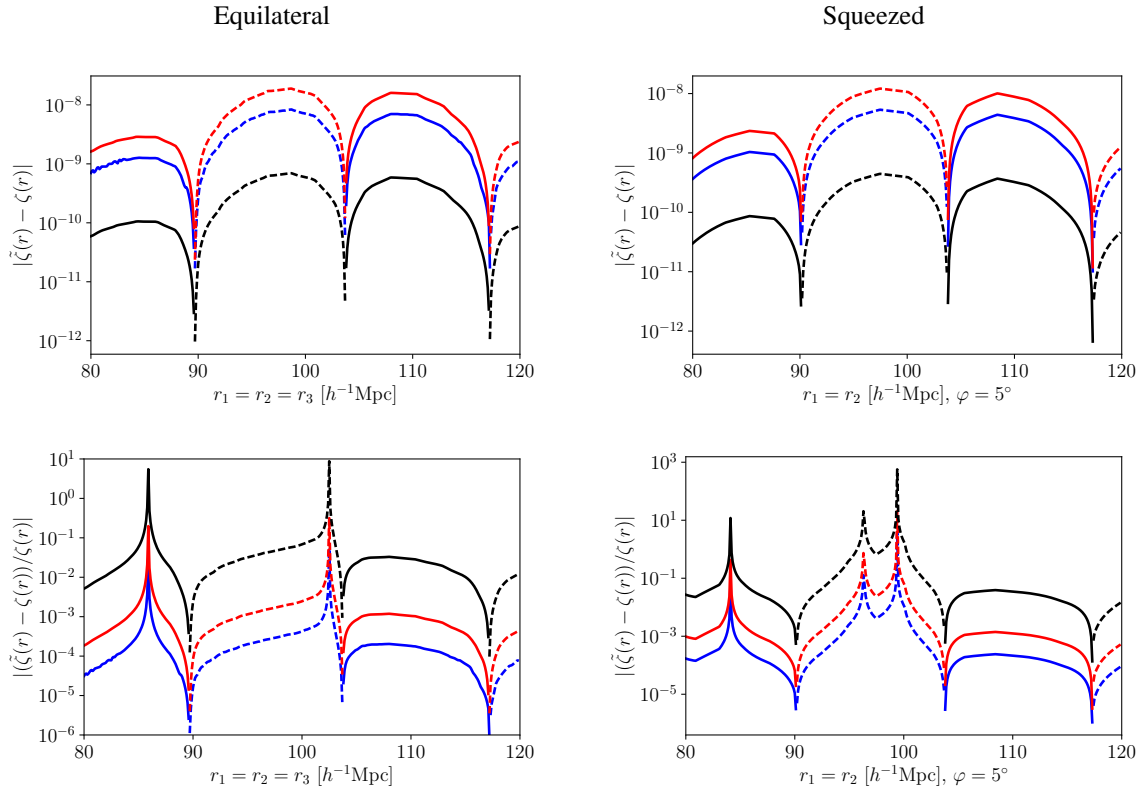

Figure 2.7: Lensing deflection of 3PCF near the BAO feature for different redshifts (blue: $z = 0.5$, red: $z = 1.0$, black: $z = 10.0$). *Top*: Lensed - unlensed 3PCF. *Bottom*: As proportion of unlensed 3PCF. *Left*: Equilateral triangles. *Right*: Squeezed triangles. Dashed lines indicate negative values.

2.4 Observability of the lensing deflection

The previous section shows that at the BAO scale the absolute value of the lensing deflection in the matter 3PCF is around 10^{-8} at $z = 1$. To assess whether a signal of this magnitude could be

detected in current or future galaxy surveys, we assume that the uncertainty in survey measurements is entirely due to Poisson shot noise, ignoring cosmic variance and other Gaussian and non-Gaussian errors which contribute to the full covariance. Thus our error estimates assume a minimum level of statistical error.

To derive an estimate of the shot noise in the 3PCF we build on expressions for the shot noise in measurements of the bispectrum, for 3D fields (Scoccimarro et al. 2004; Chan and Blot 2017) and for projected fields (Joachimi et al. 2009).

Assuming the Gaussian limit, the shot noise $\sigma^2(B)$ in the Gaussian elements of the 3D bispectrum covariance can be estimated as (Scoccimarro et al. 2004)

$$\sigma^2(B) = \frac{s_{123}}{V_s V_B \bar{n}^3} \,, \tag{2.25}$$

where V_s is the survey volume, $\bar{n}$ is the number density, $s_{123} = 1$, 2 or 6 for general, isosceles or equilateral triangles respectively, and V_B quantifies the number of Fourier modes satisfying the triangle constraint $\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 = 0$. It is the integral over triangle side lengths k_1, k_2, k_3 of three spherical shells of width Δk. An analytical expression exists for V_B (Chan and Blot 2017):

$$V_B = \int_{k_1} \mathrm{d}^3 p \int_{k_2} \mathrm{d}^3 q \int_{k_3} \mathrm{d}^3 s \; \delta_{\mathrm{D}}(\mathbf{p} + \mathbf{q} + \mathbf{s}) \tag{2.26}$$

$$= 8\pi^2 k_1 k_2 k_3 (\Delta k)^3 \,. \tag{2.27}$$

The quantity V_B is a purely geometric measure, so similar reasoning applies to the 3PCF covariance. In this case we count modes satisfying the triangle constraint $\mathbf{r}_1 + \mathbf{r}_2 + \mathbf{r}_3 = 0$ within the bin width Δr in real space to produce a quantity V_Z, analogous to V_B. Thus the shot noise in the 3PCF covariance, $\sigma^2(\zeta)$, is given by

$$\sigma^2(\zeta) = \frac{(2\pi)^3 s_{123}}{V_s V_Z \bar{n}^3} \tag{2.28}$$

$$= \frac{6\pi}{V_s r^3 (\Delta r)^3 \bar{n}^3} \text{ for equilateral triangles.} \tag{2.29}$$

For the *Euclid* spectroscopic survey $V_s \approx 100\, h^{-3}\mathrm{Gpc}^3$ and the expected number density of Hα galaxies at $z = 1$ is $\bar{n} \approx 1.7 \times 10^{-3}\, h^3\mathrm{Mpc}^{-3}$ (Amendola et al. 2013). We take $r = 100\, h^{-1}$Mpc (approximately the BAO scale, which is the scale of most interest) and $\Delta r = 10\, h^{-1}$Mpc. Inserting these values into Equation 2.29 implies that $\sigma^2(\zeta)$ is around 10^{-12} and $\sigma(\zeta) \sim 10^{-6}$. These order

of magnitude estimates are consistent with estimates for the uncertainty in the galaxy 3PCF in a survey similar to SDSS DR12 (Slepian and Eisenstein 2015).

Thus, at the BAO scale and for equilateral triangles, the shot noise in a *Euclid*-like survey is greater than the deflection effect. Since we have ignored several sources of error, in practice the errors will exceed the deflection effect by an even greater amount than calculated here.

The deflection is larger for squeezed triangles, but so too is the shot noise. Deflection peaks at around $z = 2$ (Figure 2.5) but the number density of objects in *Euclid*-like surveys falls rapidly up to and beyond this redshift. Other spectroscopic surveys, in particular DESI, will provide samples with similar number densities at $z < 2$ and much sparser QSO samples beyond. Photometric surveys like the LSST in principle observe deep tracer samples with high spatial densities, but the large line-of-sight uncertainties due to broadband photometric redshifts wash out small features like BAO signatures and the lensing deflection modification to the signal. Thus the deflection effect is too small to be detected by forthcoming galaxy and quasar surveys. To put this into context, Schmidt et al. (2008) investigated the effect on the 3PCF of lensing magnification in a magnitude-limited survey. Lensing magnifies or demagnifies objects so that intrinsically faint objects may be brought into the survey, and intrinsically bright objects may be excluded. This can clearly affect measured correlation functions. Schmidt et al. (2008) found the uncertainty in lensing magnification corrections to the 3PCF to be around 10^{-5} across a range of scales at $z = 1$. On the basis of fairly optimistic assumptions they considered the effects to be just detectable in planned surveys.

2.5 Conclusions

We have considered the displacement of sources due to weak gravitational lensing, which we refer to as lensing deflection, and have derived an expression for the effect of lensing deflection on the three-point correlation function for three sources at the same comoving distance. The derivation is quite general: it could be applied to any physical observables and is based only on the assumptions that terms above second order in the lensing deflection can be neglected and that the observables are not correlated with the lensing deflection field.

The resulting expression, given by Equation 2.12, shows that the lensing deflection depends on partial derivatives of the unlensed 3PCF. This causes the 3PCF to be smoothed by lensing, but also makes the effect large when the 3PCF is rapidly varying. If the 3PCF is approximately a power law (which will often be the case), each term of the lensing deflection in Equation 2.12 behaves like

$\langle ABC \rangle / r^2$. Thus as a proportion of the unlensed 3PCF, the deflection effect is highest for sources which are close to each other and decreases as the source separation increases. At all scales the relative effect increases monotonically with redshift.

We have calculated the size of this effect for the matter density contrast and show that lensing deflection is around 10^{-8} at $z = 1$. We have confirmed that the effect is highest at small scales ($r < 20\ h^{-1}$ Mpc), and is also especially noticeable around the BAO feature where it smooths out the peaks and troughs, reducing the amplitude of the oscillations. This could potentially affect the use of the 3PCF for cosmological parameter estimation. The deflection is greatest at around $z = 2$ but as a proportion of the unlensed 3PCF the deflection increases with redshift, reaching 10^{-2} at $z = 10$. A similar results holds for the 2PCF. Here the peak deflection is greater but occurs at higher redshift.

The effect on the 3PCF is too small to be detected in forthcoming surveys such as *Euclid* or DESI. Detections would require much higher number densities of galaxies or quasars at redshifts $z \sim 2$ where the deflection is greatest. While we cannot directly compare our findings with results obtained in Fourier space which were also considered undetectable (Di Dio et al. 2016), we expect them to be of the same order of magnitude.

Although we found that the effect on the matter 3PCF will be undetectable by forthcoming surveys, it remains possible that it is not negligible for other three-point statistics, for example measures of primordial non-Gaussianity or of the CMB bispectrum induced by lensing.

Appendices

2A Effect of lensing on the three-point correlation function

Consider three physical observables $A(\mathbf{x}_a)$, $B(\mathbf{x}_b)$ and $C(\mathbf{x}_c)$ observed at points $\mathbf{x}_a$, $\mathbf{x}_b$ and $\mathbf{x}_c$. The corresponding observed (lensed) values are $\tilde{A}(\mathbf{x}_a)$, $\tilde{B}(\mathbf{x}_b)$ and $\tilde{C}(\mathbf{x}_c)$. Then if $\boldsymbol{\lambda}_a$ is the lensing deflection vector, $\tilde{A}(\mathbf{x}_a)$ and $A(\mathbf{x}_a + \boldsymbol{\lambda}_a)$ are related by Equation (2.1), and similarly for B and C.

The unlensed 3PCF is

$$\langle A(\mathbf{x}_a)B(\mathbf{x}_b)C(\mathbf{x}_c)\rangle = \int \frac{\mathrm{d}^3k_1}{(2\pi)^3}\mathrm{e}^{\mathrm{i}\mathbf{k}_1\cdot\mathbf{x}_a}\int \frac{\mathrm{d}^3k_2}{(2\pi)^3}\mathrm{e}^{\mathrm{i}\mathbf{k}_2\cdot\mathbf{x}_b}\int \frac{\mathrm{d}^3k_3}{(2\pi)^3}\mathrm{e}^{\mathrm{i}\mathbf{k}_3\cdot\mathbf{x}_c} \times (2\pi)^3 B_{ABC}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)\delta_{\mathrm{D}}(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)\,, \tag{2.30}$$

and the lensed 3PCF is

$$\langle \tilde{A}(\mathbf{x}_a)\tilde{B}(\mathbf{x}_b)\tilde{C}(\mathbf{x}_c)\rangle = \langle A(\mathbf{x}_a+\boldsymbol{\lambda_a})B(\mathbf{x}_b+\boldsymbol{\lambda_b})C(\mathbf{x}_c+\boldsymbol{\lambda_c})\rangle \tag{2.31}$$

$$= \int \frac{\mathrm{d}^3k_1}{(2\pi)^3}\mathrm{e}^{\mathrm{i}\mathbf{k}_1\cdot(\mathbf{x}_a+\boldsymbol{\lambda_a})}\int \frac{\mathrm{d}^3k_2}{(2\pi)^3}\mathrm{e}^{\mathrm{i}\mathbf{k}_2\cdot(\mathbf{x}_b+\boldsymbol{\lambda_b})}\int \frac{\mathrm{d}^3k_3}{(2\pi)^3}\mathrm{e}^{\mathrm{i}\mathbf{k}_3\cdot(\mathbf{x}_c+\boldsymbol{\lambda_c})} \times (2\pi)^3 B_{ABC}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)\delta_{\mathrm{D}}(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)\,, \tag{2.32}$$

where δ_{D} is the Dirac delta function and $B_{ABC}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)$ is the bispectrum of the three observables, defined as $\langle A(\mathbf{k}_1)B(\mathbf{k}_2)C(\mathbf{k}_3)\rangle = (2\pi)^3\delta_{\mathrm{D}}(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)B_{ABC}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)$.

We now derive an expression for the lensed 3PCF in terms of the unlensed 3PCF and a lensing deflection term which we denote $\langle ABC\rangle_2$.

Dropping the arguments of $\tilde{A}(\mathbf{x}_a)$ etc for simplicity and making use of the delta function in Equation 2.32 leads to

$$\langle \tilde{A}\tilde{B}\tilde{C}\rangle = \int \frac{\mathrm{d}^3k_1}{(2\pi)^3}\int \frac{\mathrm{d}^3k_2}{(2\pi)^3}\langle \mathrm{e}^{\mathrm{i}\mathbf{k}_1\cdot(\mathbf{x}_a+\boldsymbol{\lambda_a}-\mathbf{x}_c-\boldsymbol{\lambda_c})}\mathrm{e}^{\mathrm{i}\mathbf{k}_2\cdot(\mathbf{x}_b+\boldsymbol{\lambda_b}-\mathbf{x}_c-\boldsymbol{\lambda_c})}\rangle B_{ABC}(\mathbf{k}_1,\mathbf{k}_2,-\mathbf{k}_1-\mathbf{k}_2) \tag{2.33}$$

$$= \int \frac{\mathrm{d}^3k_1}{(2\pi)^3}\int \frac{\mathrm{d}^3k_2}{(2\pi)^3}B_{ABC}(\mathbf{k}_1,\mathbf{k}_2,-\mathbf{k}_1-\mathbf{k}_2) \times \mathrm{e}^{\mathrm{i}\mathbf{k}_1\cdot(\mathbf{x}_a-\mathbf{x}_c)}\mathrm{e}^{\mathrm{i}\mathbf{k}_2\cdot(\mathbf{x}_b-\mathbf{x}_c)}\langle \mathrm{e}^{\mathrm{i}\mathbf{k}_1\cdot(\boldsymbol{\lambda}_a-\boldsymbol{\lambda}_c)}\mathrm{e}^{\mathrm{i}\mathbf{k}_2\cdot(\boldsymbol{\lambda}_b-\boldsymbol{\lambda}_c)}\rangle\,, \tag{2.34}$$

assuming the observables are not correlated with the lensing deflection field.

We expand the exponential factors in the expectation value up to second order in k to get

$$\begin{aligned}\langle\tilde{A}\tilde{B}\tilde{C}\rangle \approx \int \frac{\mathrm{d}^3k_1}{(2\pi)^3}\int \frac{\mathrm{d}^3k_2}{(2\pi)^3} B_{ABC}(\mathbf{k}_1,\mathbf{k}_2,-\mathbf{k}_1-\mathbf{k}_2)\mathrm{e}^{\mathrm{i}\mathbf{k}_1\cdot(\mathbf{x}_a-\mathbf{x}_c)}\mathrm{e}^{\mathrm{i}\mathbf{k}_2\cdot(\mathbf{x}_b-\mathbf{x}_c)} \\ \times\langle[1+\mathrm{i}\mathbf{k}_1\cdot(\boldsymbol{\lambda}_a-\boldsymbol{\lambda}_c)-\frac{1}{2}\mathbf{k}_1\cdot(\boldsymbol{\lambda}_a-\boldsymbol{\lambda}_c)\mathbf{k}_1\cdot(\boldsymbol{\lambda}_a-\boldsymbol{\lambda}_c)] \\ \times[1+\mathrm{i}\mathbf{k}_2\cdot(\boldsymbol{\lambda}_b-\boldsymbol{\lambda}_c)-\frac{1}{2}\mathbf{k}_2\cdot(\boldsymbol{\lambda}_b-\boldsymbol{\lambda}_c)\mathbf{k}_2\cdot(\boldsymbol{\lambda}_b-\boldsymbol{\lambda}_c)]\rangle\,.\end{aligned} \tag{2.35}$$

The zeroth order term is the unlensed 3PCF, $\langle ABC\rangle$. The terms like $\mathrm{i}\mathbf{k}_2\cdot(\boldsymbol{\lambda}_b-\boldsymbol{\lambda}_c)$ are zero because the expectation value of the deflection field is zero. So we have

$$\begin{aligned}\langle\tilde{A}\tilde{B}\tilde{C}\rangle &\approx \langle ABC\rangle + \int \frac{\mathrm{d}^3k_1}{(2\pi)^3}\int \frac{\mathrm{d}^3k_2}{(2\pi)^3} B_{ABC}(\mathbf{k}_1,\mathbf{k}_2,-\mathbf{k}_1-\mathbf{k}_2)\mathrm{e}^{\mathrm{i}\mathbf{k}_1\cdot(\mathbf{x}_a-\mathbf{x}_c)}\mathrm{e}^{\mathrm{i}\mathbf{k}_2\cdot(\mathbf{x}_b-\mathbf{x}_c)} \\ &\quad\times\langle-\frac{1}{2}\mathbf{k}_1\cdot(\boldsymbol{\lambda}_a-\boldsymbol{\lambda}_c)\mathbf{k}_1\cdot(\boldsymbol{\lambda}_a-\boldsymbol{\lambda}_c)-\frac{1}{2}\mathbf{k}_2\cdot(\boldsymbol{\lambda}_b-\boldsymbol{\lambda}_c)\mathbf{k}_2\cdot(\boldsymbol{\lambda}_b-\boldsymbol{\lambda}_c) \\ &\qquad -\mathbf{k}_1\cdot(\boldsymbol{\lambda}_a-\boldsymbol{\lambda}_c)\mathbf{k}_2\cdot(\boldsymbol{\lambda}_b-\boldsymbol{\lambda}_c)\rangle \qquad (2.36)\\ &\equiv \langle ABC\rangle + \langle ABC\rangle_2\,. \qquad (2.37)\end{aligned}$$

The expectation value in Equation 2.36 can be written

$$\begin{aligned}&\langle-\frac{1}{2}\mathbf{k}_1\cdot(\boldsymbol{\lambda}_a-\boldsymbol{\lambda}_c)\mathbf{k}_1\cdot(\boldsymbol{\lambda}_a-\boldsymbol{\lambda}_c)-\frac{1}{2}\mathbf{k}_2\cdot(\boldsymbol{\lambda}_b-\boldsymbol{\lambda}_c)\mathbf{k}_2\cdot(\boldsymbol{\lambda}_b-\boldsymbol{\lambda}_c)-\mathbf{k}_1\cdot(\boldsymbol{\lambda}_a-\boldsymbol{\lambda}_c)\mathbf{k}_2\cdot(\boldsymbol{\lambda}_b-\boldsymbol{\lambda}_c)\rangle \\ &=-\frac{1}{2}\big[k_{1i}k_{1j}[\langle\lambda_a^i\lambda_a^j\rangle-2\langle\lambda_a^i\lambda_c^j\rangle+\langle\lambda_c^i\lambda_c^j\rangle]+k_{2i}k_{2j}[\langle\lambda_b^i\lambda_b^j\rangle-2\langle\lambda_b^i\lambda_c^j\rangle+\langle\lambda_c^i\lambda_c^j\rangle]\big] \\ &\qquad -k_{1i}k_{2j}[\langle\lambda_a^i\lambda_b^j\rangle-\langle\lambda_a^i\lambda_c^j\rangle-\langle\lambda_b^i\lambda_c^j\rangle+\langle\lambda_c^i\lambda_c^j\rangle]\,,\end{aligned} \tag{2.38}$$

where summation over i and j is implied. So the lensing deflection term in Equation 2.36 becomes

$$\begin{aligned}\langle ABC\rangle_2 \approx \int \frac{\mathrm{d}^3k_1}{(2\pi)^3}\int \frac{\mathrm{d}^3k_2}{(2\pi)^3} B_{ABC}(\mathbf{k}_1,\mathbf{k}_2,-\mathbf{k}_1-\mathbf{k}_2)\mathrm{e}^{\mathrm{i}\mathbf{k}_1\cdot(\mathbf{x}_a-\mathbf{x}_c)}\mathrm{e}^{\mathrm{i}\mathbf{k}_2\cdot(\mathbf{x}_b-\mathbf{x}_c)} \\ \times\Big[-\frac{1}{2}\big[k_{1i}k_{1j}[\langle\lambda_a^i\lambda_a^j\rangle-2\langle\lambda_a^i\lambda_c^j\rangle+\langle\lambda_c^i\lambda_c^j\rangle]+k_{2i}k_{2j}[\langle\lambda_b^i\lambda_b^j\rangle-2\langle\lambda_b^i\lambda_c^j\rangle+\langle\lambda_c^i\lambda_c^j\rangle]\big] \\ -k_{1i}k_{2j}[\langle\lambda_a^i\lambda_b^j\rangle-\langle\lambda_a^i\lambda_c^j\rangle-\langle\lambda_b^i\lambda_c^j\rangle+\langle\lambda_c^i\lambda_c^j\rangle]\Big]\,.\end{aligned} \tag{2.39}$$

The deflection vectors do not depend on the integration variables and can be taken out of the integrals to give

$$\begin{aligned}\langle ABC\rangle_2 \approx -\frac{1}{2}[\langle\lambda_a^i\lambda_a^j\rangle - 2\langle\lambda_a^i\lambda_c^j\rangle + \langle\lambda_c^i\lambda_c^j\rangle]\int\frac{\mathrm{d}^3k_1}{(2\pi)^3}\int\frac{\mathrm{d}^3k_2}{(2\pi)^3}k_{1i}k_{1j}B_{ABC}(\mathbf{k}_1,\mathbf{k}_2,-\mathbf{k}_1-\mathbf{k}_2)\\ \times \mathrm{e}^{\mathrm{i}\mathbf{k}_1\cdot(\mathbf{x}_a-\mathbf{x}_c)}\mathrm{e}^{\mathrm{i}\mathbf{k}_2\cdot(\mathbf{x}_b-\mathbf{x}_c)}\\ -\frac{1}{2}[\langle\lambda_b^i\lambda_b^j\rangle - 2\langle\lambda_b^i\lambda_c^j\rangle + \langle\lambda_c^i\lambda_c^j\rangle]\int\frac{\mathrm{d}^3k_1}{(2\pi)^3}\int\frac{\mathrm{d}^3k_2}{(2\pi)^3}k_{2i}k_{2j}B_{ABC}(\mathbf{k}_1,\mathbf{k}_2,-\mathbf{k}_1-\mathbf{k}_2)\\ \times \mathrm{e}^{\mathrm{i}\mathbf{k}_1\cdot(\mathbf{x}_a-\mathbf{x}_c)}\mathrm{e}^{\mathrm{i}\mathbf{k}_2\cdot(\mathbf{x}_b-\mathbf{x}_c)}\\ -[\langle\lambda_a^i\lambda_b^j\rangle - \langle\lambda_a^i\lambda_c^j\rangle - \langle\lambda_b^i\lambda_c^j\rangle + \langle\lambda_c^i\lambda_c^j\rangle]\int\frac{\mathrm{d}^3k_1}{(2\pi)^3}\int\frac{\mathrm{d}^3k_2}{(2\pi)^3}k_{1i}k_{2j}B_{ABC}(\mathbf{k}_1,\mathbf{k}_2,-\mathbf{k}_1-\mathbf{k}_2)\\ \times \mathrm{e}^{\mathrm{i}\mathbf{k}_1\cdot(\mathbf{x}_a-\mathbf{x}_c)}\mathrm{e}^{\mathrm{i}\mathbf{k}_2\cdot(\mathbf{x}_b-\mathbf{x}_c)}\,.\end{aligned} \tag{2.40}$$

Now define the correlators as in Equation 2.9. Then the final correlator term in Equation 2.38 can be written as

$$\begin{aligned}\langle\lambda_a^i\lambda_b^j\rangle - \langle\lambda_a^i\lambda_c^j\rangle - \langle\lambda_b^i\lambda_c^j\rangle + \langle\lambda_c^i\lambda_c^j\rangle &= -\left(\frac{\langle\lambda_a^i\lambda_a^j\rangle + \langle\lambda_b^i\lambda_b^j\rangle}{2} - \langle\lambda_a^i\lambda_b^j\rangle\right) + \left(\frac{\langle\lambda_a^i\lambda_a^j\rangle + \langle\lambda_b^i\lambda_b^j\rangle}{2}\right)\\ &\quad + \left(\frac{\langle\lambda_a^i\lambda_a^j\rangle + \langle\lambda_c^i\lambda_c^j\rangle}{2} - \langle\lambda_a^i\lambda_c^j\rangle\right) - \left(\frac{\langle\lambda_a^i\lambda_a^j\rangle + \langle\lambda_c^i\lambda_c^j\rangle}{2}\right)\\ &\quad + \left(\frac{\langle\lambda_b^i\lambda_b^j\rangle + \langle\lambda_c^i\lambda_c^j\rangle}{2} - \langle\lambda_b^i\lambda_c^j\rangle\right) - \left(\frac{\langle\lambda_b^i\lambda_b^j\rangle + \langle\lambda_c^i\lambda_c^j\rangle}{2}\right)\\ &\quad + \langle\lambda_c^i\lambda_c^j\rangle\\ &= Z_{ac}^{ij} + Z_{bc}^{ij} - Z_{ab}^{ij}\,.\end{aligned} \tag{2.41}$$

So the lensing deflection is given by

$$\begin{aligned}\langle ABC\rangle_2 \approx -Z_{ac}^{ij}\int\frac{\mathrm{d}^3k_1}{(2\pi)^3}\int\frac{\mathrm{d}^3k_2}{(2\pi)^3}k_{1i}k_{1j}B_{ABC}(\mathbf{k}_1,\mathbf{k}_2,-\mathbf{k}_1-\mathbf{k}_2)\\ \times \mathrm{e}^{\mathrm{i}\mathbf{k}_1\cdot(\mathbf{x}_a-\mathbf{x}_c)}\mathrm{e}^{\mathrm{i}\mathbf{k}_2\cdot(\mathbf{x}_b-\mathbf{x}_c)}\\ -Z_{bc}^{ij}\int\frac{\mathrm{d}^3k_1}{(2\pi)^3}\int\frac{\mathrm{d}^3k_2}{(2\pi)^3}k_{2i}k_{2j}B_{ABC}(\mathbf{k}_1,\mathbf{k}_2,-\mathbf{k}_1-\mathbf{k}_2)\\ \times \mathrm{e}^{\mathrm{i}\mathbf{k}_1\cdot(\mathbf{x}_a-\mathbf{x}_c)}\mathrm{e}^{\mathrm{i}\mathbf{k}_2\cdot(\mathbf{x}_b-\mathbf{x}_c)}\\ -[Z_{ac}^{ij}+Z_{bc}^{ij}-Z_{ab}^{ij}]\int\frac{\mathrm{d}^3k_1}{(2\pi)^3}\int\frac{\mathrm{d}^3k_2}{(2\pi)^3}k_{1i}k_{2j}B_{ABC}(\mathbf{k}_1,\mathbf{k}_2,-\mathbf{k}_1-\mathbf{k}_2)\\ \times \mathrm{e}^{\mathrm{i}\mathbf{k}_1\cdot(\mathbf{x}_a-\mathbf{x}_c)}\mathrm{e}^{\mathrm{i}\mathbf{k}_2\cdot(\mathbf{x}_b-\mathbf{x}_c)}\,.\end{aligned} \tag{2.42}$$

We now define $\mathbf{r}_b$ as the observed distance between $\mathbf{x}_a$ and $\mathbf{x}_c$, and $\mathbf{r}_a$ as the observed distance between $\mathbf{x}_b$ and $\mathbf{x}_c$, so that $\mathbf{r}_b \equiv (\mathbf{x}_a - \mathbf{x}_c)$ and $\mathbf{r}_a \equiv (\mathbf{x}_b - \mathbf{x}_c)$ (see Figure 2.1). Then from Equation 2.30

$$\begin{aligned}\langle A(\mathbf{x}_a)B(\mathbf{x}_b)C(\mathbf{x}_c)\rangle = \int\frac{\mathrm{d}^3k_1}{(2\pi)^3}\mathrm{e}^{\mathrm{i}\mathbf{k}_1\cdot(\mathbf{r}_b+\mathbf{x}_c)}\int\frac{\mathrm{d}^3k_2}{(2\pi)^3}\mathrm{e}^{\mathrm{i}\mathbf{k}_2\cdot(\mathbf{r}_a+\mathbf{x}_c)}\int\frac{\mathrm{d}^3k_3}{(2\pi)^3}\mathrm{e}^{\mathrm{i}\mathbf{k}_3\cdot\mathbf{x}_c}\\ \times(2\pi)^3B_{ABC}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)\delta_{\mathrm{D}}(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)\,,\end{aligned} \tag{2.43}$$

and so partial derivatives with respect to components of the triangle sides can be written as, for example,

$$\frac{\partial^2\langle ABC\rangle}{\partial r_{bi}\partial r_{bj}} = -k_{1i}k_{1j}\langle ABC\rangle\,. \tag{2.44}$$

It follows that the lensing deflection, $\langle ABC\rangle_2$, can be expressed in terms of partial derivatives of the unlensed 3PCF, $\langle ABC\rangle$:

$$\begin{aligned}\langle ABC\rangle_2 = &\frac{\partial^2\langle ABC\rangle}{\partial r_{bi}\partial r_{bj}}Z_{ac}^{ij}\\ &+\frac{\partial^2\langle ABC\rangle}{\partial r_{ai}\partial r_{aj}}Z_{bc}^{ij}\\ &+\frac{\partial^2\langle ABC\rangle}{\partial r_{ai}\partial r_{bj}}[Z_{ac}^{ij}+Z_{bc}^{ij}-Z_{ab}^{ij}]\,.\end{aligned} \tag{2.45}$$

Following Dodelson et al. (2008) we now make the simplifying assumption that all three sources are at the same comoving distance. Without loss of generality coordinates can be chosen as in Figure 2.1. This means that the only elements of the distortion correlators Z^{ij} which we need to consider are those with $i = j = 1$ and $i = j = 2$. These represent deflections in two orthogonal directions in the plane of the sky. With these coordinates Equation 2.45 becomes

$$\begin{aligned}
\langle ABC\rangle_2 = &\frac{\partial^2\langle ABC\rangle}{\partial r_{b1}^2} Z_{ac}^{11} + \frac{\partial^2\langle ABC\rangle}{\partial r_{b2}^2} Z_{ac}^{22} \\
&+ \frac{\partial^2\langle ABC\rangle}{\partial r_{a1}^2} Z_{bc}^{11} + \frac{\partial^2\langle ABC\rangle}{\partial r_{a2}^2} Z_{bc}^{22} \\
&+ \frac{\partial^2\langle ABC\rangle}{\partial r_{a1}\partial r_{b1}} [Z_{ac}^{11} + Z_{bc}^{11} - Z_{ab}^{11}] \\
&+ \frac{\partial^2\langle ABC\rangle}{\partial r_{a2}\partial r_{b2}} [Z_{ac}^{22} + Z_{bc}^{22} - Z_{ab}^{22}] \qquad (2.46) \\
= &\frac{\partial^2\langle ABC\rangle}{\partial r_{b1}^2} Z_{ac}^{11} + \frac{\partial^2\langle ABC\rangle}{\partial r_{a1}^2} Z_{bc}^{11} + \frac{\partial^2\langle ABC\rangle}{\partial r_{a2}^2} Z_{bc}^{22} \\
&+ \frac{\partial^2\langle ABC\rangle}{\partial r_{a1}\partial r_{b1}} [Z_{ac}^{11} + Z_{bc}^{11} - Z_{ab}^{11}] \qquad (2.47)
\end{aligned}$$

because all derivatives with respect to r_{b2} are zero through the choice of coordinates.

We next express these derivatives in terms of the distances r_a and r_b and the angle φ between $\mathbf{r}_a$ and $\mathbf{r}_b$. We have $r_{b1} \equiv r_b$ so partial derivatives with respect to r_{b1} are straightforward. We now derive the other partial derivatives which appear in Equation 2.47. For brevity we write $\langle ABC\rangle$ as $f \equiv f(r_a, r_b, \varphi)$.

1. Partial derivative with respect to r_{a1}.

$$\begin{aligned}
\frac{\partial f(r_a, r_b, \varphi)}{\partial r_{a_1}} &= \frac{\partial f}{\partial r_a}\frac{\partial r_a}{\partial r_{a_1}} + \frac{\partial f}{\partial r_b}\frac{\partial r_b}{\partial r_{a_1}} + \frac{\partial f}{\partial \varphi}\frac{\partial \varphi}{\partial r_{a_1}} \\
&= \cos\varphi \frac{\partial f}{\partial r_a} - \frac{\sin\varphi}{r_a}\frac{\partial f}{\partial \varphi} \, . \qquad (2.48)
\end{aligned}$$

This uses $r_a = (r_{a1}^2 + r_{a2}^2)^{1/2}$ which means $\frac{\partial r_a}{\partial r_{a1}} = \frac{r_{a1}}{(r_{a1}^2 + r_{a2}^2)^{1/2}} = {}^{r_{a1}}\!/_{r_a} = \cos\varphi$,
and $\varphi = \arctan({}^{r_{a2}}\!/_{r_{a1}})$ which means $\frac{\partial \varphi}{\partial r_{a1}} = \left(\frac{1}{1+({}^{r_{a2}}/_{r_{a1}})^2}\right)({}^{-r_{a2}}\!/_{r_{a1}^2}) = ({}^{r_{a1}^2}\!/_{r_a^2})({}^{-r_{a2}}\!/_{r_{a1}^2}) = {}^{-\sin\varphi}\!/_{r_a}$.

2. Second partial derivative with respect to r_{a1}.

$$\begin{aligned}\frac{\partial^2 f(r_a, r_b, \varphi)}{\partial r_{a_1}^2} &= \left[\cos\varphi\frac{\partial}{\partial r_a} - \frac{\sin\varphi}{r_a}\frac{\partial}{\partial\varphi}\right]\left[\cos\varphi\frac{\partial f}{\partial r_a} - \frac{\sin\varphi}{r_a}\frac{\partial f}{\partial\varphi}\right] \\ &= \cos^2\varphi\frac{\partial^2 f}{\partial r_a^2} - \frac{2\sin\varphi\cos\varphi}{r_a}\frac{\partial^2 f}{\partial r_a\partial\varphi} + \frac{\sin^2\varphi}{r_a^2}\frac{\partial^2 f}{\partial\varphi^2} \\ &\quad + \frac{\sin^2\varphi}{r_a}\frac{\partial f}{\partial r_a} + \frac{2\sin\varphi\cos\varphi}{r_a^2}\frac{\partial f}{\partial\varphi}\,. \end{aligned} \tag{2.49}$$

3. Second partial derivative with respect to r_{a1} and r_{b1}.

$$\begin{aligned}\frac{\partial^2 f(r_a, r_b, \varphi)}{\partial r_{a_1}\partial r_{b_1}} &= \frac{\partial^2 f(r_a, r_b, \varphi)}{\partial r_{a_1}\partial r_b} \\ &= \cos\varphi\frac{\partial^2 f}{\partial r_a\partial r_b} - \frac{\sin\varphi}{r_a}\frac{\partial^2 f}{\partial r_b\partial\varphi}\,. \end{aligned} \tag{2.50}$$

4. Partial derivative with respect to r_{a2}.

$$\begin{aligned}\frac{\partial f(r_a, r_b, \varphi)}{\partial r_{a_2}} &= \frac{\partial f}{\partial r_a}\frac{\partial r_a}{\partial r_{a_2}} + \frac{\partial f}{\partial r_b}\frac{\partial r_b}{\partial r_{a_2}} + \frac{\partial f}{\partial\varphi}\frac{\partial\varphi}{\partial r_{a_2}} \\ &= \sin\varphi\frac{\partial f}{\partial r_a} + \frac{\cos\varphi}{r_a}\frac{\partial f}{\partial\varphi}\,. \end{aligned} \tag{2.51}$$

This uses $\frac{\partial r_a}{\partial r_{a2}} = \frac{r_{a2}}{(r_{a1}^2+r_{a2}^2)^{1/2}} = \sin\varphi$, and $\frac{\partial\varphi}{\partial r_{a2}} = \left(\frac{1}{1+(r_{a2}/r_{a1})^2}\right)\left(\frac{1}{r_{a1}}\right) = \frac{\cos\varphi}{r_a}$.

5. Second partial derivative with respect to r_{a2}.

$$\begin{aligned}\frac{\partial^2 f(r_a, r_b, \varphi)}{\partial r_{a_2}^2} &= \left[\sin\varphi\frac{\partial}{\partial r_a} + \frac{\cos\varphi}{r_a}\frac{\partial}{\partial\varphi}\right]\left[\sin\varphi\frac{\partial f}{\partial r_a} + \frac{\cos\varphi}{r_a}\frac{\partial f}{\partial\varphi}\right] \\ &= \sin^2\varphi\frac{\partial^2 f}{\partial r_a^2} + \frac{2\sin\varphi\cos\varphi}{r_a}\frac{\partial^2 f}{\partial r_a\partial\varphi} + \frac{\cos^2\varphi}{r_a^2}\frac{\partial^2 f}{\partial\varphi^2} \\ &\quad + \frac{\cos^2\varphi}{r_a}\frac{\partial f}{\partial r_a} - \frac{2\sin\varphi\cos\varphi}{r_a^2}\frac{\partial f}{\partial\varphi}\,. \end{aligned} \tag{2.52}$$

Having assembled these ingredients we can substitute into Equation 2.47 to get

$$\begin{aligned}\langle ABC\rangle_2 = Z_{ac}^{11}&\frac{\partial^2\langle ABC\rangle}{\partial r_b^2}\\
&+ Z_{bc}^{11}\left[\cos^2\varphi\frac{\partial^2\langle ABC\rangle}{\partial r_a^2} - \frac{2\sin\varphi\cos\varphi}{r_a}\frac{\partial^2\langle ABC\rangle}{\partial r_a\partial\varphi} + \frac{\sin^2\varphi}{r_a^2}\frac{\partial^2\langle ABC\rangle}{\partial\varphi^2}\right.\\
&\left.+ \frac{\sin^2\varphi}{r_a}\frac{\partial\langle ABC\rangle}{\partial r_a} + \frac{2\sin\varphi\cos\varphi}{r_a^2}\frac{\partial\langle ABC\rangle}{\partial\varphi}\right]\\
&+ Z_{bc}^{22}\left[\sin^2\varphi\frac{\partial^2\langle ABC\rangle}{\partial r_a^2} + \frac{2\sin\varphi\cos\varphi}{r_a}\frac{\partial^2\langle ABC\rangle}{\partial r_a\partial\varphi} + \frac{\cos^2\varphi}{r_a^2}\frac{\partial^2\langle ABC\rangle}{\partial\varphi^2}\right.\\
&\left.+ \frac{\cos^2\varphi}{r_a}\frac{\partial\langle ABC\rangle}{\partial r_a} - \frac{2\sin\varphi\cos\varphi}{r_a^2}\frac{\partial\langle ABC\rangle}{\partial\varphi}\right]\\
&+ [Z_{ac}^{11} + Z_{bc}^{11} - Z_{ab}^{11}]\left[\cos\varphi\frac{\partial^2\langle ABC\rangle}{\partial r_a\partial r_b} - \frac{\sin\varphi}{r_a}\frac{\partial^2\langle ABC\rangle}{\partial r_b\partial\varphi}\right]. \qquad (2.53)\end{aligned}$$

To finish we collect together terms in each partial derivative and obtain Equation 2.12, the final result for the deflection contribution to the lensed 3PCF.

2B The unlensed matter three-point correlation function

The three-point correlation function $\zeta(\mathbf{r_1}, \mathbf{r_2}, \mathbf{r_3})$ of the matter density field is defined as in Equation 2.30, with the matter bispectrum in place of the general bispectrum. The matter bispectrum can be computed using Eulerian perturbation theory (Fry 1984; Bernardeau et al. 2002) as

$$B_{\rm PT}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = \left[\frac{10}{7} + \frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{k_1 k_2}\left(\frac{k_1^2 + k_2^2}{k_1 k_2}\right) + \frac{4}{7}\left(\frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{k_1 k_2}\right)^2\right] P_\delta(k_1) P_\delta(k_2) + 2 \text{ perms.} , \tag{2.54}$$

where $P_\delta(k)$ is the matter power spectrum.

From this it is possible to derive the following expression for the three-point correlation function (Jing and Borner 1997):

$$\begin{aligned}\zeta(\mathbf{r_1}, \mathbf{r_2}, \mathbf{r_3}) = \frac{10}{7}\xi(r_{21})\xi(r_{31}) - \big[\eta_2(r_{21})\eta_0(r_{31}) + \eta_0(r_{21})\eta_2(r_{31})\big]\mathbf{r}_{21} \cdot \mathbf{r}_{31} \\ + \frac{4}{7}\big[\epsilon(r_{21})\epsilon_2(r_{31})(\mathbf{r}_{21} \cdot \mathbf{r}_{31})^2 + \epsilon(r_{21})\eta_2(r_{31})r_{21}^2 \\ + \eta_2(r_{21})\epsilon(r_{31})r_{31}^2 + 3\eta_2(r_{21})\eta_2(r_{31})\big] + 2 \text{ perms.}\end{aligned} \tag{2.55}$$

where $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$.

The functions $\xi(r)$, $\eta_l(r)$ and $\epsilon(r)$ are given by

$$\xi(r) = \frac{1}{2\pi^2}\int_0^\infty \mathrm{d}k\, k^2 P_\delta(k) j_0(kr) , \tag{2.56}$$

$$\eta_l(r) = -\frac{1}{2\pi^2}\int_0^\infty \mathrm{d}k\, k^2 P_\delta(k) \frac{k}{k^l r} j_1(kr) , \tag{2.57}$$

$$\epsilon(r) = \frac{1}{2\pi^2}\int_0^\infty \mathrm{d}k\, k^2 P_\delta(k) \frac{k^2}{r^2} j_2(kr) . \tag{2.58}$$

$\xi(r)$ is the two-point correlation function. $\eta_l(r)$ has two variants with $l = 0$ and $l = 2$.
These expressions can be problematic to integrate numerically because the Bessel functions are oscillatory. However they can in fact be evaluated efficiently with Fast Fourier Transforms (FFT) (Hamilton 2000; Simon 2005). To achieve this we can use the fact that it is possible to transform

equations of the form

$$f(r) = \int_0^\infty \mathrm{d}k\, rF(kr)\hat{f}(k)\ , \tag{2.59}$$

for some function F, with $f(r)$ and $\hat{f}(k)$ a Hankel transform pair,

$$\hat{f}(k) = \int_0^\infty \mathrm{d}r\, kF(kr)f(r)\ , \tag{2.60}$$

in a way that makes them readily integrable. To see this we make a change of variables $r \equiv \mathrm{e}^x$ and $k \equiv \mathrm{e}^y$ and define $g(x) \equiv f(\mathrm{e}^x)$ and $\hat{g}(y) \equiv \hat{f}(\mathrm{e}^y)$. Then Equation 2.59 becomes

$$g(x) = \int_{-\infty}^\infty \mathrm{d}y\ \mathrm{e}^{x+y}F(e^{x+y})\hat{g}(y) \tag{2.61}$$

$$= \int_{-\infty}^\infty \mathrm{d}y\ G(x+y)\hat{g}(y)\ , \tag{2.62}$$

where $G(z)$ is defined as $\mathrm{e}^z F(\mathrm{e}^z)$.

The integral in Equation 2.62 is the cross-correlation $G * \hat{g}(y)$ which is equivalent to the convolution $G * \hat{g}^*(-y)$ where $\hat{g}^*$ is the complex conjugate of $\hat{g}$. We can thus avoid the need for integration by transforming to Fourier space where convolutions become products.

To transform Equations 2.56 to 2.58 to the required form we change the spherical Bessel functions to cylindrical Bessel functions using the identity

$$j_\mu(z) = \sqrt{\frac{\pi}{2z}} J_{\mu+\frac{1}{2}}(z)\ . \tag{2.63}$$

We can then write the equations as:

$$\xi(r) = \sqrt{\frac{\pi}{2}}\frac{1}{2\pi^2} r^{-1} \int_0^\infty \mathrm{d}k\ k^2 P_\delta(k) r (kr)^{-\frac{1}{2}} J_{\frac{1}{2}}(kr)\ , \tag{2.64}$$

$$\eta_0(r) = -\sqrt{\frac{\pi}{2}}\frac{1}{2\pi^2} r^{-3} \int_0^\infty \mathrm{d}k\ k^2 P_\delta(k) r (kr)^{\frac{1}{2}} J_{\frac{3}{2}}(kr)\ , \tag{2.65}$$

$$\eta_2(r) = -\sqrt{\frac{\pi}{2}}\frac{1}{2\pi^2} r^{-1} \int_0^\infty \mathrm{d}k\ k^2 P_\delta(k) r (kr)^{-\frac{3}{2}} J_{\frac{3}{2}}(kr)\ , \tag{2.66}$$

$$\epsilon(r) = \sqrt{\frac{\pi}{2}}\frac{1}{2\pi^2} r^{-3} \int_0^\infty \mathrm{d}k\ k^2 P_\delta(k) r (kr)^{\frac{1}{2}} J_{\frac{5}{2}}(kr)\ . \tag{2.67}$$

These have the form of Equation 2.59.

3 Information from the weak lensing bispectrum

3.1 Introduction

One of the primary aims of modern cosmology is to constrain the cosmological parameters. Weak gravitational lensing is becoming an increasingly reliable tool for this purpose, with statistical and systematic errors continually being reduced.

The next generation of galaxy surveys such as *Euclid*[1] and LSST[2] will represent a step change in the quantity and precision of weak lensing data. It is important to plan how best to use this data to estimate parameters, including investigating methods which are not feasible for current surveys. In this chapter we consider the potential for using three-point weak lensing statistics to analyse this new survey data.

Despite formidable observational and analytical challenges, there is no question that a *Euclid*-like survey should be able to provide three-point measurements. A three-point weak lensing signal was detected as early as 2003 (Bernardeau et al. 2003; Pen et al. 2003). Subsequently Semboloni et al. (2010) used three-point aperture mass statistics from the Cosmic Evolution Survey[3] to estimate cosmological parameters. This work was an important proof of concept. Despite a small, less than ideal survey, three-point statistics produced a modest improvement in parameter constraints. More recently the usefulness of three-point measures was confirmed by Fu et al. (2014) using the larger Canada France Hawaii Telescope Lensing Survey[4].

With the advent of high-quality measurements from surveys specifically designed for weak lensing, there are several reasons to believe that three-point statistics could help improve parameter constraints. Most obviously, the shear field is non-Gaussian so two-point statistics do not capture all potential information. Including three-point statistics should therefore help to reduce statistical errors. Secondly, the matter bispectrum and power spectrum depend on cosmological parameters in different ways, offering the prospect of breaking degeneracies such as those between $\Omega_{\rm m}$ and σ_8. For the weak lensing power spectrum and bispectrum this degeneracy-breaking may be strengthened because they are differently-weighted projections of their matter counterparts. Finally, three-point

[1] http://sci.esa.int/euclid/
[2] lsst.org
[3] http://cosmos.astro.caltech.edu
[4] http://www.cfhtlens.org

statistics may help to control systematic uncertainties: there is evidence that systematics affect the bispectrum and power spectrum in different ways (Semboloni et al. 2008; Foreman et al. 2019).

To date most weak lensing studies have used two-point statistics. This is often a purely pragmatic choice. The bispectrum is significantly more complicated than the power spectrum which means that the (possibly modest) improvement in parameter constraints must be weighed against increased observational, analytical and computational costs. In the first place, the data vector is much larger for the bispectrum than for the power spectrum. The power spectrum is characterised by a single wavenumber but the bispectrum depends on triangle configurations defined by three pieces of information, for example the lengths of three wavevectors or two wavevectors and an angle. For a *Euclid*-like survey with 10 tomographic bins there are $\sim 10^5$ different triangle configurations, compared with only $\sim 10^3$ combinations of single wavevectors. Moreover to estimate parameter values we need to evaluate the covariance matrix, which again is much larger for the bispectrum. It is also more complex: if all Gaussian and non-Gaussian contributions are considered the bispectrum covariance depends on up to a six-point function whereas the power spectrum covariance depends only on a trispectrum. These problems of size and complexity affect observations, simulations and numerical analysis alike.

On top of this, several poorly-understood systematic effects degrade the weak lensing signal. Quantifying and controlling these systematics even for two-point statistics is difficult and their effect on three-point statistics is not well understood. Principal amongst these systematics are the intrinsic alignments of galaxies which can mimic the shear signal (Troxel and Ishak 2015), uncertainty related to the use of photometric rather than spectroscopic redshifts, and the impact of baryonic effects at the small non-linear scales which will be probed by future surveys. As statistical uncertainties are reduced, control of systematic uncertainties will become an increasingly pressing issue.

Several studies, in particular Kayo et al. (2012), Kayo and Takada (2013) and Rizzato et al. (2019), have investigated the theoretical gain in information from using the weak lensing bispectrum as well as the power spectrum. Importantly, these studies did not take account of systematic uncertainties so their conclusions must be considered to be optimistic. In this chapter we first confirm the main findings of these studies and then consider how systematic uncertainties affect the bispectrum.

In Section 3.2 we review the matter bispectrum covariance matrix. We clarify the supersample covariance and in Section 3.2.3 we use the halo model, perturbation theory and the so-called 'integrated trispectrum' to derive all the matter bispectrum supersample covariance terms for

equilateral triangles. Other authors have reported similar derivations and we contrast our methods and results with those of Adhikari et al. (2016) and Chan et al. (2018). Finally we extend the results to the weak lensing supersample covariance. These are summarised in Appendix 3A.

In Section 3.3 we consider statistical errors in a next-generation survey. We present Fisher matrix forecasts which confirm previous findings (Kayo et al. 2012; Rizzato et al. 2019) about the benefit of tomography and the potential reduction in statistical errors from including the bispectrum. Our Fisher matrix analysis was mostly complete before the publication of Rizzato et al. (2019).

In Section 3.4 we discuss two major sources of systematic error, intrinsic alignments and uncertainty in redshift binning, and show that using the bispectrum as well as the power spectrum can help to mitigate the effect of these systematics.

Section 3.5 concludes with a discussion of further work which would build on our results.

Throughout we consider a *Euclid*-like survey. Details of survey and modelling assumptions and of fiducial parameter values are given in Section 3.3.1.

3.2 Matter power spectrum and bispectrum covariance

3.2.1 Power spectrum and bispectrum estimators

Before addressing the power spectrum and bispectrum covariance matrices we first define power spectrum and bispectrum estimators. We initially discuss the matter density contrast $\delta(\mathbf{x})$ and later extend this to the weak lensing field, since this is just a weighted projection of the matter field. We assume a survey with volume V_W and define a survey window function $W(\mathbf{x})$ which is a binary mask with $W(\mathbf{x}) = 1$ within the survey volume and zero otherwise. Thus

$$V_W \equiv \int \mathrm{d}^3\mathbf{x} W(\mathbf{x}) \,. \tag{3.1}$$

The observed field is then $\delta_W(\mathbf{x}) \equiv \delta(\mathbf{x})W(\mathbf{x})$.

We assume that observations are binned into concentric shells of width Δk with $\Delta k \ll k$ and that the volume of the shell labelled by k is $V_P(k)$. Then we can define the matter power spectrum estimator $\hat{P}(k)$ as (Scoccimarro et al. 1999; Sefusatti et al. 2006; Takada and Hu 2013)

$$\hat{P}(k) = \frac{1}{V_W V_P(k)} \int_k \mathrm{d}^3 q_1 \int_k \mathrm{d}^3 q_2 \; \tilde{\delta}_W(\mathbf{q}_1)\tilde{\delta}_W(-\mathbf{q}_2)\delta_\mathrm{D}(\mathbf{q}_{12}) \,, \tag{3.2}$$

where $\tilde{\delta}_W(\mathbf{k})$ is the Fourier transform of the matter density contrast convolved with the survey window function and $\delta_{\rm D}(\mathbf{q}_{12}) \equiv \delta_{\rm D}(\mathbf{q}_1 + \mathbf{q}_2)$ is the Dirac delta function. The integrals are over all modes which fall into the spherical shell defined by k, that is all modes with $|\mathbf{q}_i| \in k$. $V_P(k)$ is given by (Sefusatti et al. 2006)

$$V_P(k) = \int_k \mathrm{d}q_1 \int_k \mathrm{d}q_2 \; \delta_{\rm D}(\mathbf{q}_{12}) \tag{3.3}$$

$$\approx 4\pi k^2 \Delta k \; . \tag{3.4}$$

For modes with wavenumber much larger than the length scale of the survey $\hat{P}$ is an unbiased estimator (Takada and Hu 2013).

We can use a similar approach to define a bispectrum estimator $\hat{B}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)$ (Sefusatti et al. 2006):

$$\hat{B}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = \frac{1}{V_W V_B(k_1, k_2, k_3)} \int_{k_1} \mathrm{d}^3 q_1 \int_{k_2} \mathrm{d}^3 q_2 \int_{k_3} \mathrm{d}^3 q_3 \; \delta_{\rm D}(\mathbf{q}_{123}) \times \tilde{\delta}_W(\mathbf{q_1}) \tilde{\delta}_W(\mathbf{q_2}) \tilde{\delta}_W(\mathbf{q_3}) \tag{3.5}$$

where $\delta_{\rm D}(\mathbf{q}_{123}) \equiv \delta_{\rm D}(\mathbf{q}_1 + \mathbf{q}_2 + \mathbf{q}_3)$ imposes a triangle constraint on the wavevectors. Again assuming $\Delta k \ll k_i$ for all i, $V_B(k_1, k_2, k_3)$ is given by (Scoccimarro et al. 2004)

$$V_B(k_1, k_2, k_3) = \int \mathrm{d}^3 q_1 \int \mathrm{d}^3 q_2 \int \mathrm{d}^3 q_3 \; \delta_{\rm D}(\mathbf{q}_{123}) \tag{3.6}$$

$$\approx 8\pi^2 k_1 k_2 k_3 \Delta k^3 \; . \tag{3.7}$$

To extend these estimators to weak lensing we can use the Limber and flat-sky approximations and integrate over the comoving distance χ. The Limber approximation relates the three-dimensional wavevectors $\mathbf{k}$ to the two-dimensional angular multipoles $\boldsymbol{\ell}$ by $\boldsymbol{\ell} = \mathbf{k}\chi(z)$.

Assuming a flat Universe, for a tomographic survey the convergence field $\kappa_{(i)}$ for the ith tomographic bin is then

$$\kappa_{(i)}(\boldsymbol{\theta}) = \int_0^{\chi_{\rm lim}} \mathrm{d}\chi \; q_{(i)}(\chi) \delta(\chi, \chi\boldsymbol{\theta}) \; , \tag{3.8}$$

where $\boldsymbol{\theta}$ is the angular position on the sky, $\chi_{\rm lim}$ is the maximum comoving distance of the survey, δ is the matter density contrast and $q_{(i)}(\chi)$ is defined as (Joachimi and Bridle 2010)

$$q_{(i)}(\chi) = \frac{3H_0^2\Omega_{\rm m}}{2c^2}\frac{\chi}{a(\chi)}g_{(i)}(\chi)\,. \tag{3.9}$$

Here $a(\chi)$ is the scale factor and $g_{(i)}(\chi)$ is the lensing efficiency:

$$g_{(i)}(\chi) = \int_{\chi}^{\chi_{\rm lim}} {\rm d}\chi'\, p_{(i)}(\chi')\frac{(\chi'-\chi)}{\chi'}\,, \tag{3.10}$$

where $p_{(i)}(\chi)$ is the distribution of galaxies in the ith tomographic bin. This leads to the tomographic weak lensing power spectrum and bispectrum:

$$P^{\kappa}_{(ij)}(\ell) = \int_0^{\chi_{\rm lim}} {\rm d}\chi\; q_{(i)}(\chi)q_{(j)}(\chi)\chi^{-2}P_{\delta}(k;\chi)\,, \tag{3.11}$$

and

$$B^{\kappa}_{(ijk)}(\boldsymbol{\ell}_1,\boldsymbol{\ell}_2,\boldsymbol{\ell}_3) = \int_0^{\chi_{\rm lim}} {\rm d}\chi\; q_{(i)}(\chi)q_{(j)}(\chi)q_{(k)}(\chi)\chi^{-4}B_{\delta}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3;\chi)\,. \tag{3.12}$$

3.2.2 Covariance matrices

We now derive the power spectrum and bispectrum auto-covariances and their cross-covariance. Again we start with the matter power spectrum and bispectrum. We ignore shot noise which would of course exist in any realistic situation. In what follows shot noise can generally be assumed to be Gaussian and simply added to the power spectrum.

The covariances can be expressed as

$$\begin{aligned}
{\rm Cov}[P(k_i),P(k_j)] &= \langle\hat{P}(k_i)\hat{P}(k_j)\rangle - \langle\hat{P}(k_i)\rangle\langle\hat{P}(k_j)\rangle &(3.13)\\
{\rm Cov}[B(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3),B(\mathbf{k}_4,\mathbf{k}_5,\mathbf{k}_6)] &= \langle\hat{B}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)\hat{B}(\mathbf{k}_4,\mathbf{k}_5,\mathbf{k}_6)\rangle &(3.14)\\
&\quad - \langle\hat{B}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)\rangle\langle\hat{B}(\mathbf{k}_4,\mathbf{k}_5,\mathbf{k}_6)\rangle\,.
\end{aligned}$$

These covariances both have the general form

$${\rm Cov}_{\rm full} = {\rm Cov}_{\rm Gauss} + {\rm Cov}_{\rm NG} + {\rm Cov}_{\rm SSC}\,, \tag{3.15}$$

where the subscripts denote 'Gaussian', 'non-Gaussian' and 'supersample covariance'. The power spectrum–bispectrum cross-covariance takes a similar form but has no Gaussian term.

We can use Wick's theorem to express the covariances in terms of correlators of $\tilde{\delta}$, the Fourier transform of the underlying density contrast. Thus the power spectrum covariance has a term $\langle\tilde{\delta}\tilde{\delta}\rangle\langle\tilde{\delta}\tilde{\delta}\rangle$ which is the product of two power spectra and also a term involving the connected 4-point function or trispectrum $\langle\tilde{\delta}\tilde{\delta}\tilde{\delta}\tilde{\delta}\rangle_{\rm c}$. Similarly the bispectrum covariance has terms involving $\langle\tilde{\delta}\tilde{\delta}\rangle\langle\tilde{\delta}\tilde{\delta}\rangle\langle\tilde{\delta}\tilde{\delta}\rangle$, $\langle\tilde{\delta}\tilde{\delta}\tilde{\delta}\rangle_{\rm c}\langle\tilde{\delta}\tilde{\delta}\tilde{\delta}\rangle_{\rm c}$, $\langle\tilde{\delta}\tilde{\delta}\tilde{\delta}\tilde{\delta}\rangle_{\rm c}\langle\tilde{\delta}\tilde{\delta}\rangle$ and $\langle\tilde{\delta}\tilde{\delta}\tilde{\delta}\tilde{\delta}\tilde{\delta}\tilde{\delta}\rangle_{\rm c}$, and the power spectrum–bispectrum cross-covariance involves $\langle\tilde{\delta}\tilde{\delta}\rangle\langle\tilde{\delta}\tilde{\delta}\tilde{\delta}\rangle_{\rm c}$ and $\langle\tilde{\delta}\tilde{\delta}\tilde{\delta}\tilde{\delta}\tilde{\delta}\rangle_{\rm c}$.The terms which depend only on the power spectrum are referred to as Gaussian. If the underlying field is Gaussian then these are the only non-zero parts of the covariance so the matrices are diagonal.

The remaining terms generate non-zero off-diagonal elements. These terms can be split into 'standard' non-Gaussian terms, which arise from coupling between small-scale modes within the survey window, and supersample terms caused by coupling between in-survey modes and long-wavelength modes longer than the survey window dimension (Rimes and Hamilton 2006; Hamilton et al. 2006). This supersample covariance is generated by the 4-point correlator in the power spectrum covariance and the 6-point correlator in the bispectrum covariance. It has been further divided into 'beat coupling' (BC) at large scales, and 'halo sample variance' (HSV) at intermediate to small scales. At large scales there is also another much smaller effect known as linear dilation (Li et al. 2014*a*). This terminology has sometimes been used rather loosely. In particular the label SSC has sometimes been used for only part of the full supersample covariance. Here we use SSC to include all these terms: BC, HSV and dilation (but excluding additional tidal terms discussed for example in Barreira et al. (2018*b*)). We use the label NG to refer to the 'standard' non-Gaussian terms only.

Gaussian terms

The Gaussian terms of the covariance matrices are (Scoccimarro et al. 1999; Sefusatti et al. 2006; Chan and Blot 2017)

$$\mathrm{Cov}^{\rm PP}_{\rm Gauss} = \frac{(2\pi)^3}{V_W}\frac{2}{V_P(k_i)}P(k_i)P(k_j)\delta^{\rm K}_{k_ik_j}\ , \tag{3.16}$$

where V_W is the survey volume, $V_P(k_i)$ is defined in Section 3.2.1 and $\delta^{\rm K}_{k_i k_j}$ is the Kronecker delta.

$$\mathrm{Cov}^{\rm BB}_{\rm Gauss} = \frac{(2\pi)^3}{V_W}\frac{1}{V_B(k_1,k_2,k_3)}P(k_1)P(k_2)P(k_3)\delta^{\rm K}_{k_1k_4}\delta^{\rm K}_{k_2k_5}\delta^{\rm K}_{k_3k_6} + 5 \text{ perms.} \tag{3.17}$$

$$= \frac{(2\pi)^3}{V_W}\frac{s_B}{V_B(k_1,k_2,k_3)}P(k_1)P(k_2)P(k_3)\ , \tag{3.18}$$

where again $V_B(k_1,k_2,k_3)$ is defined in Section 3.2.1, and $s_B = 1, 2, 6$ for general, isosceles and equilateral triangles respectively. The Kronecker deltas in Equation 3.17 show that Gaussian terms of the bispectrum covariance are zero unless the two triangles of the bispectra coincide in shape, size and orientation. Thus the Gaussian terms contribute only to the diagonals of the covariance matrix.

Non-Gaussian terms

The 'standard' non-Gaussian terms are generated by higher-order polyspectra. In the case of the power spectrum covariance the non-Gaussian term arises from the trispectrum, specifically from parallelogram configurations because of the constraint $\mathbf{k}_1 + \mathbf{k}_2 = 0$, leading to (Scoccimarro et al. 1999)

$$\mathrm{Cov}^{\rm PP}_{\rm NG} = \frac{1}{V_W}\int_{k_1}\frac{\mathrm{d}^3q_1}{V_P(k_1)}\int_{k_2}\frac{\mathrm{d}^3q_2}{V_P(k_2)}T(\mathbf{q}_1,-\mathbf{q}_1,\mathbf{q}_2,-\mathbf{q}_2)\ . \tag{3.19}$$

The non-Gaussian terms of the bispectrum covariance come from bispectrum-bispectrum and trispectrum-power spectrum correlations and the six-point function or pentaspectrum, P_6. Thus

$$\mathrm{Cov}^{\rm BB}_{\rm NG} = \mathrm{Cov}^{\rm BB}_{\rm BB} + \mathrm{Cov}^{\rm BB}_{\rm TP} + \mathrm{Cov}^{\rm BB}_{\rm P_6}\ . \tag{3.20}$$

The possible configurations are determined by triangle conditions, giving (Sefusatti et al. 2006)

$$\begin{aligned}\mathrm{Cov}_{\mathrm{NG}}^{\mathrm{BB}} = \delta^{\mathrm{K}}_{q_3q_4} C_{123456} \int_1\int_2\int_3\int_5\int_6 \delta_{\mathrm{D}}(\mathbf{q}_{123})\delta_{\mathrm{D}}(\mathbf{q}_{356}) \\ \times B(\mathbf{q}_1,\mathbf{q}_2,\mathbf{q}_3)B(\mathbf{q}_3,\mathbf{q}_5,\mathbf{q}_6) + 8 \text{ perms.} \\ + \delta^{\mathrm{K}}_{q_3q_4} C_{123456} \int_1\int_2\int_3\int_5\int_6 \delta_{\mathrm{D}}(\mathbf{q}_{123})\delta_{\mathrm{D}}(\mathbf{q}_{356}) \\ \times T(\mathbf{q}_1,\mathbf{q}_2,\mathbf{q}_5,\mathbf{q}_6)P(q_3) + 8 \text{ perms.} \\ + C_{123456} \int_1\int_2\int_3\int_4\int_5\int_6 \delta_{\mathrm{D}}(\mathbf{q}_{123})\delta_{\mathrm{D}}(\mathbf{q}_{456}) \\ \times P_6(\mathbf{q}_1,\mathbf{q}_2,\mathbf{q}_3,\mathbf{q}_4,\mathbf{q}_5,\mathbf{q}_6)\,, \end{aligned} \tag{3.21}$$

where $\int_i$ is shorthand for integration over $\mathbf{k}_i$, and

$$C_{123456} = \frac{(2\pi)^3}{V_W V_B(q_1,q_2,q_3) V_B(q_4,q_5,q_6)}\,. \tag{3.22}$$

The permutations in the first (BB) and second (TP) terms arise from the different possible combinations of two sides, one from each triangle (Kayo et al. 2012). The BB term is zero unless the two triangles have one side in common. The TP term is zero unless the triangles have one side of equal length but opposite direction.

The non-Gaussian part of the cross-covariance between the power spectrum and bispectrum can be derived in a similar way as (Sefusatti et al. 2006; Kayo et al. 2012)

$$\begin{aligned}\mathrm{Cov}_{\mathrm{NG}}^{\mathrm{PB}} = \delta^{\mathrm{K}}_{q_1q_2} \frac{(2\pi)^3}{V_W} \frac{2}{V_P(q_1)V_B(q_2,q_3,q_4)} P(q_1)B(\mathbf{q}_2,\mathbf{q}_3,\mathbf{q}_4) + 2 \text{ perms.} \\ + \frac{(2\pi)^3}{V_W} \frac{2}{V_P(q_1)V_B(q_2,q_3,q_4)} \int_1\int_2\int_3\int_4 \delta_{\mathrm{D}}(\mathbf{q}_{234}) P_5(\mathbf{q}_1,-\mathbf{q}_1,\mathbf{q}_2,\mathbf{q}_3,\mathbf{q}_4)\,, \end{aligned} \tag{3.23}$$

where P_5 is the connected five-point power spectrum.

It is straightforward to extend these results to the weak lensing field using Equations 3.11 and 3.12. We state the results in Appendix 3A.

Supersample covariance

In this section we summarise the origin of supersample covariance and discuss its effect on the power spectrum covariance as a prelude to deriving corresponding results for the bispectrum.

Supersample covariance arises because Fourier modes within the survey volume interact with modes longer than the survey window dimension. These long modes are essentially constant across the survey window so their effect can be equated to a change in the mean density within the survey region (Rimes and Hamilton 2006; Hamilton et al. 2006; Takada and Hu 2013; Li et al. 2014*a*,*b*; Baldauf et al. 2016; Chan et al. 2018; Barreira et al. 2018*b*).

Looked at from this point of view a super-survey mode δ_b has three effects on the local growth of structure which feed through into the power spectrum and bispectrum and their covariances. First the local mean density is rescaled by a factor $(1+\delta_b)$ compared to the global mean. As a result within the survey window the power spectrum is rescaled by $(1+\delta_b)^2$ and the bispectrum is rescaled by $(1+\delta_b)^3$.

Secondly an overdense region expands less quickly than the Universe as a whole does. The local scale factor is reduced so that (Li et al. 2014*a*,*b*; Baldauf et al. 2016)

$$a^3_{\rm local} = \frac{a^3_{\rm global}}{1+\delta_b}\,. \tag{3.24}$$

This rescales local wavenumbers as $k_{\rm local} = (1+\delta_b)^{\frac{1}{3}} k_{\rm global}$.

Finally the supersample mode changes the intrinsic growth inside the survey window so that the growth of small-scale structure is enhanced. From the perspective of the halo model this means that the halo number density is enhanced within the survey region.

Supersample covariance shows up through the response of the power spectrum and bispectrum to all these effects. It can equivalently be considered to be a result of the convolution of the underlying density field with the survey window function which couples in-survey Fourier modes with super-survey modes.

As background to deriving the bispectrum supersample covariance we first summarise the derivation of the supersample contribution to the power spectrum covariance in Takada and Hu (2013). Starting from Equation 3.19 these authors considered the convolution of the trispectrum with the survey window function. This convolution means that the trispectrum takes the form

$$\begin{aligned} T^W(\mathbf{k},\mathbf{k}') = \frac{1}{V_W}\int_k \frac{{\rm d}^3\mathbf{k}_1}{V_P(k)}\int_{k'}\frac{{\rm d}^3\mathbf{k}_2}{V_P(k')}\int\frac{{\rm d}^3\mathbf{q}_1}{(2\pi)^3}\int\frac{{\rm d}^3\mathbf{q}_2}{(2\pi)^3}\int\frac{{\rm d}^3\mathbf{q}_3}{(2\pi)^3}\int\frac{{\rm d}^3\mathbf{q}_4}{(2\pi)^3} \\ \times (2\pi)^3\delta_{\rm D}(\mathbf{q}_{1234})T(\mathbf{k}_1+\mathbf{q}_1,-\mathbf{k}_1+\mathbf{q}_2,\mathbf{k}_2+\mathbf{q}_3,-\mathbf{k}_2+\mathbf{q}_4)\,, \end{aligned} \tag{3.25}$$

where $\mathbf{q}_i$ are long-wavelength super-survey modes with $|\mathbf{q_i}| \ll k$ and $\delta_{\rm D}(\mathbf{q}_{1234})$ is shorthand for $\delta_{\rm D}(\mathbf{q}_1+\mathbf{q}_2+\mathbf{q}_3+\mathbf{q}_4)$. This trispectrum depends on only two wavevectors $\mathbf{k}$ and $\mathbf{k}'$. This follows

from the fact that the trispectrum in the in-survey non-Gaussian covariance involves a parallelogram configuration. The presence of the background mode turns this into a configuration which is nearly but not exactly a parallelogram.

To understand how this configuration gives rise to the supersample covariance we can make the substitutions $\mathbf{k}_1 + \mathbf{q}_1 \to \mathbf{k}_1$, $\mathbf{q}_1 + \mathbf{q}_2 \to \mathbf{q}$. Then, using the Dirac delta function in the trispectrum definition we can write the trispectrum of Equation 3.25 as $T(\mathbf{k}_1, -\mathbf{k}_1 + \mathbf{q}, \mathbf{k}_2, -\mathbf{k}_2 - \mathbf{q})$.

This is a squeezed trispectrum with two very small equal and opposite sides. It can be seen to represent the connection between two power spectra, $P(k_1)$ and $P(k_2)$, through a shared long-wavelength mode $\mathbf{q}$ caused by the background density fluctuation δ_b. An alternative way to think of this is that the squeezed trispectrum expresses the response of the power spectra to the background mode so that

$$T(\mathbf{k}_1, -\mathbf{k}_1 + \mathbf{q}, \mathbf{k}_2, -\mathbf{k}_2 - \mathbf{q}) \approx T(\mathbf{k}_1, -\mathbf{k}_1, \mathbf{k}_2, -\mathbf{k}_2) + \frac{\partial P(k_1)}{\partial \delta_b} \frac{\partial P(k_2)}{\partial \delta_b} P_{\rm L}(q) \ , \quad (3.26)$$

where $P_{\rm L}$ is the linear power spectrum (because the long-wavelength mode is in the linear regime). Inserting this into Equation 3.25 gives

$$\begin{aligned} T^W(\mathbf{k}, \mathbf{k}') = \frac{1}{V_W} \int_k \frac{{\rm d}^3\mathbf{k}_1}{V_P(k)} \int_{k'} \frac{{\rm d}^3\mathbf{k}_2}{V_P(k')} \int \frac{{\rm d}^3\mathbf{q}_1}{(2\pi)^3} \int \frac{{\rm d}^3\mathbf{q}_2}{(2\pi)^3} \int \frac{{\rm d}^3\mathbf{q}_3}{(2\pi)^3} \int \frac{{\rm d}^3\mathbf{q}_4}{(2\pi)^3} \\ \times (2\pi)^3 \delta_{\rm D}(\mathbf{q}_{1234}) \left[T(\mathbf{k}_1, -\mathbf{k}_1, \mathbf{k}_2, -\mathbf{k}_2) + \frac{\partial P(k_1)}{\partial \delta_b} \frac{\partial P(k_2)}{\partial \delta_b} P_{\rm L}(q) \right] \\ = {\rm Cov}^{\rm PP}_{\rm NG} + (\sigma^L_W)^2 \frac{\partial P(k)}{\partial \delta_b} \frac{\partial P(k')}{\partial \delta_b} \end{aligned} \quad (3.27)$$

where $(\sigma^L_W)^2$ is the variance of the background density field within the survey window.

$$(\sigma^L_W)^2 = \frac{1}{V_W^2} \int \frac{{\rm d}^3\mathbf{q}}{(2\pi)^3} |\tilde{W}(\mathbf{q})|^2 P_{\rm L}(q) \ , \quad (3.28)$$

where $\tilde{W}(\mathbf{q})$ is the Fourier transform of the window function defined by Equation 3.1.

Thus the supersample covariance is given by

$${\rm Cov}^{\rm PP}_{\rm SSC} = (\sigma^L_W)^2 \frac{\partial P(k)}{\partial \delta_b} \frac{\partial P(k')}{\partial \delta_b} \ . \quad (3.29)$$

Clearly long-wavelength super-survey modes also have analogous effects on the bispectrum covariance. In fact for completely general estimators $\hat{f}$ and $\hat{g}$ which depend on the halo number count

$$\mathrm{Cov}(\hat{f},\hat{g}) \approx \mathrm{Cov}(\hat{f},\hat{g})_{\mathrm{G}} + \mathrm{Cov}(\hat{f},\hat{g})_{\mathrm{NG}} + (\sigma_W^L)^2 \frac{\partial f}{\partial \delta_b}\frac{\partial g}{\partial \delta_b} \,, \tag{3.30}$$

(Schaan et al. 2014). Thus we can write the bispectrum supersample covariance as

$$\mathrm{Cov}^{\mathrm{BB}}_{\mathrm{SSC}} = (\sigma_W^L)^2 \frac{\partial B(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)}{\partial \delta_b}\frac{\partial B(\mathbf{k}_4,\mathbf{k}_5,\mathbf{k}_6)}{\partial \delta_b} \,, \tag{3.31}$$

and the power spectrum–bispectrum supersample cross-covariance as

$$\mathrm{Cov}^{\mathrm{PB}}_{\mathrm{SSC}} = (\sigma_W^L)^2 \frac{\partial P(k_1)}{\partial \delta_b}\frac{\partial B(\mathbf{k}_2,\mathbf{k}_3,\mathbf{k}_4)}{\partial \delta_b} \,. \tag{3.32}$$

The task now is to determine the response functions. To do this we use the halo model and perturbation theory.

3.2.3 Halo model power spectrum and bispectrum supersample covariance

Halo model expressions for the Gaussian and non-Gaussian parts of the covariance matrices in terms of the power spectrum, bispectrum and higher-level polyspectra are straightforward and are given, for example, in Kayo et al. (2012) and Takada and Hu (2013). Here we discuss the supersample terms. We review the power spectrum supersample covariance in the halo model formalism and derive an expression for the complete bispectrum supersample covariance for equilateral triangles.

The halo model is discussed in more detail in Section 1.3.2 of the introduction to this thesis. It is based round the integrals :

$$I^{\beta}_{\mu}(k_1,k_2,....,k_\mu) \equiv \int \mathrm{d}M \frac{\mathrm{d}n}{\mathrm{d}M}\left(\frac{M}{\bar{\rho}}\right)^{\mu} b_\beta \tilde{u}_M(k_1)\tilde{u}_M(k_2)\ldots\tilde{u}_M(k_\mu) \,. \tag{3.33}$$

Here $M(z)$ is halo mass, $n(M,z)$ is the number density of halos, $\tilde{u}_M(k)$ is the Fourier transform of the halo density profile, μ is the number of points being correlated, and $b_\beta(M)$ is the halo bias. The bias quantifies the β-th order response of the halo mass function $\mathrm{d}n/\mathrm{d}M$ to the long-wavelength mode δ_b (Mo and White 1996; Schmidt et al. 2013):

$$b_\beta(M) = \frac{1}{f}\frac{\partial^\beta f}{\partial \delta_b^\beta}\bigg|_{\delta_b=0} \,, \tag{3.34}$$

where $f(M,z) \equiv \mathrm{d}n/\mathrm{d}M$. We assume linear bias so that $b_0 = 1$, $b_1 = b(M)$ and $b_\beta = 0$ for $\beta > 1$.

The halo model expression for the power spectrum is

$$P(k) = I_2^0(k,k) + [I_1^1(k)]^2 P_{\rm L}(k) \,, \tag{3.35}$$

so

$$\frac{\partial P(k)}{\partial \delta_b} = \frac{\partial I_2^0(k,k)}{\partial \delta_b} + [I_1^1(k)]^2 \frac{\partial P_{\rm L}(k)}{\partial \delta_b} \,, \tag{3.36}$$

where we assume that the one-halo term $I_1^1(k)$ is not affected by the background mode δ_b (Chiang et al. 2014). We further assume that in the presence of δ_b the halo mass function changes from f to $(1+\delta_b)f$ but the halo profile does not change (Schaan et al. 2014), so that

$$\left.\frac{\partial I_2^0(k,k)}{\partial \delta_b}\right|_{\delta_b=0} = \int \mathrm{d}M \left(\frac{M}{\bar{\rho}}\right)^2 \left(\frac{\partial f}{\partial \delta_b} + f\right) \tilde{u}_M(k_1)\tilde{u}_M(k_2) \,. \tag{3.37}$$

From Equation 3.34

$$\left.\frac{\partial f}{\partial \delta_b}\right|_{\delta_b=0} = f b(M) \,. \tag{3.38}$$

Substituting into Equation 3.37 gives

$$\frac{\partial I_2^0(k_1,k_2)}{\partial \delta_b} = \int \mathrm{d}M \frac{\mathrm{d}n}{\mathrm{d}M} \left(\frac{M}{\bar{\rho}}\right)^2 (1 + b(M)) \tilde{u}_M(k_1)\tilde{u}_M(k_2) \tag{3.39}$$

$$= I_2^1(k_1,k_2) \,. \tag{3.40}$$

Finally

$$\frac{\partial P(k)}{\partial \delta_b} = I_2^1(k,k) + [I_1^1(k)]^2 \frac{\partial P_{\rm L}(k)}{\partial \delta_b} \,. \tag{3.41}$$

Thus we need the response of the linear power spectrum to the background fluctuation δ_b. This was derived by Takada and Hu (2013) using perturbation theory and consistency relation arguments. Similar results were also obtained by Li et al. (2014*a*) using a separate universe approach and by Chiang et al. (2014) using the position-dependent power spectrum and integrated bispectrum. All

these authors expressed the power spectrum in the presence of the long-wavelength mode as

$$P_{\rm L}(k|\delta_b) \approx \left(1+\frac{26}{21}\delta_b\right)P_{\rm L}(k) + \delta_b\left(P_{\rm L}(k) - \frac{1}{3}k\frac{\partial P_{\rm L}(k)}{\partial k}\right) \tag{3.42}$$

$$= \left(1+\frac{47}{21}\delta_b\right)P_{\rm L}(k) - \frac{\delta_b}{3}\frac{\partial P_{\rm L}(k)}{\partial \ln k}\,. \tag{3.43}$$

From this we obtain

$$\frac{\partial P_{\rm L}(k|\delta_b)}{\partial \delta_b} = \frac{47}{21}P_{\rm L}(k|\delta_b) - \frac{1}{3}\frac{\partial P_{\rm L}(k)}{\partial \ln k}\,. \tag{3.44}$$

We can restate the final partial derivative and substitute into Equation 3.41 to get

$$\frac{\partial P(k)}{\partial \delta_b} = I_2^1(k,k) + [I_1^1(k)]^2\left[\frac{68}{21} - \frac{1}{3}\frac{\partial \ln k^3(P_{\rm L}(k))}{\partial \ln k}\right]P_{\rm L}(k)\,. \tag{3.45}$$

Thus the halo model expression for the power spectrum supersample covariance is

$$\begin{aligned}\mathrm{Cov}_{\rm SSC}^{\rm PP} = (\sigma_W^L)^2&\left[I_2^1(k_i,k_i) + [I_1^1(k_i)]^2\left[\frac{68}{21} - \frac{1}{3}\frac{\partial \ln k_i^3(P_{\rm L}(k_i))}{\partial \ln k_i}\right]P_{\rm L}(k_i)\right]\\ &\times\left[I_2^1(k_j,k_j) + [I_1^1(k_j)]^2\left[\frac{68}{21} - \frac{1}{3}\frac{\partial \ln k_j^3(P_{\rm L}(k_j))}{\partial \ln k_j}\right]P_{\rm L}(k_j)\right]\end{aligned} \tag{3.46}$$

The $I_2^1(k_i,k_i)I_2^1(k_j,k_j)$ and $(I_1^1(k_i))^2(I_1^1(k_j))^2$ terms correspond respectively to halo sample variance and beat coupling and there is also a BC-HSV cross term. The derivative terms represent linear dilation.

3.2.4 Matter bispectrum response function

In the halo model the bispectrum response in Equation 3.31 is

$$\frac{\partial B(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)}{\partial\delta_b} = \frac{\partial}{\partial\delta_b}\Big(B^{\rm 1h}+B^{\rm 2h}+B^{\rm 3h}\Big) \tag{3.47}$$

$$= \frac{\partial}{\partial\delta_b}\Big(I_3^0(k_1,k_2,k_3)+I_1^1(k_1)I_2^1(k_2,k_3)P(k_1)+\text{ 2 perms.} + I_1^1(k_1)I_1^1(k_2)I_1^1(k_3)B_{\rm PT}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)\Big) \tag{3.48}$$

$$= I_3^1(k_1,k_2,k_3)+I_1^1(k_1)I_2^1(k_2,k_3)\frac{\partial P(k_{23})}{\partial\delta_b}+\text{ 2 perms.} + I_1^1(k_1)I_1^1(k_2)I_1^1(k_3)\frac{\partial B_{\rm PT}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)}{\partial\delta_b} \tag{3.49}$$

$$= I_3^1(k_1,k_2,k_3)+\frac{47}{21}\big[I_1^1(k_1)I_2^1(k_2,k_3)P(k_1)+\text{ 2 perms.}\big] + I_1^1(k_1)I_1^1(k_2)I_1^1(k_3)\frac{\partial B_{\rm PT}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)}{\partial\delta_b}\,, \tag{3.50}$$

where $B_{\rm PT}$ is the perturbation theory bispectrum. To determine the response of $B_{\rm PT}$ to the background fluctuation δ_b we follow the method of Chiang et al. (2014). These authors derived the power spectrum response from the position-dependent power spectrum: the correlation between the power spectrum measured in a sub-volume with length scale L and volume V_L and the mean density contrast at the centre of the sub-volume, $\mathbf{r}$. They showed that this is equivalent to an integrated bispectrum which is straightforward to calculate.

The position-dependent power spectrum is

$$\langle P(\mathbf{k}),\bar\delta(\mathbf{r})\rangle = \frac{1}{V_L^2}\int\frac{{\rm d}^3\mathbf{q}_1}{(2\pi)^3}\int\frac{{\rm d}^3\mathbf{q}_2}{(2\pi)^3}\int\frac{{\rm d}^3\mathbf{q}_3}{(2\pi)^3}\langle\delta(\mathbf{k}-\mathbf{q}_1)\delta(-\mathbf{k}-\mathbf{q}_2)\delta(-\mathbf{q}_3)\rangle \times W_L(\mathbf{q}_1)W_L(\mathbf{q}_2)W_L(\mathbf{q}_3){\rm e}^{-{\rm I}\mathbf{r}\cdot(\mathbf{q}_1+\mathbf{q}_2+\mathbf{q}_3)} \tag{3.51}$$

$$= \frac{1}{V_L^2}\int\frac{{\rm d}^3\mathbf{q}_1}{(2\pi)^3}\frac{{\rm d}^3\mathbf{q}_3}{(2\pi)^3}B(\mathbf{k}-\mathbf{q}_1,-\mathbf{k}+\mathbf{q}_1+\mathbf{q}_3,-\mathbf{q}_3) \tag{3.52}$$

$$\times W_L(\mathbf{q}_1)W_L(\mathbf{q}_2)W_L(\mathbf{q}_3) \tag{3.53}$$

where $\bar\delta(\mathbf{r})$ is the local mean density fluctuation within the sub-volume and $W_L(\mathbf{q})$ is the sub-volume window function. $W_L(\mathbf{q})=1$ inside the sub-volume and zero otherwise. Thus the position-dependent power spectrum can be viewed as an integrated bispectrum which we denote as $iB(\mathbf{k})$.

For modes well within the sub-volume, that is with $q_1, q_3 \ll k$, the dominant contribution to $iB(\mathbf{k})$ comes from the bispectrum in squeezed configurations with one triangle side very much shorter than the other two: $B(\mathbf{k}, -\mathbf{k}+\mathbf{q}, -\mathbf{q})$.

We can further simplify by spherically averaging over the solid angle Ω_k between $\mathbf{k}$ and $\mathbf{q}$ to get the angle-averaged integrated bispectrum

$$iB(k) = \frac{1}{V_L^2} \int \frac{\mathrm{d}^3\mathbf{q}}{(2\pi)^3} |W_L(\mathbf{q})|^2 \int \frac{\mathrm{d}^2\Omega_k}{4\pi} B(\mathbf{k}, -\mathbf{k}+\mathbf{q}, -\mathbf{q}) \; . \tag{3.54}$$

In perturbation theory the bispectrum can be expressed in terms of the power spectrum as

$$\begin{aligned} B(\mathbf{k}, -\mathbf{k}+\mathbf{q}, -\mathbf{q}) = 2[&F_2(\mathbf{k}, -\mathbf{k}+\mathbf{q}) P(k) P(|-\mathbf{k}+\mathbf{q}|) \\ &+ F_2(-\mathbf{k}+\mathbf{q}, -\mathbf{q}) P(|-\mathbf{k}+\mathbf{q}|) P(q) \\ &+ F_2(-\mathbf{q}, \mathbf{k}) P(q) P(k)] \end{aligned} \tag{3.55}$$

As a consequence, to leading order the angle-averaged integrated bispectrum has the form (Chiang et al. 2014)

$$iB(k) \approx \sigma_L^2 f(k) P(k) \; , \tag{3.56}$$

where σ_L^2 is the variance of the density field on the sub-volume scale

$$\sigma_L^2 = \frac{1}{V_L^2} \int \frac{\mathrm{d}^3\mathbf{q}}{(2\pi)^3} W_L^2(q) P(q) \; . \tag{3.57}$$

Thus $f(k)$ is the required response of the power spectrum to the long-wavelength mode.

$$f(k) \equiv \left. \frac{\partial \ln P(k)}{\partial \delta_b} \right|_{\delta_b = 0} \; . \tag{3.58}$$

Chiang et al. (2014) used perturbation theory to derive an expression for the squeezed limit bispectrum in Equation 3.54, and showed that this produced the result for the power spectrum response in Equation 3.44.

We now define the integrated trispectrum $iT(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)$ in an analogous way to the integrated bispectrum and use this to determine the bispectrum response, on the assumption that we can obtain

an expression, analogous to Equation 3.56, of the form

$$iT(k_1, k_2, k_3) \approx \sigma_L^2 g(k_1 k_2, k_3) B(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) \,, \tag{3.59}$$

where now

$$g(k_1 k_2, k_3) = \left. \frac{\partial \ln B_{\rm PT}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)}{\partial \delta_b} \right|_{\delta_b = 0} \,, \tag{3.60}$$

and $iT(k_1, k_2, k_3)$, independent of the directions of the wavevectors, is an angular average of $iT(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)$.

The integrated trispectrum has also been used by Adhikari et al. (2016) in the context of primordial non-Gaussianity and is discussed in detail by Munshi and Coles (2017). It is defined as a position-dependent bispectrum:

$$\begin{aligned} iT(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) &\equiv \langle B(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) \bar{\delta}(\mathbf{r}) \rangle & (3.61) \\ &= \frac{1}{V_L^2} \int \frac{\mathrm{d}^3 \mathbf{q}_1}{(2\pi)^3} \int \frac{\mathrm{d}^3 \mathbf{q}_2}{(2\pi)^3} \int \frac{\mathrm{d}^3 \mathbf{q}_3}{(2\pi)^3} \int \frac{\mathrm{d}^3 \mathbf{q}_4}{(2\pi)^3} \\ &\quad \times \langle \delta(\mathbf{k}_1 - \mathbf{q}_1) \delta(\mathbf{k}_2 - \mathbf{q}_2) \delta(\mathbf{k}_3 - \mathbf{q}_3) \delta(-\mathbf{q}_4) \rangle \\ &\quad \times W_L(\mathbf{q}_1) W_L(\mathbf{q}_2) W_L(\mathbf{q}_3) W_L(\mathbf{q}_4) \mathrm{e}^{-\mathrm{I}\mathbf{r} \cdot (\mathbf{q}_1 + \mathbf{q}_2 + \mathbf{q}_3 + \mathbf{q}_4)} \,. & (3.62) \end{aligned}$$

We can now use the definition of the trispectrum

$$\langle \delta(\mathbf{q}_1) \delta(\mathbf{q}_2) \delta(\mathbf{q}_3) \delta(\mathbf{q}_4) \rangle_c = (2\pi)^3 \delta_{\rm D}(\mathbf{q}_1 + \mathbf{q}_2 + \mathbf{q}_3 + \mathbf{q}_4) T(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \mathbf{q}_4) \tag{3.63}$$

and a change of variables $\mathbf{q}_4 \to \mathbf{q}_3, \mathbf{q}_3 \to -(\mathbf{q}_1 + \mathbf{q}_2 + \mathbf{q}_3)$. Then performing the $\mathbf{q}_4$ integral leads to

$$\begin{aligned} iT(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) &= \frac{1}{V_L^2} \int \frac{\mathrm{d}^3 \mathbf{q}_1}{(2\pi)^3} \int \frac{\mathrm{d}^3 \mathbf{q}_2}{(2\pi)^3} \int \frac{\mathrm{d}^3 \mathbf{q}_3}{(2\pi)^3} \\ &\quad \times T(\mathbf{k}_1 - \mathbf{q}_1, \mathbf{k}_2 - \mathbf{q}_2, \mathbf{k}_3 + \mathbf{q}_1 + \mathbf{q}_2 + \mathbf{q}_3, -\mathbf{q}_3) \\ &\quad \times W_L(\mathbf{q}_2) W_L(-\mathbf{q}_1 - \mathbf{q}_2 - \mathbf{q}_3) W_L(\mathbf{q}_3) \,. & (3.64) \end{aligned}$$

Following the method used for the integrated bispectrum we can remove two of the $\mathbf{q}$ integrals using the property of the binary window function

$$\int \frac{\mathrm{d}^3\mathbf{q}_2}{(2\pi)^3} W_L(-\mathbf{q}_1-\mathbf{q}_2)W_L(\mathbf{q}_2) = W_L(\mathbf{q}_1) \tag{3.65}$$

and then make another change of variables $\mathbf{k}_1 - \mathbf{q}_1 \to \mathbf{k}_1$, $\mathbf{k}_2 - \mathbf{q}_2 \to \mathbf{k}_2$ and $\mathbf{q}_3 = \mathbf{q}$, to arrive at

$$iT(\mathbf{k}_1, \mathbf{k}_2) \approx \frac{1}{V_L^2} \int \frac{\mathrm{d}^3\mathbf{q}}{(2\pi)^3} |W_L(\mathbf{q})|^2 T(\mathbf{k}_1, \mathbf{k}_2, -\mathbf{k}_{12} + \mathbf{q}, -\mathbf{q}) \ , \tag{3.66}$$

where $\mathbf{k}_{12} = \mathbf{k}_1 + \mathbf{k}_2$ and we have used the fact that $\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 = 0$.

We now average this trispectrum over the solid angles Ω_{12} and Ω_{23} between two pairs of wavevectors to obtain $iT(k_1, k_2)$, independent of the directions of the $\mathbf{k}$ vectors.

$$iT(k_1, k_2) = \int \frac{\mathrm{d}^2\Omega_{12}}{4\pi} \int \frac{\mathrm{d}^2\Omega_{13}}{4\pi} iT(\mathbf{k}_1, \mathbf{k}_2) \tag{3.67}$$

$$= \frac{1}{V_L^2} \int \frac{\mathrm{d}^3\mathbf{q}}{(2\pi)^3} |W_L(\mathbf{q})|^2 \int \frac{\mathrm{d}^2\Omega_{12}}{4\pi} T(\mathbf{k}_1, \mathbf{k}_2, -\mathbf{k}_2 + \mathbf{q}, -\mathbf{q}) \ . \tag{3.68}$$

In the last line we have removed one angular integral by fixing the direction of one $\mathbf{k}$ vector. Thus once again, as with the integrated bispectrum, we must evaluate a trispectrum with one wavevector much smaller than the other three. However the derivation above makes some strong assumptions. Strictly speaking the step leading to Equation 3.66 is not legitimate because the squeezed limit of the trispectrum cannot be defined by four sides alone. It also requires a fifth item, either an angle or a diagonal. As a consequence Equation 3.66 is not in general the exact squeezed limit of Equation 3.64 (Adhikari et al. 2016).

Nevertheless we proceed using perturbation theory to derive the squeezed trispectrum which appears in Equation 3.68, specialising to 'equilateral' configurations with three sides of equal length and a fourth much smaller side. This derivation has not explicitly been presented by previous authors.

In perturbation theory the general trispectrum $T_{\rm PT}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4)$ can be expressed as (Bernardeau et al. 2002; Pielorz et al. 2010):

$$T_{\rm PT} = 4T_a + 6T_b \tag{3.69}$$

where T_a is the sum of 12 terms like $F_2(\mathbf{k}_1, -\mathbf{k}_1 - \mathbf{k}_3)F_2(\mathbf{k}_2, \mathbf{k}_1 + \mathbf{k}_3)P_{\rm L}(k_1)P_{\rm L}(k_2)P_{\rm L}|\mathbf{k}_1 + \mathbf{k}_3|)$ and T_b is the sum of 4 terms like $F_3(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)P_{\rm L}(k_1)P_{\rm L}(k_2)P_{\rm L}(k_3)$.

For the trispectrum $T(\mathbf{k}_1, \mathbf{k}_2, -\mathbf{k}_{12}+\mathbf{q}, -\mathbf{q})$ this leads to

$$
\begin{aligned}
T_a(\mathbf{k}_1,&\mathbf{k}_2, -\mathbf{k}_{12}+\mathbf{q}, -\mathbf{q}) = \\
&P_1P_2[P(|-\mathbf{k}_2+\mathbf{q}|)F_2(\mathbf{k}_1, \mathbf{k}_2-\mathbf{q})F_2(\mathbf{k}_2, -\mathbf{k}_2+\mathbf{q}) \\
&\quad + P(|\mathbf{k}_1-\mathbf{q}|)F_2(\mathbf{k}_1, -\mathbf{k}_1+\mathbf{q})F_2(\mathbf{k}_2, \mathbf{k}_1-\mathbf{q})] \\
&+ P_1P(|-\mathbf{k}_{12}+\mathbf{q}|)[P(|\mathbf{k}_1+\mathbf{k}_2|)F_2(\mathbf{k}_1, -\mathbf{k}_{12})F_2(-\mathbf{k}_{12}+\mathbf{q}, \mathbf{k}_{12}) \\
&\quad + P(|\mathbf{k}_1-\mathbf{q}|)F_2(\mathbf{k}_1, -\mathbf{k}_1+\mathbf{q})F_2(-\mathbf{k}_{12}+\mathbf{q}, \mathbf{k}_1-\mathbf{q})] \\
&+ P_1P_{\rm L}(q)[P(k_{12})F_2(\mathbf{k}_1, -\mathbf{k}_{12})F_2(-\mathbf{q}, \mathbf{k}_{12}) \\
&\quad + P(|-\mathbf{k}_2+\mathbf{q}|)F_2(\mathbf{k}_1, \mathbf{k}_2-\mathbf{q})F_2(-\mathbf{q}, -\mathbf{k}_2+\mathbf{q})] \\
&+ P_2P(|-\mathbf{k}_{12}+\mathbf{q}|)[P(k_{12})F_2(\mathbf{k}_2, -\mathbf{k}_{12})F_2(-\mathbf{k}_{12}+\mathbf{q}, \mathbf{k}_{12}) \\
&\quad + P(|\mathbf{k}_2-\mathbf{q}|)F_2(\mathbf{k}_2, -\mathbf{k}_2+\mathbf{q})F_2(-\mathbf{k}_{12}+\mathbf{q}, \mathbf{k}_2-\mathbf{q})] \\
&+ P_2P_{\rm L}(q)[P(k_{12})F_2(\mathbf{k}_2, -\mathbf{k}_{12})F_2(-\mathbf{q}, \mathbf{k}_{12}) \\
&\quad P(|-\mathbf{k}_1+\mathbf{q}|)F_2(\mathbf{k}_2, \mathbf{k}_1-\mathbf{q})F_2(-\mathbf{q}, -\mathbf{k}_1+\mathbf{q})] \\
&+ P(|-\mathbf{k}_{12}+\mathbf{q}|)P_{\rm L}(q)[P(|-\mathbf{k}_2+\mathbf{q}|)F_2(-\mathbf{k}_{12}+\mathbf{q}, \mathbf{k}_2-\mathbf{q})F_2(-\mathbf{q}, -\mathbf{k}_2+\mathbf{q}) \\
&\quad + P(|-\mathbf{k}_1+\mathbf{q}|)F_2(-\mathbf{k}_{12}+\mathbf{q}, \mathbf{k}_1-\mathbf{q})F_2(-\mathbf{q}, -\mathbf{k}_1+\mathbf{q})] \qquad (3.70)
\end{aligned}
$$

and

$$
\begin{aligned}
T_b(\mathbf{k}_1, \mathbf{k}_2, -\mathbf{k}_{12}+\mathbf{q}, -\mathbf{q}) = &P_1P_2P(|-\mathbf{k}_{12}+\mathbf{q}|)F_3(\mathbf{k}_1, \mathbf{k}_2, -\mathbf{k}_{12}+\mathbf{q}) \\
&+ P_1P_2P_{\rm L}(q)F_3(\mathbf{k}_1, \mathbf{k}_2, -\mathbf{q}) \\
&+ P_1P(|-\mathbf{k}_{12}+\mathbf{q}|)P_{\rm L}(q)F_3(\mathbf{k}_1, -\mathbf{k}_{12}+\mathbf{q}, -\mathbf{q}) \\
&+ P_2P(|-\mathbf{k}_{12}+\mathbf{q}|)P_{\rm L}(q)F_3(\mathbf{k}_2, -\mathbf{k}_{12}+\mathbf{q}, -\mathbf{q})\,. \qquad (3.71)
\end{aligned}
$$

Here $\mathbf{k}_{12} = \mathbf{k}_1 + \mathbf{k}_2 = -\mathbf{k}_3$, P_i is shorthand for $P_{\rm L}(k_i)$ and $P(k_{12}) = P_{\rm L}(|\mathbf{k}_1 + \mathbf{k}_2|)$.

We use the computer algebra package Mathematica[5] to evaluate these expressions, and then specialise to configurations with three sides of equal length k and one side of length q, retaining only leading order terms in q/k. Further details are given in Appendix 3B. Our final result, ignoring

[5] https://www.wolfram.com/mathematica

terms which vanish as $q \to 0$, is

$$T_{\rm PT}^{\rm equi} = P(q)P_{\rm L}(k)\left[\left(\frac{2055}{98} + \frac{1354(\mathbf{k}\cdot\mathbf{q})^2}{98k^2q^2}\right)P_{\rm L}(k) + \left(\frac{327}{147k} - \frac{2268(\mathbf{k}\cdot\mathbf{q})^2}{147k^3q^2}\right)\frac{\partial P_{\rm L}(k)}{\partial k}\right]. \tag{3.72}$$

We substitute this into Equation 3.68 and take the average over the solid angle Ω between k and q. The angular average of $(\mathbf{k}\cdot\mathbf{q})^2/k^2q^2$ is 1/3 and so we get

$$iT^{\rm equi} = \frac{1}{V_L^2}\int\frac{{\rm d}^3\mathbf{q}}{(2\pi)^3}|W_L(\mathbf{q})|^2\int\frac{{\rm d}^2\Omega}{4\pi}T_{\rm PT}^{\rm equi} \tag{3.73}$$

$$= \sigma_L^2 P_{\rm L}(k)^2\left[\frac{7519}{294} - \frac{143}{49}\frac{\partial\ln P_{\rm L}(k)}{\partial\ln k}\right]. \tag{3.74}$$

Then from Equations 3.50 and 3.66 the bispectrum response is:

$$\frac{\partial B^{\rm equi}}{\partial\delta_b} = I_3^1(k,k,k) + \frac{47}{7}\left[I_1^1(k)I_2^1(k,k)P_{\rm L}(k)\right] + I_1^1(k)I_1^1(k)I_1^1(k)\left[\frac{7519}{294} - \frac{143}{49}\frac{\partial\ln P_{\rm L}(k)}{\partial\ln k}\right]. \tag{3.75}$$

Figure 3.1 shows the shapes of our calculated response functions at $z = 0$ using Equation 3.45 for the matter power spectrum and Equation 3.75 for equilateral matter bispectrum configurations. The total response functions are shown in black (excluding the dilation terms which are small except at large scales). We also separately show the one-halo, two-halo and three-halo terms which make up the totals. As would be expected, the one-halo terms dominate at small scales where correlations are between different points in a single halo. At large scales points in different halos are being correlated so that the power spectrum response is dominated by the two-halo term and the bispectrum response by the three-halo term.

Substituting for the bispectrum response in Equation 3.31 gives the bispectrum supersample covariance. The power spectrum–bispectrum supersample covariance can be derived in a similar way from the power spectrum and bispectrum response functions.

In Section 3.2.6 we extend these results to the components of the weak lensing supersample covariance.

3.2.5 Other derivations of the matter bispectrum response function

Several recent authors have published analysis relevant to the bispectrum response function. Adhikari et al. (2016) used the integrated trispectrum method to constrain bias parameters and the amplitude of primordial trispectra, Chan et al. (2018) used perturbation theory to derive the response function, and the bispectrum response has been considered by Kehagias et al. (2014), Valageas (2014), Wagner et al. (2015) and Barreira et al. (2018*b*). These works have used differing methods and assumptions and have arrived at different answers. This may not be surprising given the complexity of three-point statistics and the need to approximate to make the analysis tractable.

As an intermediate step in their analysis Adhikari et al. (2016) derived the late-time angle-averaged integrated trispectrum for equilateral triangles as

$$iT(k) = P_{\rm L}(k)^2 P(q) \left[\frac{579}{98} - \frac{8}{7} \frac{\partial \ln P_{\rm L}(k)}{\partial \ln k} \right] . \tag{3.76}$$

This is significantly different from our result. It is difficult to explore why the difference arises because Adhikari et al. (2016) give no detail of their derivation which was also generated by Mathematica (private communication).

Chan et al. (2018) used perturbation theory in a thorough analysis of the matter bispectrum response and arrived at

$$\left. \frac{\partial B(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3 | \delta_b)}{\partial \delta_b} \right|_{\delta_b = 0} = \frac{433}{126} B_{\rm PT}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) + \frac{5}{126} B_{\rm G}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) - \frac{1}{3} \sum_{i=1}^{3} \frac{\partial B_{\rm PT}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)}{\partial \ln k_i} , \tag{3.77}$$

where $B_{\rm PT}$ is the tree-level matter bispectrum and $B_{\rm G}$ is identical to $B_{\rm PT}$ but with the F_2 kernels replaced by G_2. We show in Appendix 3C that for equilateral triangles this becomes

$$\frac{\partial B_{\rm PT}^{\rm equi}}{\partial \delta_b} = \left[\frac{2623}{98} - \frac{36}{7} \frac{\partial \ln P(k)}{\partial \ln k} \right] P(k)^2 . \tag{3.78}$$

This is is close to, but not identical with our result. Once again it is difficult to unearth the source of the discrepancy, although it may be partly due to the assumptions we made when taking the angular average of the integrated trispectrum (Equations 3.64 and 3.66). Certainly with these assumptions we would not expect to obtain exactly the result of Equation 3.77. Throughout the remainder of this chapter we have chosen to use Equation 3.77 since this is the more complete and general result

(even though we restrict our analysis to equilateral triangles). In the next section we show that using Equation 3.74 instead (or indeed Equation 3.76) would make only a very small difference to the weak lensing supersample covariance.

3.2.6 Weak lensing power spectrum and bispectrum covariance

We can use Equations 3.11 and 3.12 to relate the lensing power spectrum and bispectrum to their matter counterparts. Appendix 3A gives details of the resulting expressions. To illustrate the dominant terms of the resulting weak lensing covariance matrices, Figure 3.2 shows the Gaussian, non-Gaussian and supersample terms calculated for a single redshift bin for a 15,000 deg^2 survey. In the case of the bispectrum we show results for equilateral configurations only.

Figure 3.3 splits the supersample terms of the equilateral-triangle bispectrum covariance and the power spectrum-bispectrum cross-covariance into their one-halo, two-halo and three-halo components. (For the cross-covariance we retain both the one-halo and two-halo terms for the power spectrum but vary the bispectrum terms). In both the bispectrum covariance and the cross-covariance the one-halo term is dominant, especially at small scales, and for the cross-covariance it is always the largest term. In the rest of this chapter we always use all three terms, but in fact the dominance of the one-halo terms means that the other terms could safely be dropped. Moreover the fact that the three-halo component is sub-dominant except at very large scales means that the discrepancy between our calculation of the bispectrum response (Equation 3.74) and the response derived from the formula in Chan et al. (2018) (Equation 3.77 (which we have used in later analysis) is immaterial. This is shown in Figure 3.4 which compares the three-halo contributions and also the total supersample covariance using each of the three derivations of the matter bispectrum response discussed above (this work, Chan et al. (2018) and Adhikari et al. (2016)). Even using a very different result for the response function makes almost no difference to the total weak lensing supersample covariance at the scales of interest for weak lensing. An obvious corollary is that the total covariance shown in Figure 3.2 is also hardly affected.

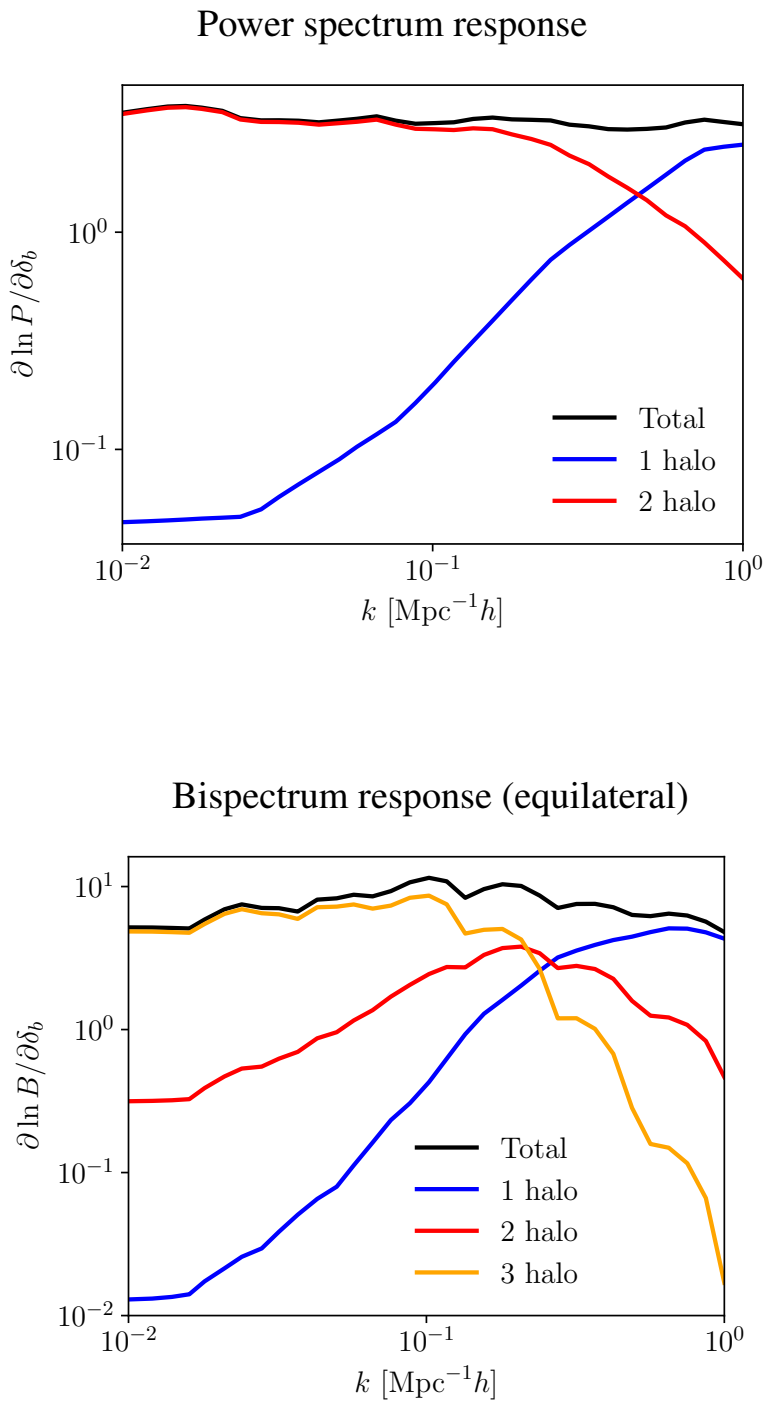

Figure 3.1: Responses of the linear matter power spectrum (top) and tree-level matter equilateral bispectrum (bottom) to a long-wavelength super-survey mode δ_b. Shown are the individual halo model terms and the total response at $z = 0$. The dilation terms, which are sub-dominant, are not included.

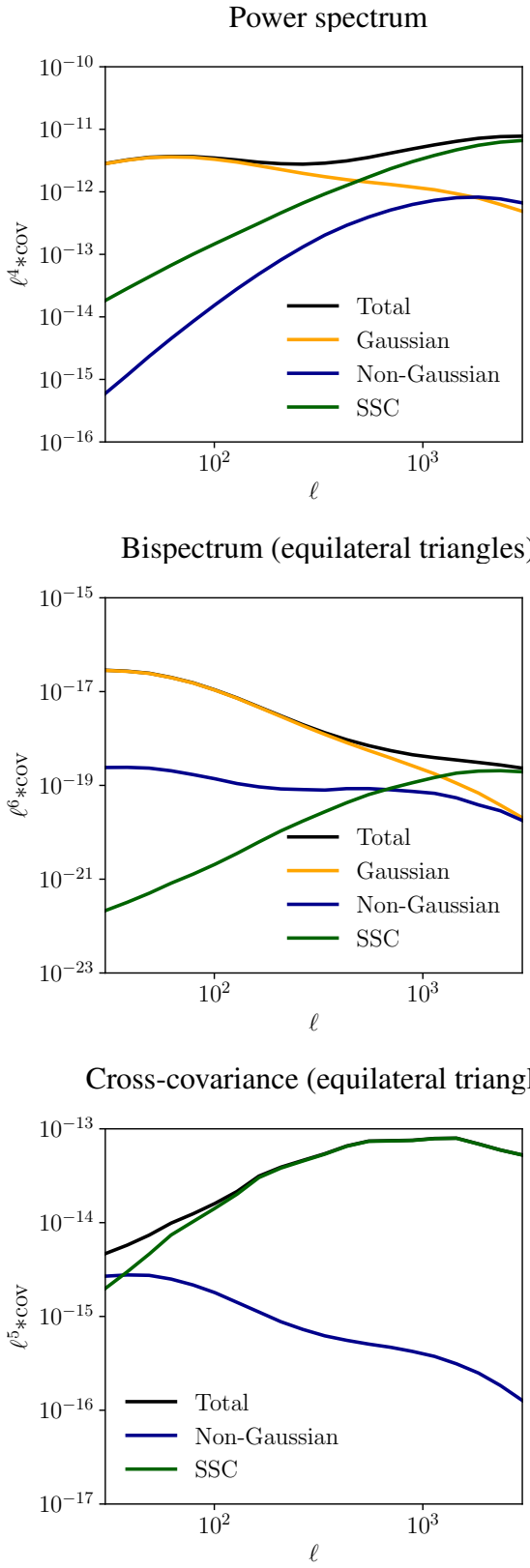

Figure 3.2: Diagonal terms of the weak lensing power spectrum and bispectrum covariance matrices and their cross-covariance, calculated for a single redshift bin at $z = 0.2$. The Gaussian (including shot noise), non-Gaussian and supersample terms are shown. The bispectrum covariance and cross-covariance include equilateral triangles only.

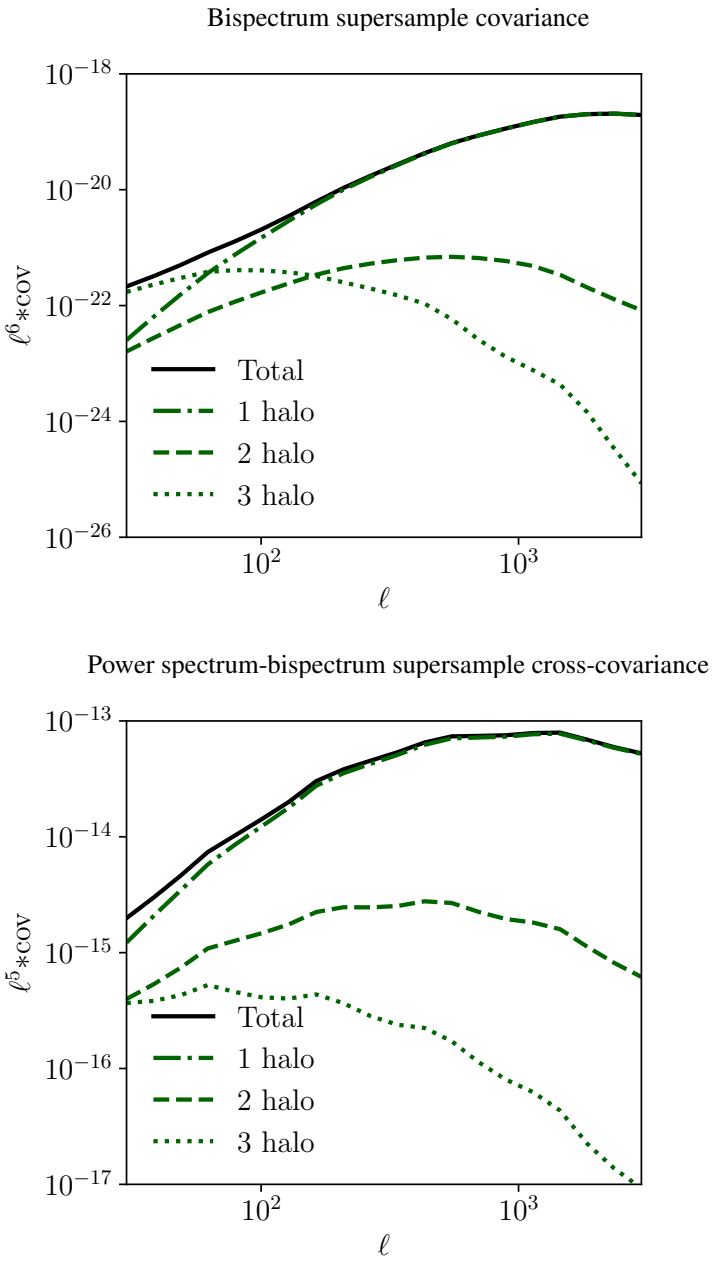

Figure 3.3: Individual terms of the weak lensing supersample covariance, calculated for a single redshift bin at $z = 0.2$. *Top*: Bispectrum supersample covariance. *Bottom*: Power spectrum-bispectrum supersample cross-covariance. The power spectrum includes both the one-halo and two-halo terms.

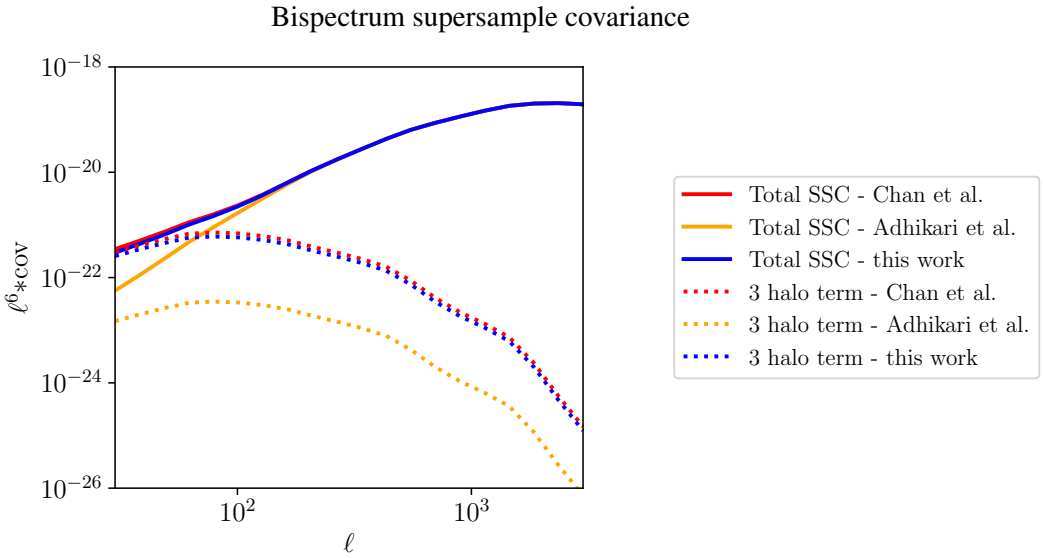

Figure 3.4: The impact on the weak lensing bispectrum supersample covariance of alternative expressions for the matter bispectrum response: Chan et al. (2018) (red), Adhikari et al. (2016) (orange) and our Equation 3.74 (blue), calculated for equilateral triangles and a single redshift bin at $z = 0$. Dotted lines indicate the three-halo term of the weak lensing bispectrum supersample covariance and solid lines show the total supersample covariance.

3.3 Statistical uncertainties

In this section we consider statistical errors only and therefore assume perfect knowledge of, for example, redshifts and galaxy shapes. This approach is consistent with earlier discussions of the weak lensing bispectrum, in particular in Takada and Jain (2004), Kayo et al. (2012), Kayo and Takada (2013), Sato and Nishimichi (2013) and Rizzato et al. (2019). The first four of these considered only relatively small surveys, with areas ranging from 25 to 1,500 square degrees. In keeping with this they either used no tomography or only 3 or 4 redshift bins. Only Rizzato et al. (2019) discussed a survey as large as *Euclid* with 10 tomographic bins, but presented only signal-to-noise forecasts and not Fisher matrix analysis.

Some broad and consistent conclusions arose from all these studies: in particular the importance of including shot noise, the need to include supersample covariance in the power spectrum and bispectrum covariances (other studies had considered the Gaussian covariance only), and the value of a larger survey and many redshift bins. Finally, all agreed that including the bispectrum as well as the power spectrum could make a worthwhile improvement to parameter constraints. For example Kayo et al. (2012) estimated a 20-40% improvement in the signal to noise ratio, Kayo and Takada (2013) forecast a 60% improvement in the dark energy figure of merit (which measures how well dark energy parameters are constrained), and Rizzato et al. (2019) forecast an improvement in the signal to noise ratio of around 10%. (The differences in the results can be attributed to different survey specifications and tomographic set-ups). Kayo et al. (2012) and Sato and Nishimichi (2013) additionally compared their analytical results to results from simulations and found good agreement.

It might appear counter-intuitive that supersample covariance is important for a survey like *Euclid* which covers a large fraction of the sky so that there are relatively few super-survey modes. There are two reasons why this intuition may be wrong, especially for a tomographic weak lensing survey. Firstly, in a tomographic survey only a fraction of the survey volume is available at low redshifts, so many Fourier modes are larger than the effective survey area for a given tomographic bin. Secondly, the supersample covariance depends on the survey area in a different way from the Gaussian and 'standard' components which both scale as the inverse of the survey area. In contrast the (matter) supersample covariance depends on the variance of the background mode within the survey footprint (Equation 3.28). Thus supersample covariance does not decrease in a straightforward way as the survey area increases.

3.3.1 Survey and modelling assumptions

We base our survey assumptions on requirements for *Euclid* specified in Laureijs et al. (2011). We assume a survey with area 15,000 deg^2, total galaxy density 30 arcmin^{-2} and redshift range $z_{\rm min} = 0.2$ to $z_{\rm max} = 2.0$. We allow for up to 10 redshift bins with overall redshift distribution given by

$$p(z) \propto z^{\alpha} \exp\left(- \left(\frac{z}{z_0} \right)^{\beta} \right), \tag{3.79}$$

with $\alpha = 2.0$, $\beta = 1.5$, $z_0 = z_{\rm med}/\sqrt{2}$, $z_{\rm med} = 0.8$. We assume sources are equally distributed between bins and have a Gaussian distribution within each bin. We use 20 angular bins equally logarithmically spaced from $\ell_{\rm min} = 30$ to $\ell_{\rm max} = 3000$. This is a conservative $\ell_{\rm max}$, which avoids non-linear scales where the matter bispectrum in particular is not well understood. For many *Euclid* forecasts $\ell_{\rm max}$ as high as 5000 is used (Laureijs et al. 2011). For ease of computation we consider only triangles which are equilateral in the angular multipole ℓ within angular and redshift bin widths.

We model the non-linear matter power spectrum with the fitting formula from Takahashi et al. (2012) and the matter bispectrum with the formula from Gil-Marín et al. (2012), which was the most recent and accurate available at the time this work was done. The fitting range of this formula is comparatively narrow and we note that Takahashi et al. (2019)) recently derived an improved formula which is accurate further into the nonlinear regime and which explicitly takes account of baryonic effects on the bispectrum.

For the halo model we assume an NFW halo matter density profile, using the mass-concentration relation given in Duffy et al. (2008). We model the halo mass function using the formula in Tinker et al. (2008).

We employ the transfer function from Eisenstein and Hu (1998).

We assume a spatially flat ΛCDM model and consider cosmological parameters with fiducial values $\Omega_{\rm b} = 0.05$, $\Omega_{\rm m} = 0.27$, $\Omega_{\Lambda} = 0.728$, $\sigma_8 = 0.81$, $h = 0.71$, $n_{\rm s} = 0.96$.

To calculate the power spectrum and bispectrum supersample covariance we need to model the survey window function which determines the interaction between in-survey and super-survey modes (see Appendix 3A). For example for the tomographic power spectrum supersample

covariance:

$$\mathrm{Cov}[P^{\kappa}_{ij}(\ell_1), P^{\kappa}_{i'j'}(\ell_2)]_{\mathrm{SSC}} = \frac{1}{\Omega_{\mathrm{s}}^2}\int \mathrm{d}\chi\, q_{(i)}(\chi) q_{(j)}(\chi) q_{(i')}(\chi) q_{(j')}(\chi)\chi^{-6} \times \frac{\partial P_\delta(\ell_1/\chi;\chi)}{\partial \delta_b}\frac{\partial P_\delta(\ell_2/\chi;\chi)}{\partial \delta_b}\int \frac{\mathrm{d}^2\boldsymbol{\ell}}{(2\pi)^2}|\tilde{W}(\boldsymbol{\ell})|^2 P_{\mathrm{L}}(\ell/\chi;\chi)\,, \tag{3.80}$$

where Ω_{s} is the survey area, $q_{(i)}(\chi)$ is the lensing weight function in redshift bin i (Equation 3.9), $\tilde{W}$ is the Fourier transform of the survey mask, P_{L} is the linear matter power spectrum, and we assume the Limber and flat-sky approximations.

We make the simple assumption of a single square survey patch so $\tilde{W}(\boldsymbol{\ell};\chi)$ is evaluated as

$$\tilde{W}(\boldsymbol{\ell};\chi) = \Omega_{\mathrm{s}}\,\mathrm{sinc}\left(\frac{\ell_x}{2}\sqrt{\Omega_{\mathrm{s}}}\right)\mathrm{sinc}\left(\frac{\ell_y}{2}\sqrt{\Omega_{\mathrm{s}}}\right)\,, \tag{3.81}$$

where ℓ_x and ℓ_y are the Cartesian components of the vector $\boldsymbol{\ell}$ and $\mathrm{sinc}(x) = \sin(x)/x$. The right hand side of Equation 3.81 implicitly depends on redshift and the integral over $\boldsymbol{\ell}$ in Equation 3.80 is evaluated separately for each tomographic bin.

3.3.2 Fisher matrix analysis

We quantify the improvement in parameter constraints in two ways. Firstly we calculate the Fisher matrix which was introduced in Section 1.3.4 of this thesis. For the power spectrum this is defined (in simplified notation) by

$$F_{\alpha\beta} = \frac{\partial \mathbf{P}}{\partial p_\alpha}\mathrm{Cov}_{\mathrm{P}}^{-1}\frac{\partial \mathbf{P}}{\partial p_\beta}\,, \tag{3.82}$$

where $\mathbf{P}$ is the power spectrum data vector, $\mathrm{Cov}_{\mathrm{P}}$ is the power spectrum covariance, and p_α and p_β are cosmological parameters. In detail this matrix multiplication is a sum over all combinations of angular frequencies and tomographic bins. Similar definitions apply to the bispectrum and to the power spectrum and bispectrum combined. The Fisher matrix measures how peaked the posterior is around its maximum likelihood value and thus quantifies the information content (Tegmark et al. 1997). Throughout we show ellipses corresponding to 1σ credible regions.

An alternative way to quantify the improvement in parameter constraints is to use the inverse of the area of the Fisher ellipse as a figure of merit (FoM) as defined by the Dark Energy Task Force

(Albrecht et al. 2006). In the plane of the parameters p_α and p_β this is

$$\mathrm{FoM}_{\alpha\beta} = \left[(\mathbf{F}^{-1})_{\alpha\alpha}(\mathbf{F}^{-1})_{\beta\beta} - ((\mathbf{F}^{-1})_{\alpha\beta})^2\right]^{-1/2} . \tag{3.83}$$

We focus on figures of merit in the $\Omega_\mathrm{m} - \sigma_8$ and $w_0 - w_a$ planes, since this is where we might hope that the bispectrum can do most to break degeneracies.

Constraining power

Because of the complexity of the bispectrum covariance it has sometimes been simplified by omitting some or all of its constituent terms. In this section we consider the effect on constraining power of shape noise, supersample covariance and the 'standard' non-Gaussian covariance term and hence determine which of these should be included in the covariance matrix and which can be dropped.

In Figure 3.5 we show parameter constraints obtained with the Gaussian power spectrum covariance only (blue) and with the supersample covariance as well (red). For illustration this figure is based two tomographic bins only; similar effects would be seen with more bins. The top panel does not include shape noise and shows that the supersample term significantly increases the statistical uncertainty compared with using the Gaussian covariance only. This confirms previous results for example in Sato and Nishimichi (2013); Barreira et al. (2018*b*). Ignoring the supersample term would significantly underestimate parameter uncertainties.

In the bottom panel we add to the power spectrum shape noise of the form $\sigma_\epsilon^2/2\bar{n}_i$ where σ_ϵ^2 is the total intrinsic ellipticity dispersion and $\bar{n}_i$ is the galaxy number density in redshift bin i. With our *Euclid*-like survey parameters the inclusion of shape noise can increase the Gaussian errors significantly especially for the Ω_m–σ_8 ellipse. Similar results are shown in Sato and Nishimichi (2013) for a smaller survey. In the remainder of this chapter we always include Gaussian shape noise.

Figure 3.6 additionally includes the full non-Gaussian terms of the covariance and shows these make little difference to parameter constraints over and above the constraints obtained with the Gaussian and supersample terms only. Illustrative results are shown using five tomographic bins for the power spectrum and a single bin for the bispectrum. Varying the number of bins does not affect this conclusion which is consistent with results in Rizzato et al. (2019) and Barreira (2019) who found the non-Gaussian term to be sub-dominant. This result can also be expected from examining the relative sizes of the non-Gaussian terms shown in Figure 3.2. In the light of this we use only the Gaussian and supersample terms in the rest of this chapter.

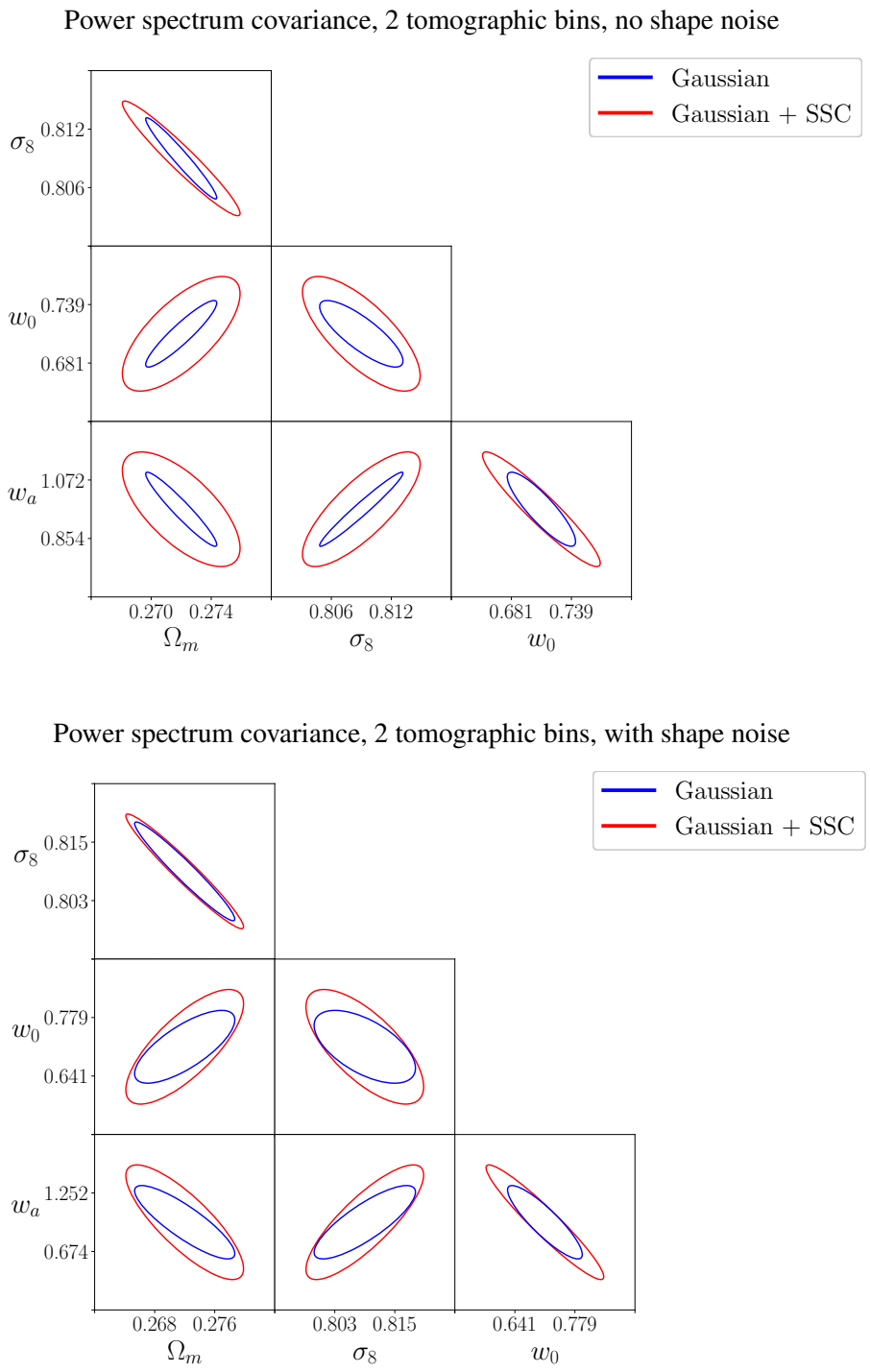

Figure 3.5: Effect of shape noise - power spectrum. Parameter constraints from weak lensing power spectrum using two tomographic bins between $z = 0.2$ and $z = 2.0$. Constraints using only the Gaussian covariance are shown in blue and using Gaussian plus supersample covariance matrices are shown in red. *Top*: Without shape noise. The supersample term increases most parameter uncertainties. *Bottom*: With Gaussian shape noise. shape noise reduces the constraining power of the Gaussian covariance.

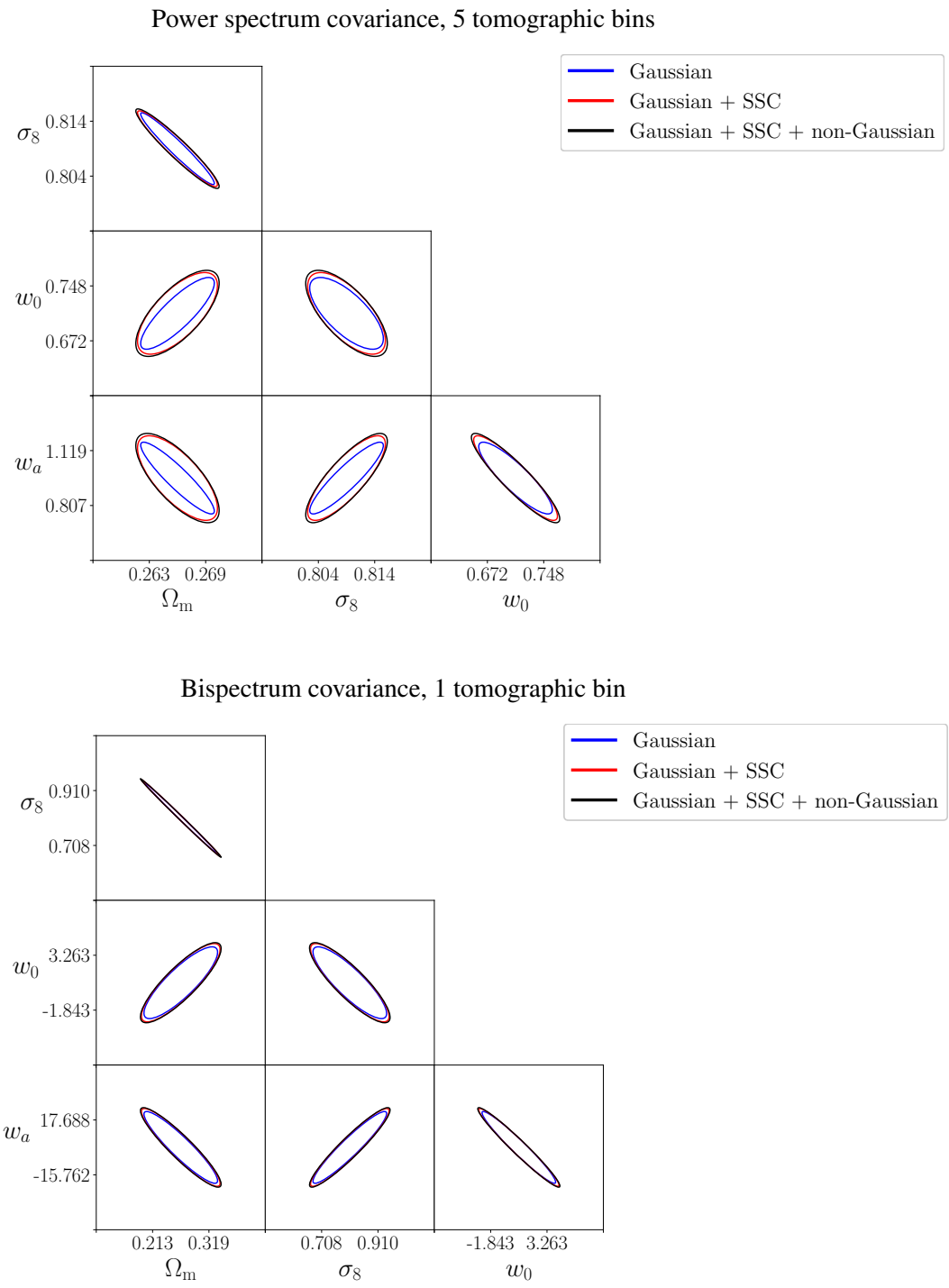

Figure 3.6: Effect of including non-Gaussian covariance. Parameter constraints from weak lensing power spectrum and bispectrum. Constraints obtained with Gaussian only (red) are compared with constraints using Gaussian + SSC (blue) and Gaussian + SSC + non-Gaussian (orange). Shape noise is included. *Top*: Power spectrum - five tomographic bins between $z = 0.2$ and $z = 2.0$. *Bottom*: Bispectrum - one tomographic bin only. (The blue and black ellipses effectively coincide in this case.)

Effect of tomography

To investigate the effect of tomography we use the $\Omega_{\rm m} - \sigma_8$ and $w_0 - w_a$ figures of merit defined in Section 3.3.2.

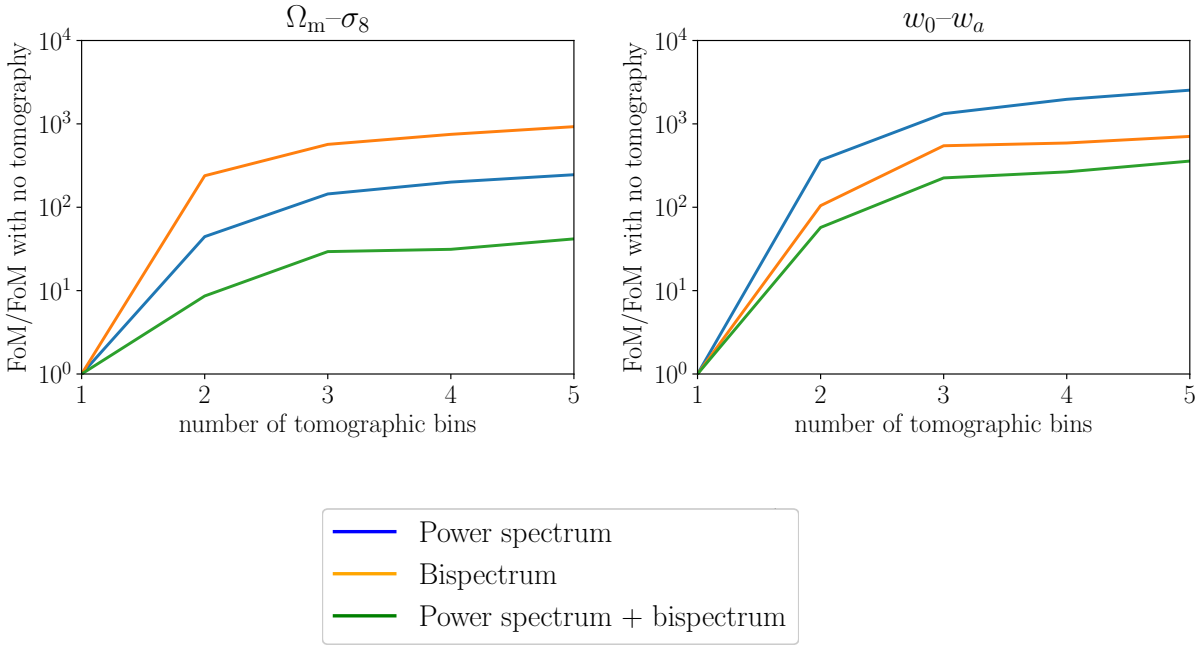

Figure 3.7: Improvement in figures of merit from using more tomographic bins. We plot the figure of merit relative to the figure of merit with no tomography. Increasing the number of bins used for the power spectrum and bispectrum simultaneously yields large improvements in the figures of merit. *Left*: $\Omega_{\rm m}$–σ_8. *Right*: $w_0 - w_a$.

Figure 3.7 shows how increasing the number of tomographic bins from one to five improves the figure of merit when using only the power spectrum, only the bispectrum and both together. In all cases tomography makes difference of several orders of magnitude to the figures of merit. However again the improvement levels off after 3-5 bins, a finding also noted by Rizzato et al. (2019). For the power spectrum covariance we found that there is little further improvement from using more than five bins. This result is already well-established, for example in Ma et al. (2006) and Joachimi and Bridle (2010).

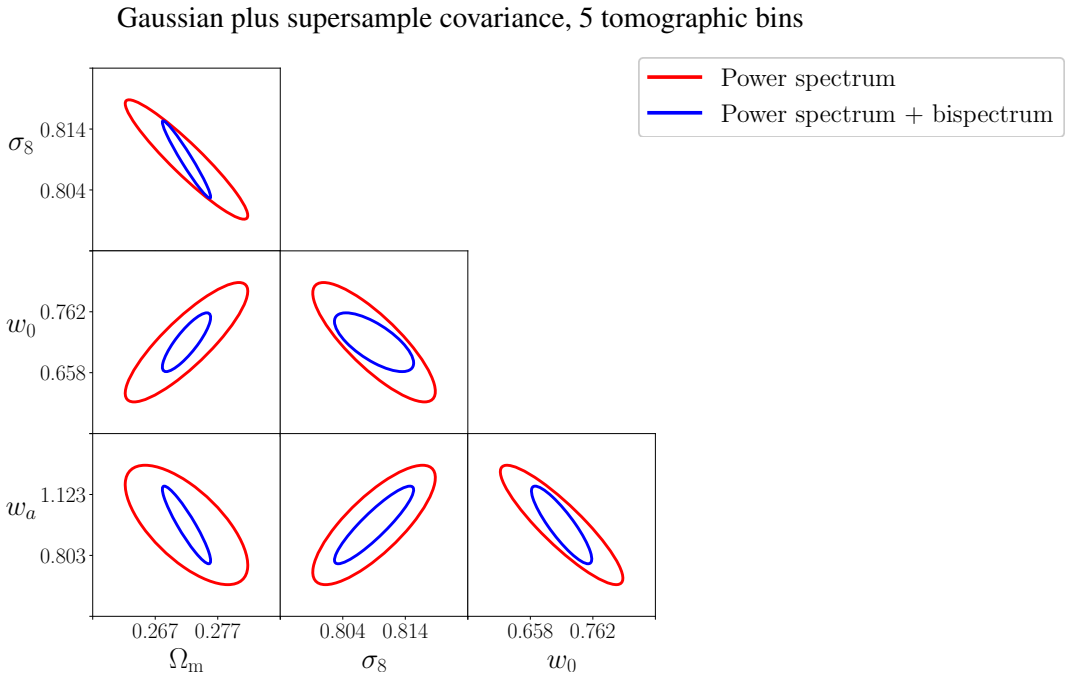

Figure 3.8: Effect of including bispectrum. Parameter constraints from weak lensing power spectrum and power spectrum plus bispectrum. Constraints using only the power spectrum are shown in red and using the power spectrum plus the bispectrum are shown in blue. Five tomographic bins between $z = 0.2$ and $z = 2.0$ are used. Gaussian shape noise is included.

Additional information provided by the bispectrum

Figure 3.8 shows how parameter constraints are improved when the combined power spectrum-bispectrum covariance is used. The improvement is quantified in Table 3.1 which shows how the inclusion of the bispectrum covariance reduces errors on individual parameters. This shows the marginalised error on each parameter using the power spectrum covariance only and also with the combined power spectrum and bispectrum. With five tomographic bins including the bispectrum can reduce errors by up to a third.

Given the rapid increase in size of the bispectrum covariance as the number of redshift bins increases it is interesting to see whether a worthwhile improvement in the figures of merit can be obtained if the power spectrum is based on many bins but fewer bins are used for the bispectrum, and whether this is a better strategy than using a moderate number of bins for both. In Figure 3.9 we fix the number of power spectrum bins at ten and show the figures of merit as a function of the number of bispectrum bins, compared with benchmark values obtained with five power spectrum and five bispectrum bins. It is clearly better, and a more efficient use of computational resources, to

Marginalised statistical errors

	$\Omega_{\rm m}$	σ_8	w_0	w_a
Power spectrum	0.006	0.006	0.067	0.207
Bispectrum	0.019	0.048	0.419	1.64
Power spectrum + bispectrum	0.004	0.005	0.044	0.167
% improvement from bispectrum	34	27	34	19

Table 3.1: Improvement in statistical errors from including the bispectrum covariance as well as the power spectrum. Results are for a *Euclid*-like survey with 5 tomographic bins.

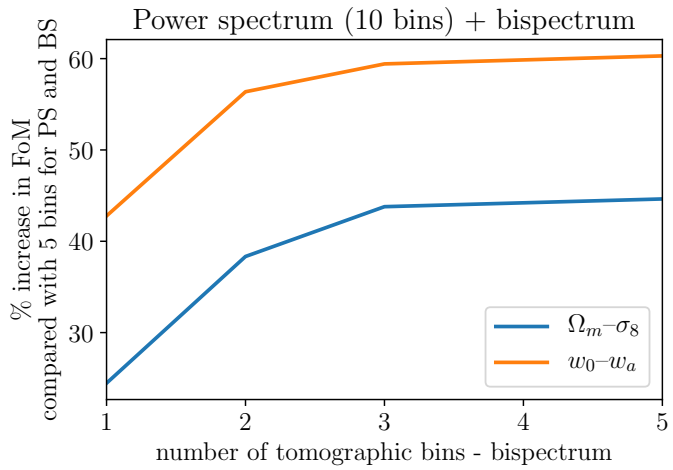

Figure 3.9: We compare the figures of merit obtained with ten power spectrum tomographic bins and increasing numbers of bispectrum bins with the values obtained with five bins for both the power spectrum and bispectrum.

use many power spectrum bins. We find that with ten power spectrum bins, using only two redshift bins for the bispectrum produces an improvement of around 10% in the figures of merit compared with no tomography. The improvement with five bispectrum bins is about 15%. This suggests that an effective strategy is to use as many power spectrum bins as feasible but only two or three bins for the bispectrum. Little is gained by using more bins for the bispectrum, at great computational expense.

3.3.3 Conclusions - statistical uncertainties

Our results confirm previous findings about the information content of the weak lensing bispectrum, specifically for a *Euclid*-like survey, and set the stage for the exploration of systematic errors in

Section 3.4. We have shown that it is important to include both shape noise and the supersample covariance in the covariance calculation. Without these, parameter uncertainties can be severely underestimated. Although we used the one-halo, two-halo and three-halo terms of the bispectrum supersample covariance, in fact we showed that the one-halo term is dominant and it is safe to neglect the other terms. Furthermore, the 'standard' non-Gaussian covariance makes little difference to parameter constraints once the supersample terms are included. For these reasons all subsequent analysis in this chapter is based on the Gaussian plus supersample terms only and includes shape noise.

Next we showed that, for both the power spectrum covariance and the bispectrum covariance, tomography greatly reduces statistical uncertainties. However in the absence of systematic errors this improvement levels off once more than about five tomographic bins are used. We also showed that, for a *Euclid*-like survey with five redshift bins, including the bispectrum covariance can decrease statistical errors in $\Omega_{\rm m}$, σ_8 and the dark energy parameters by around 20-30%. Moreover if enough tomographic bins are used to estimate the power spectrum covariance, most of the gain from the bispectrum can be achieved even if only two or three bins are used to estimate the bispectrum covariance. This is a useful result because the size and complexity of the bispectrum covariance and the computational cost of estimating it increase rapidly as the number of tomographic bins increases. Overall these results suggest that if systematic errors can be sufficiently well controlled the bispectrum can provide useful information over and above the power spectrum.

3.4 Control of systematics

The results in the previous section suggest that in the absence of systematic errors the weak lensing bispectrum could provide worthwhile improvements to parameter constraints. However in reality weak lensing is affected by many hard-to-control systematic uncertainties. The effect of systematic errors on the weak lensing power spectrum has been studied extensively and there is a significant literature presenting general methods of estimating and controlling weak lensing systematics or discussing specific types of uncertainty, for example Huterer and Takada (2005); Huterer et al. (2006); Ma et al. (2006); Bridle and King (2007); Hearin et al. (2012); Massey et al. (2012); Cropper et al. (2013). In contrast relatively little attention has been paid to the bispectrum even though many of the concepts developed for the power spectrum can readily be adapted. In this section we address the bispectrum and model two major sources of systematic error: intrinsic alignments of galaxies and uncertainty about their redshifts. We estimate by how much these systematics degrade

the figures of merit presented in Section 3.3 and explore whether they affect the bispectrum and power spectrum in different ways. We then consider the scope for self-calibration: the possibility of estimating the systematic nuisance parameters alongside the cosmological parameters using only data from the survey itself.

In Section 3.4.1 we first discuss the basic concepts of intrinsic alignments, for simplicity starting with their effect on two-point statistics, and go on to develop a model of the intrinsic alignment bispectrum. In Section 3.4.2 we discuss our modelling of redshift uncertainties. Results for both systematics are given in Section 3.4.3 and our conclusions are in Section 3.4.4.

3.4.1 Intrinsic alignment of galaxies

Basic concepts: two-point statistics

Galaxies (even spiral galaxies) are intrinsically elliptical and the shear signal is much smaller than this intrinsic ellipticity. If the intrinsic component was randomly oriented with respect to the shear field then in principle these two sources of ellipticity could be disentangled straightforwardly. However in fact galaxies are aligned with the local tidal gravitational field which is also the source of the lensing signal. Thus intrinsic alignment is correlated with lensing, which enormously complicates the extraction of the shear signal. Useful recent reviews are Joachimi et al. (2015) on all aspects of galaxy alignment and Troxel and Ishak (2015) specifically on the effect of intrinsic alignments on weak lensing.

As an introduction we first discuss the effect of intrinsic alignments on two-point statistics. The observed galaxy ellipiticity can be considered as the sum of the true gravitational shear, $\gamma_{\rm G}$, and the intrinsic ellipticity of the galaxy, $\epsilon_{\rm I}$:

$$\epsilon_{\rm obs} = \gamma_{\rm G} + \epsilon_{\rm I} \, . \tag{3.84}$$

So the observed two-point correlation function for two galaxy samples denoted by i and j is

$$\langle \epsilon^i_{\rm obs} \epsilon^j_{\rm obs} \rangle = \langle \gamma^i_{\rm G} \gamma^j_{\rm G} \rangle + \langle \gamma^i_{\rm G} \epsilon^j_{\rm I} \rangle + \langle \gamma^j_{\rm G} \epsilon^i_{\rm I} \rangle + \langle \epsilon^i_{\rm G} \epsilon^j_{\rm I} \rangle \tag{3.85}$$

The first term is the lensing signal, generally referred to as GG, which is what we want to measure. The next two terms are labelled GI and represent cross-correlations between shear and intrinsic ellipticity, and the fourth, II, term represents intrinsic-alignment–intrinsic-alignment correlation. If

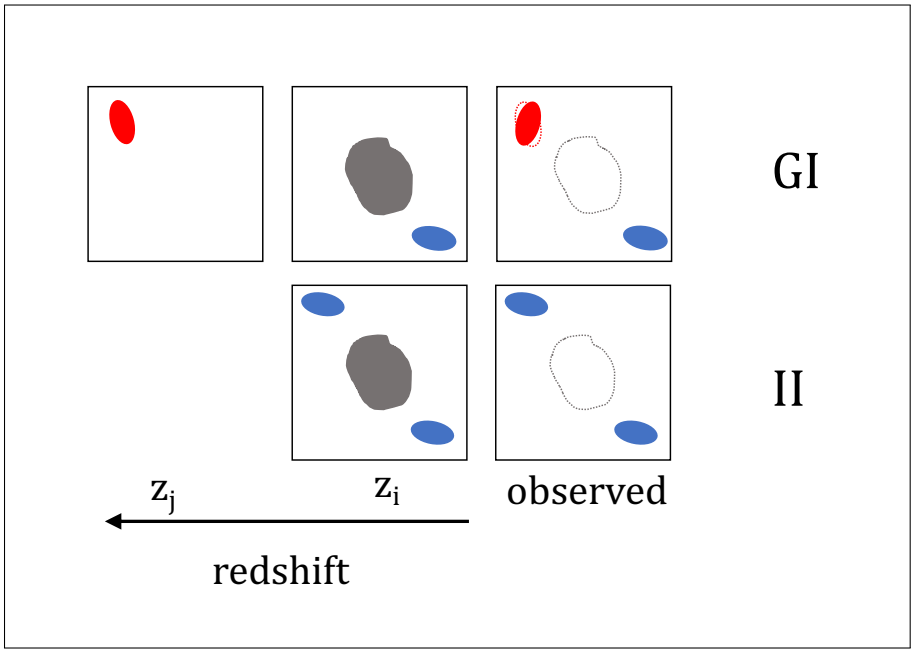

Figure 3.10: Origin of the GI and II terms. *Top*: GI term. A distant galaxy (coloured red) at redshift z_j is lensed by a structure at redshift z_i. The same galaxy is also gravitationally affected by a galaxy at redshift z_i (coloured blue) which is intrinsically aligned to the same matter structure. *Bottom*: II term. Two galaxies at redshift z_i are intrinsically aligned to the structure at the same redshift.
In each case the rightmost column shows what is observed, with the true orientation of the lensed galaxies shown by a dotted outline. Lensing is anti-correlated with intrinsic alignment.
Adapted from Figure 6 in Troxel and Ishak (2015).

we assume that $z_i < z_j$ then the second term should be negligible because it is not possible for an intrinsically-aligned galaxy at higher redshift to affect the lensing of a galaxy at lower redshift.

Figure 3.10, motivated by Figure 6 in Troxel and Ishak (2015), shows how the GI and II terms arise. The top panel shows the GI case. The grey shape represents a matter structure at redshift z_i which lenses a distant galaxy (red) at redshift z_j. The gravitational tidal field of the structure also affects the intrinsic alignment of a galaxy (blue) which is close to the structure (effectively at the same redshift and angular position). The rightmost panel shows what is observed, with the position of the matter structure shown by a dotted line: the observed ellipticities of the blue and red galaxies are correlated. In fact the lensing effect is anti-correlated with the intrinsic alignment, so that the GI term is negative.

The lower panel shows the II case - two galaxies which are both at the same redshift and angular position as the matter structure. Both are intrinsically aligned by its tidal gravitational field. The left panel shows their actual positions at redshift z_i and the right panel shows how they appear to an observer.

Controlling intrinsic alignment uncertainties

For two-point statistics three broad approaches have been used to control errors due to contamination from intrinsic alignments: modelling, self-calibration and nulling (Troxel and Ishak 2015). Modelling involves developing fitting functions for the intrinsic alignment signals with one or more free parameters (Joachimi and Bridle 2010). This can be extended to self-calibration in which systematic errors are modelled using other information from the survey, for example about the clustering of galaxies (Huterer et al. 2006; Zhang 2010). Finally, the nulling technique uses the redshift dependence of the correlation between intrinsic ellipticity and shear to remove the intrinsic alignment contamination, though with some loss of information (Joachimi and Schneider 2008).

Few studies have considered the control of intrinsic alignments for three-point shear statistics. Of these few, Troxel and Ishak (2011, 2012) used the redshift dependency within single redshift bins to inform a self-calibration method and Shi et al. (2010) studied the scope for using nulling. In this work we adapt the linear alignment model developed by Hirata and Seljak (2004) which is well-established for two-point statistics (Bridle and King 2007; Kirk et al. 2012) and has been used to inform practical investigations (Hildebrandt et al. 2016; Köhlinger et al. 2017; van Uitert et al. 2018; Hildebrandt et al. 2020).

The linear alignment model assumes that the ellipticity of a galaxy is linearly related to the gravitational potential at the time the galaxy formed. In its basic form the model is relatively simple; to make it more realistic it has been developed in various ways, for example Schneider and Bridle (2010) developed a halo model version which extended the modelling to small scales, and other models incorporate dependence on the redshift and luminosity of source galaxies, partly motivated by simulation results (Tenneti et al. 2016; Hilbert et al. 2017). We adopt some of these extensions, as explained in the next section. Appendix A of Kirk et al. (2012) usefully summarises the history and rationale for the linear alignment approach.

Modelling the intrinsic alignment bispectrum

We now extend the earlier discussion of intrinsic alignments to cover three-point statistics. From Equation 3.84 we can see that the three-point weak lensing correlation function is the sum of four terms

$$\langle \epsilon^i_{\rm obs} \epsilon^j_{\rm obs} \epsilon^k_{\rm obs} \rangle = {\rm GGG} + {\rm GGI} + {\rm GII} + {\rm III} \,, \tag{3.86}$$

where i, j and k denote different galaxies, G represents lensing and I represents intrinsic alignment.

The four terms in this equation are given by

$$\mathrm{GGG} = \langle \gamma_{\mathrm{G}}^i \gamma_{\mathrm{G}}^j \gamma_{\mathrm{G}}^k \rangle \tag{3.87}$$

$$\mathrm{GGI} = \langle \gamma_{\mathrm{G}}^i \gamma_{\mathrm{G}}^j \epsilon_{\mathrm{I}}^k \rangle + \langle \gamma_{\mathrm{G}}^j \gamma_{\mathrm{G}}^k \epsilon_{\mathrm{I}}^i \rangle + \langle \gamma_{\mathrm{G}}^k \gamma_{\mathrm{G}}^i \epsilon_{\mathrm{I}}^j \rangle \tag{3.88}$$

$$\mathrm{GII} = \langle \gamma_{\mathrm{G}}^i \epsilon_{\mathrm{I}}^j \epsilon_{\mathrm{I}}^k \rangle + \langle \gamma_{\mathrm{G}}^j \epsilon_{\mathrm{I}}^k \epsilon_{\mathrm{I}}^i \rangle + \langle \gamma_{\mathrm{G}}^k \epsilon_{\mathrm{I}}^j \epsilon_{\mathrm{I}}^i \rangle \tag{3.89}$$

$$\mathrm{III} = \langle \epsilon_{\mathrm{I}}^i \epsilon_{\mathrm{I}}^j \epsilon_{\mathrm{I}}^k \rangle \, . \tag{3.90}$$

In a similar way we can divide the observed bispectrum B_{obs} into four terms

$$B_{\mathrm{obs}} = B_{\mathrm{GGG}} + B_{\mathrm{GGI}} + B_{\mathrm{GII}} + B_{\mathrm{III}} \, . \tag{3.91}$$

Figure 3.11 shows the origin of these terms using notation similar to Figure 3.10. The configurations in the three-point case are more complicated than for two-point statistics. In particular GII terms can arise in two different ways shown in the two central panels of Figure 3.11: firstly a distant galaxy may be lensed by two structures along the line of sight each of which has an associated nearby intrinsically-aligned galaxy, and secondly lensing may be due to a single structure with two nearby intrinsically-aligned galaxies.

The first term, B_{GGG}, is the lensing bispectrum defined by

$$\langle \tilde{\gamma}_{\mathrm{G}}^i(\boldsymbol{\ell}_1) \tilde{\gamma}_{\mathrm{G}}^j((\boldsymbol{\ell}_2) \tilde{\gamma}_{\mathrm{G}}^k(\boldsymbol{\ell}_3) \rangle = (2\pi)^2 \delta_{\mathrm{D}}((\boldsymbol{\ell}_1 + \boldsymbol{\ell}_2 + \boldsymbol{\ell}_3) B_{\mathrm{GGG}}^{ijk}((\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \boldsymbol{\ell}_3) \, , \tag{3.92}$$

where $\tilde{\gamma}_{\mathrm{G}}$ is the Fourier transform of the shear. In a tomographic analysis we can associate the labels i, j and k with different redshift bins rather than different galaxy samples and assume $z_i < z_j < z_k$.

Using the lensing weight $q_{(i)}(\chi)$ defined by Equation 3.9 we can write the lensing bispectrum as a projection of the three-dimensional matter bispectrum, $B_{\delta\delta\delta}^{ijk}$:

$$B_{\mathrm{GGG}}^{ijk}(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \boldsymbol{\ell}_3) = \int_0^{\chi_{\mathrm{lim}}} \mathrm{d}\chi \, q_{(i)}(\chi) q_{(j)}(\chi) q_{(k)}(\chi) \chi^{-4} B_{\delta\delta\delta}\left(\frac{\boldsymbol{\ell}_1}{\chi}, \frac{\boldsymbol{\ell}_2}{\chi}, \frac{\boldsymbol{\ell}_3}{\chi}; \chi \right) . \tag{3.93}$$

The other three terms in Equation 3.91 can be defined similarly, replacing $\tilde{\gamma}_{\mathrm{G}}$ by $\tilde{\epsilon}_{\mathrm{I}}$ as appropriate.

For example for B_{GGI}

$$\langle \tilde{\gamma}_{\mathrm{G}}^i(\boldsymbol{\ell}_1) \tilde{\gamma}_{\mathrm{G}}^j(\boldsymbol{\ell}_2) \tilde{\epsilon}_{\mathrm{I}}^k(\boldsymbol{\ell}_3) = (2\pi)^2 \delta_{\mathrm{D}}(\boldsymbol{\ell}_1 + \boldsymbol{\ell}_2 + \boldsymbol{\ell}_3) B_{\mathrm{GGI}}^{ijk}(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \boldsymbol{\ell}_3) \, , \tag{3.94}$$

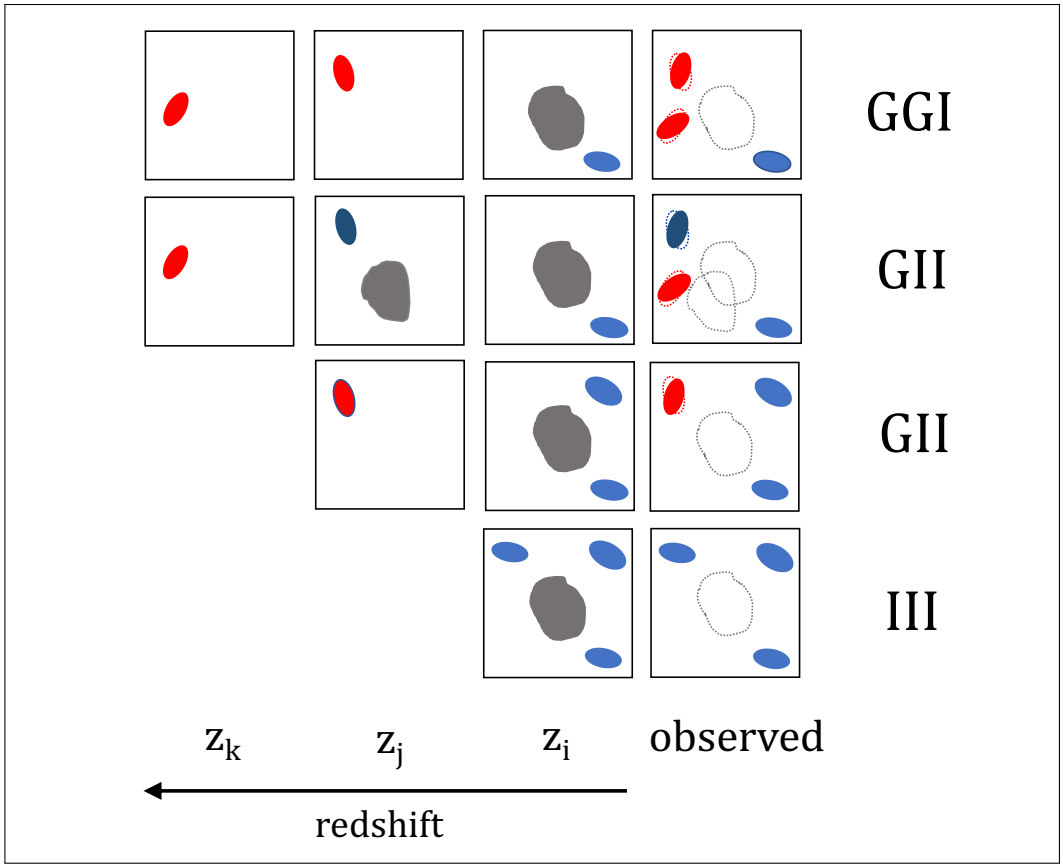

Figure 3.11: Origin of the GGI, GII and III terms. *Top panel*: GGI term. Distant galaxies (coloured red) at redshifts z_j and z_k are lensed by a structure at redshift z_i. The galaxies are also affected by a galaxy at redshift z_i (coloured blue) which is intrinsically aligned to the matter structure at the same redshift. *Centre two panels*: GII terms. These can arise in two ways. Firstly a galaxy (coloured red) at redshift z_k is lensed by structures at redshifts z_j and z_i. Two other galaxies (coloured blue) are intrinsically aligned to the same structures. This configuration can also be considered to be intermediate between GGI and GII. Secondly, the main source of the GII signal, a galaxy at redshift z_j is lensed by a structure at redshift z_i and two other galaxies at redshift z_i are aligned to the same structure. *Bottom panel*: III term. Three galaxies at redshift z_i are instrinsically aligned to a structure at the same redshift.
In each case the rightmost column shows what is observed, with the true orientation of the lensed galaxies shown by a dotted outline. Lensing is anti-correlated with intrinsic alignment.
Adapted from Figure 7 in Troxel and Ishak (2015).

and

$$B^{ijk}_{\mathrm{GGI}}(\boldsymbol{\ell}_1,\boldsymbol{\ell}_2,\boldsymbol{\ell}_3) = \int_0^{\chi_{\mathrm{lim}}} \mathrm{d}\chi\; q_{(i)}(\chi) q_{(j)}(\chi) p_{(k)}(\chi) \chi^{-4} B_{\delta\delta\delta_{\mathrm{I}}}\left(\frac{\boldsymbol{\ell}_1}{\chi},\frac{\boldsymbol{\ell}_2}{\chi},\frac{\boldsymbol{\ell}_3}{\chi};\chi\right), \tag{3.95}$$

where $p_{(k)}(\chi)$ is the distribution of galaxies in the kth tomographic bin and $B_{\delta\delta\delta_{\mathrm{I}}}$ is defined by

$$\langle \tilde{\delta}_{\mathrm{G}}(\mathbf{k}_1,\chi)\tilde{\delta}_{\mathrm{G}}(\mathbf{k}_2,\chi)\tilde{\delta}_{\mathrm{I}}(\mathbf{k}_3,\chi)\rangle = (2\pi)^3\delta_{\mathrm{D}}(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3) B_{\delta\delta\delta_{\mathrm{I}}}\left(\frac{\boldsymbol{\ell}_1}{\chi},\frac{\boldsymbol{\ell}_2}{\chi},\frac{\boldsymbol{\ell}_3}{\chi};\chi\right). \tag{3.96}$$

Here $\tilde{\delta}_{\mathrm{I}}$ denotes the Fourier transform of the matter density contrast of the field which produces the intrinsic alignment. The full expression for the B_{GGI} part of the observed bispectrum is formed by summing Equation 3.95 over all permutations of the three redshift bins.

To evaluate the matter bispectrum we use the fitting formula for the three-dimensional matter bispectrum derived by Gil-Marín et al. (2012) which is based on perturbation theory. In this formula the normal perturbation theory kernels F_2 are replaced by new effective kernels F_2^{eff} so that

$$B_{\delta\delta\delta}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3) = 2F_2^{\mathrm{eff}}(\mathbf{k}_1,\mathbf{k}_2)P_\delta(k_1)P_\delta(k_2) + 2 \text{ perms.} \tag{3.97}$$

We need similar expressions for $B_{\delta\delta\delta_{\mathrm{I}}}$, $B_{\delta\delta_{\mathrm{I}}\delta_{\mathrm{I}}}$ and $B_{\delta_{\mathrm{I}}\delta_{\mathrm{I}}\delta_{\mathrm{I}}}$. We build on the expression for $\tilde{\delta}_{\mathrm{I}}$ derived by Hirata and Seljak (2004) for intrinsic alignment power spectra. In the two-point case there are two such spectra, $P_{\delta_{\mathrm{I}}\delta_{\mathrm{I}}}(k)$ and $P_{\delta\delta_{\mathrm{I}}}(k)$, given by

$$P_{\delta_{\mathrm{I}}\delta_{\mathrm{I}}}(k) = \left(-\frac{C_1\rho_c}{(1+z)D(z)}\right)^2 P_\delta^{\mathrm{lin}}(k)\,, \tag{3.98}$$

$$P_{\delta\delta_{\mathrm{I}}}(k) = -\frac{C_1\rho_c}{(1+z)D(z)} P_\delta^{\mathrm{lin}}(k)\,, \tag{3.99}$$

where ρ_c is the critical density, $D(z)$ is the growth factor normalised to unity at the present day and $P_\delta^{\mathrm{lin}}(k)$ is the linear matter power spectrum. The parameter C_1 is a normalisation factor which in principle can be determined from observations or simulations. Thus in this model $\tilde{\delta}_{\mathrm{I}}$ is related to $\tilde{\delta}_{\mathrm{G}}$ as $\tilde{\delta}_{\mathrm{I}} = f_{\mathrm{IA}}\tilde{\delta}_{\mathrm{G}}$ with

$$f_{\mathrm{IA}} = -\frac{C_1\rho_c}{(1+z)D(z)}\,. \tag{3.100}$$

For both the power spectrum and bispectrum we choose a more flexible version of the factor $f_{\rm IA}$ than in Equation 3.100 to allow for uncertainty in the amplitude and possible redshift dependence:

$$f_{\rm IA} = -A_{\rm IA}\frac{C_1\rho_c}{(1+z)D(z)}\left(\frac{1+z}{1+z_0}\right)^{\beta}. \tag{3.101}$$

The amplitude $A_{\rm IA}$ and the exponent β are free to vary and z_0 is an arbitrary pivot value. We substitute the non-linear power spectrum in place of $P_\delta^{\rm lin}(k)$ following Bridle and King (2007) who found this was a practicable way of extending the model to smaller scales. We use the value of C_1 determined by the same authors which is $5\times10^{-14}\ h^{-4}{\rm M}_\odot^{-2}{\rm Mpc}^6$, leading to $C_1\rho_c = 0.0134$ (Joachimi, Mandelbaum, Abdalla and Bridle 2011). We take the fiducial value of the amplitude to be 1 and of β to be zero.

We extend this model to the intrinsic alignment bispectra we use the fitting function from Equation 3.97 which leads to

$$B_{\delta_{\rm I}\delta_{\rm I}\delta_{\rm I}}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3) = f_{\rm IA}^4 B_{\delta\delta\delta}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3) \tag{3.102}$$

$$\begin{aligned} B_{\delta\delta_{\rm I}\delta_{\rm I}}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3) = 2\big[&f_{\rm IA}^2 F_2^{\rm eff}(\mathbf{k}_1,\mathbf{k}_2)P_\delta(k_1)P_\delta(k_2) \\ &+ 2f_{\rm IA}^3[F_2^{\rm eff}(\mathbf{k}_2,\mathbf{k}_3)P_\delta(k_2)P_\delta(k_3) + F_2^{\rm eff}(\mathbf{k}_3,\mathbf{k}_1)P_\delta(k_3)P_\delta(k_1)]\big] \end{aligned} \tag{3.103}$$

$$\begin{aligned} B_{\delta\delta\delta_{\rm I}}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3) = 2\big[&f_{\rm IA}^2 F_2^{\rm eff}(\mathbf{k}_1,\mathbf{k}_2)P_\delta(k_1)P_\delta(k_2) \\ &+ 2f_{\rm IA}[F_2^{\rm eff}(\mathbf{k}_2,\mathbf{k}_3)P_\delta(k_2)P_\delta(k_3) + F_2^{\rm eff}(\mathbf{k}_3,\mathbf{k}_1)P_\delta(k_3)P_\delta(k_1)]\big]\,. \end{aligned} \tag{3.104}$$

Then integrating as in Equation 3.95 gives expressions for the weak lensing intrinsic alignment bispectra. Examples of resulting bispectra for arbitrary illustrative tomographic bins are shown in Figure 3.12. This plots equilateral triangle bispectra obtained with 5 redshift bins, assuming fiducial values of the intrinsic alignment parameters $A_{\rm IA}$ and β. The GGI bispectrum is negative and its magnitude can be almost as large as the GGG signal. The other bispectra are positive. In some bin combinations the GII bispectrum can also be large, but the III bispectrum is always sub-dominant, which is consistent with the findings in Semboloni et al. (2008) for a survey like CFHTLens. Overall this confirms the importance of taking intrinsic alignments into account; otherwise the weak lensing bispectrum can be seriously mis-estimated.

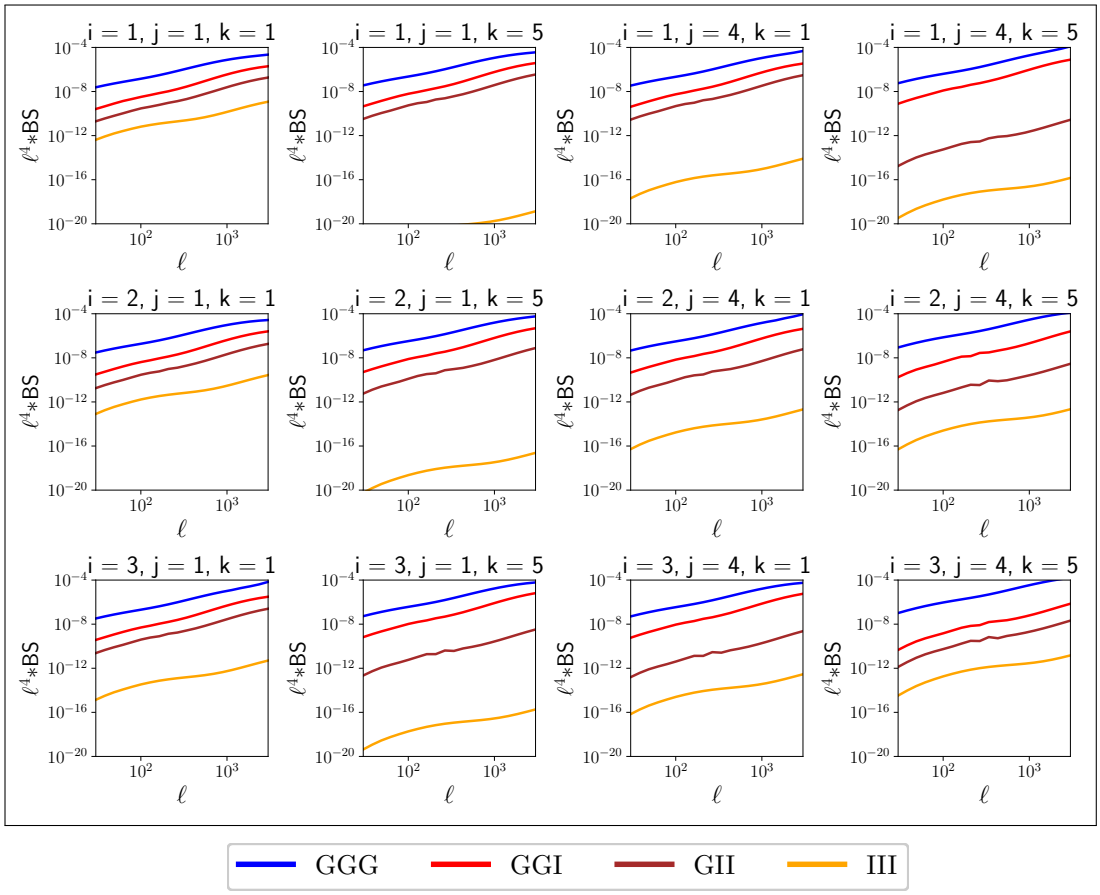

Figure 3.12: Examples of intrinsic alignment bispectra for illustrative tomographic bin combinations. Results are given for equilateral triangle configurations with 5 redshift bins, assuming fiducial values of the parameters $A_{\rm IA}$ and β, and show the absolute values of the bispectra since the intrinsic alignment bispectra can be negative in some configurations. The magnitudes of the GGI and GII bispectra can approach that of the GGG lensing signal in some bin combinations. The III bispectrum is always sub-dominant.

Figures 3.13 and 3.14 show the relative importance of the intrinsic alignment terms compared with the pure lensing signal. We find that intrinsic alignment affects the power spectrum more than the bispectrum, confirming that the bispectrum should help mitigate this systematic. However we note that this contrasts with the findings from simulations in Semboloni et al. (2008). These authors measured three-point aperture mass statistics and concluded that the III/GGG ratio was generally higher than the II/GG ratio. They also found that the III signal is negative whereas in our model it is positive. We discuss these disagreements further in Section 3.5. For now we have no reason to discard the linear alignment model which is well-established and robust for two-point statistics.

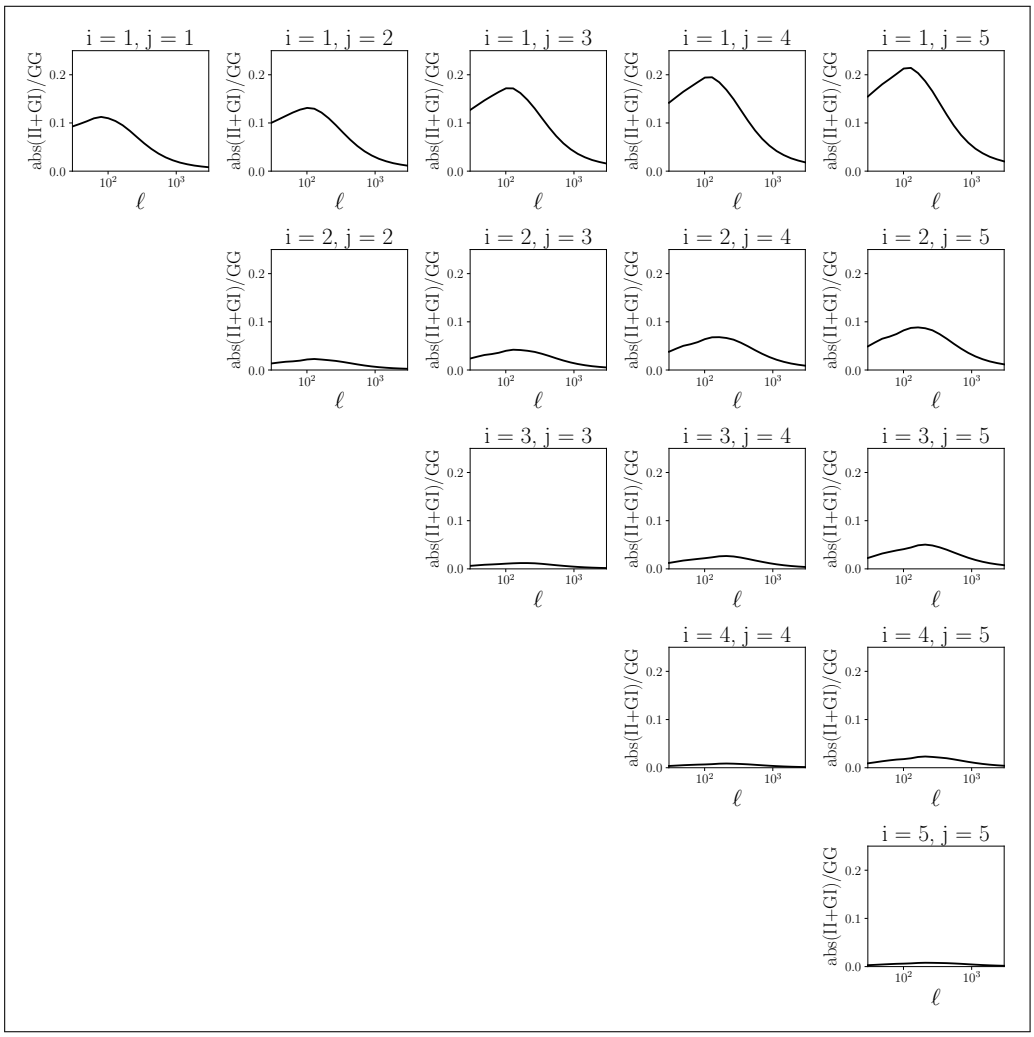

Figure 3.13: The absolute value of the total the intrinsic alignment power spectrum relative to the lensing power spectrum for all combinations of 5 tomographic bins.

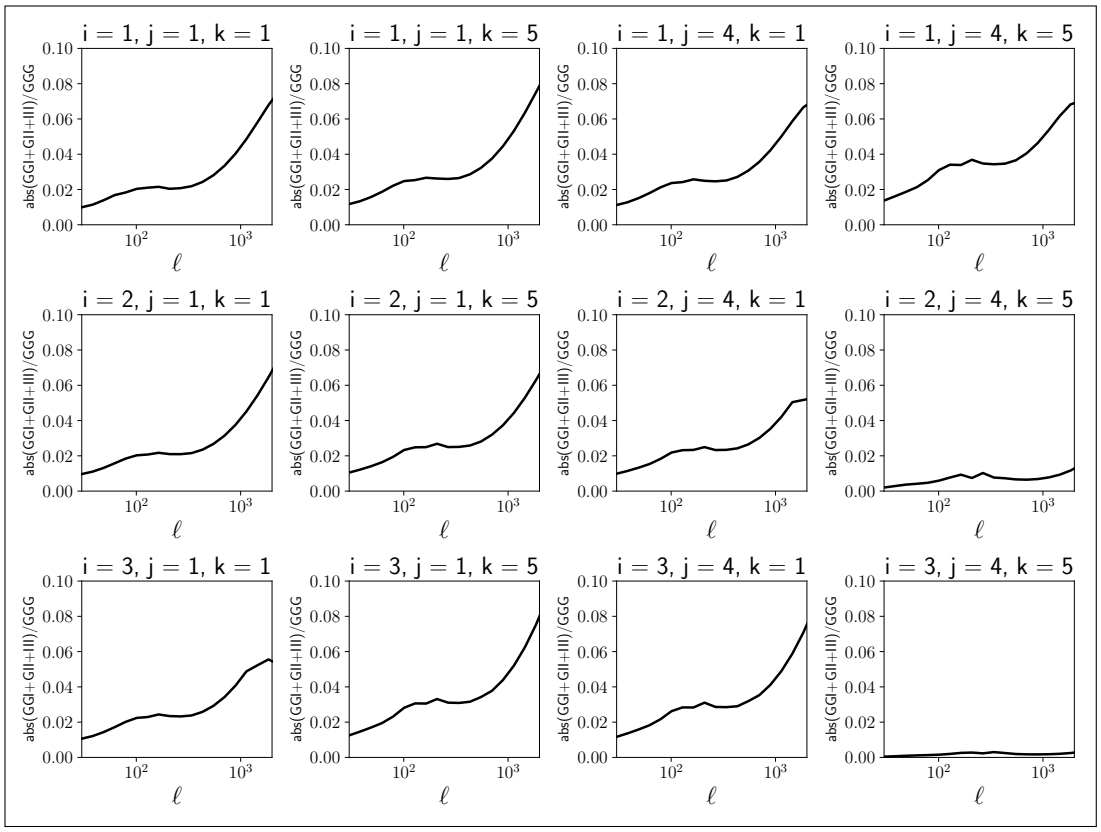

Figure 3.14: For equilateral triangles only, the absolute value of the total intrinsic alignment bispectrum relative to the lensing bispectrum for the same tomographic bin combinations as in Figure 3.12. Comparison with Figure 3.13 shows that intrinsic alignment affects the bispectrum less than the power spectrum.

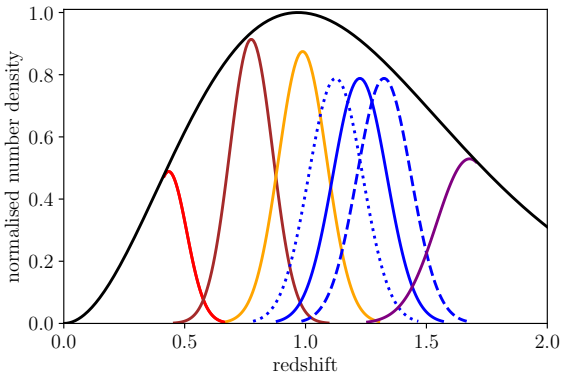

Figure 3.15: Schematic depiction of the number density distribution of galaxies. The black line shows the overall number density distribution normalised so that the maximum value is 1. Bins are defined so that each contains the same number of galaxies and coloured lines show distributions within each redshift bin. Dotted and dashed blue lines show uncertainty in the mean redshift of the fourth bin.

3.4.2 Redshift errors

Another source of systematic uncertainty is the measurement of the redshifts which are used to allocate galaxies to tomographic bins and estimate the distribution of galaxies in each bin. It is not feasible to obtain accurate spectroscopic redshifts for all the galaxies in a survey and we have to rely on photometric measurements which will always include some inescapable uncertainty. In this work we consider a single source of uncertainty: bias in the mean redshift of each tomographic bin. In other words we consider the effect of shifting the whole distribution of galaxies in a bin to higher or lower redshift, without changing the shape of the distribution. We allow for different uncertainty and hence different shifts in each tomographic bin so that with N_z tomographic bins there are N_z parameters representing the shifts in the mean of each bin. A similar method for forecasting redshift systematics was used by Huterer et al. (2006). It is also the standard approach used in all current surveys, for example Joudaki et al. (2016), Abbott et al. (2018), Hikage et al. (2019).

We assume the overall redshift probability distribution of galaxies, $p_{\rm tot}(z)$, has the form given by Equation 3.79 and that the redshift distribution within each bin is Gaussian. The redshift bin boundaries are then constructed so that each bin contains the same number of galaxies. This results in bin boundaries [0.20,0.51], [0.51,0.71], [0.71,0.91], [0.91,1.17] and [1.17,2.00]. Figure 3.15 shows this schematically, firstly assuming the mean redshift of each bin is known exactly and secondly indicating positive (dashed) and negative (dotted) shifts in the mean of the fourth redshift bin.

3.4.3 Results

We now expand the Fisher information matrix which we used in Section 3.3 to additionally take account of the two nuisance parameters for the intrinsic alignment model – the parameters $A_{\rm IA}$ and β from Equation 3.101 – and N_z nuisance parameters Δz_i denoting the shift in the mean value of the redshift bin centered on z_i, as well the seven cosmological parameters listed in 3.3.1.

As in earlier parts of this chapter we use equilateral triangle bispectra only and work with five tomographic bins. The covariance matrix includes the Gaussian (including shape noise) and supersample terms allowing better estimation of statistical errors than for example in Shi et al. (2010) who used only the Gaussian covariance.

As before we use figures of merit defined by Equation 3.83 to quantify the improvements in parameter constraints. We focus on the $\Omega_{\rm m}$–σ_8 and w_0–w_a figures of merit, since these parameter combinations are degenerate in weak lensing. For completeness, Appendix 3D contains Fisher matrix results including all cosmological and nuisance parameters.

To calculate the Fisher matrices we need the derivatives of the power spectrum and bispectrum with respect to the parameters. For the cosmological parameters and redshift shifts we use a standard five-point stencil. The derivatives with respect to the intrinsic alignment parameters are evaluated analytically.

In Table 3.2 we summarise the impact of systematic uncertainties on the figures of merit, and demonstrate the effect of combining the power spectrum and bispectrum. We compare the figures of merit with no systematics with firstly those involving intrinsic alignments only, secondly redshift uncertainties only and finally both systematics together. The presence of systematic uncertainties severely reduces the figures of merit. However, as also shown in Table 3.3, this reduction can be partly redressed by including the bispectrum as well as the power spectrum, and in fact the bispectrum is much more beneficial when systematic uncertainties are present. In both these tables we use wide, essentially uninformative, priors on all the nuisance parameters.

We next investigate the effect on the figures of merit of narrowing the priors on the nuisance parameters, starting from a very wide prior of ten. As we narrow the priors (ie increase our knowledge of the parameters) the figures of merit increase until they reach a plateau where the improvement levels off. At this point effectively no further information is obtainable: if the nuisance parameters were constrained as tightly as this then we could retrieve all information. The difference between the plateau obtained with tight priors and the 'base' level obtained with wide priors

Spectrum type	FoM $\Omega_{\rm m}$–σ_8	FoM w_0–w_a
PS no systematics	71139	160
PS with intrinsic alignments	229	0.4
PS with redshift bin shifts	7752	4.5
PS with both systematics	210	0.3
PS + BS no systematics	116734	268
PS + BS with intrinsic alignments	16198	5.2
PS + BS with redshift bin shifts	66359	40.1
PS + BS with both systematics	14170	4.8

Table 3.2: The table compares the figure of merit obtained firstly with the power spectrum only, and then using the power spectrum and bispectrum together. In each case we show the figures of merit without systematic uncertainties (the best which can theoretically be achieved), then in the presence of intrinsic alignments only, redshift uncertainties only, and finally both together.

	Ratio of figures of merit	
Spectrum type	$\Omega_{\rm m}$–σ_8	w_0–w_a
(PS + BS)/PS no systematics	1.64	1.60
(PS + BS)/PS with intrinsic alignments	70.7	13.0
(PS + BS)/PS with redshift bin shifts	8.6	8.9
(PS + BS)/PS with both systematics	67.5	15.5

Table 3.3: Effect of the bispectrum on figures of merit. Based on the data in Table 3.2, we show the ratios of the figures of merit using the combined power spectrum and bispectrum to the figures of merit using the power spectrum only. The bispectrum results in large improvements in the figure of merit if systematic uncertainties are present.

quantifies the loss of information caused by systematic uncertainties. The scope for self-calibration is indicated by the value at which the increase in figure of merit occurs. The smaller this value is, the less need for external information to help calibrate the nuisance parameters.

We first consider the intrinsic alignment parameters only. Figure 3.16 shows that if only the power spectrum is used the improvement plateaus when the priors on $A_{\rm IA}$ and β are around 0.1. The percentage improvement in the figures of merit as the priors are tightened is less for the combined power spectrum and bispectrum than for the power spectrum alone. However this is an improvement to figures which are already considerably larger, as shown in Table 3.2. The net effect is that, when systematics are present, including the bispectrum will produce a significant improvement in parameter constraints and better self-calibration. We also observe that, for the power spectrum, the $\Omega_{\rm m}$–σ_8 figure of merit is most sensitive to the amplitude parameter $A_{\mathbf{IA}}$ and the w_0–w_a figure of merit is most sensitive to the redshift exponent β. This is as expected: $\Omega_{\rm m}$ and σ_8 are related to the growth of structure which also affects the amplitude of the intrinsic alignment

signal, whereas the dark energy parameters depend on the expansion of the Universe and hence on the measured redshift.

In Figure 3.17 priors on both intrinsic alignment parameters are tightened simultaneously. In this case the self-calibration regime starts at a prior of around 0.1.

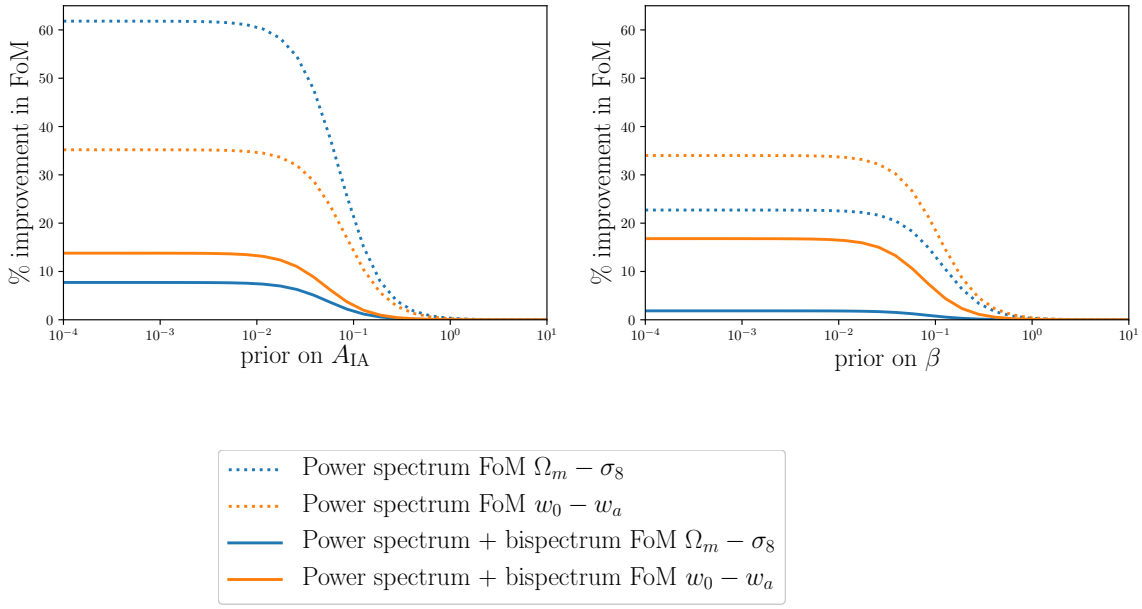

Figure 3.16: Percentage increase in figures of merit when the priors on the parameters $A_{\rm IA}$ and β are tightened, compared to a wide prior of 10. *Left*: Effect of tightening prior on $A_{\mathbf{IA}}$ only. *Right*: Effect of tightening prior on β only.

In Figure 3.18 we turn to the N_z redshift bin nuisance parameters. The top row shows the impact of tightening priors when we consider only the power spectrum and the bottom row shows the effect when the power spectrum and bispectrum are combined. The left and right panels respectively show the figures of merit for $\Omega_{\rm m}$–σ_8 and w_0–w_a. Each line shows the effect of tightening the prior on a single redshift bin. The self-calibration regime extends to a prior of around 0.001 on the redshift bin mean. This is confirmed by Figure 3.19 where we tighten the priors on all the redshift bin means simultaneously.

Finally we consider what happens if we know the accuracy of either the intrinsic alignment parameters or the redshift parameters so that we can impose fixed priors. In Figure 3.20 we show the effect of imposing priors on the redshift parameters and varying the priors on the intrinsic alignment parameters and in Figure 3.21 we show the reverse. The effects are summarised in Table 3.4 and Table 3.5. If both the power spectrum and power spectrum are used and priors on the

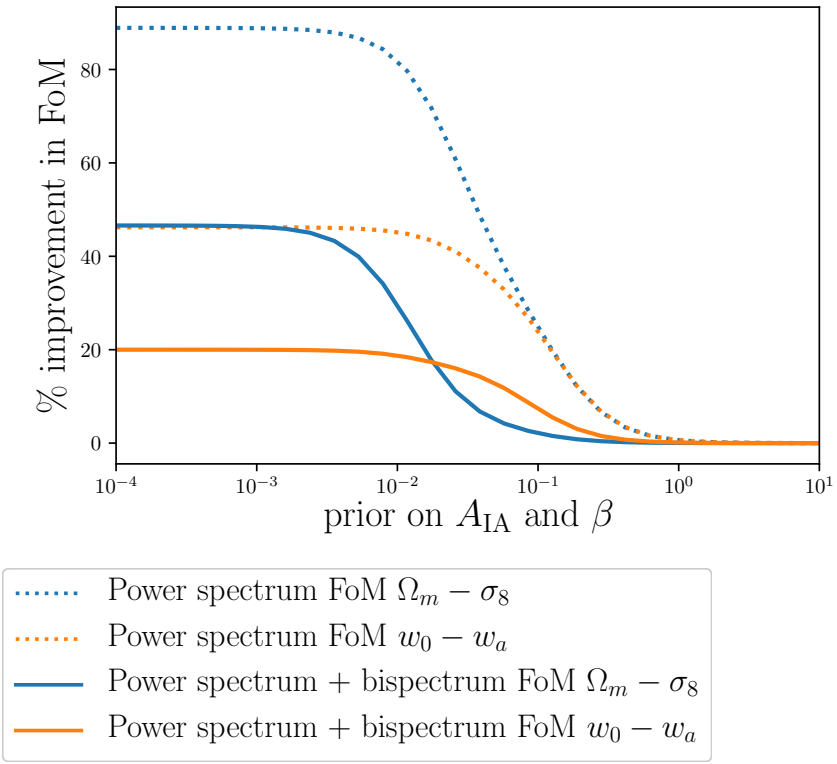

Figure 3.17: Percentage increase in figures of merit when the priors on both the parameters $A_{\rm IA}$ and β are tightened simultaneously, compared to wide priors of 10.

nuisance parameters are sufficiently tight the figures of merit can increase by as much as 70% for $\Omega_{\rm m}$–σ_8 and 15% for w_0–w_a, compared with using the power spectrum only.

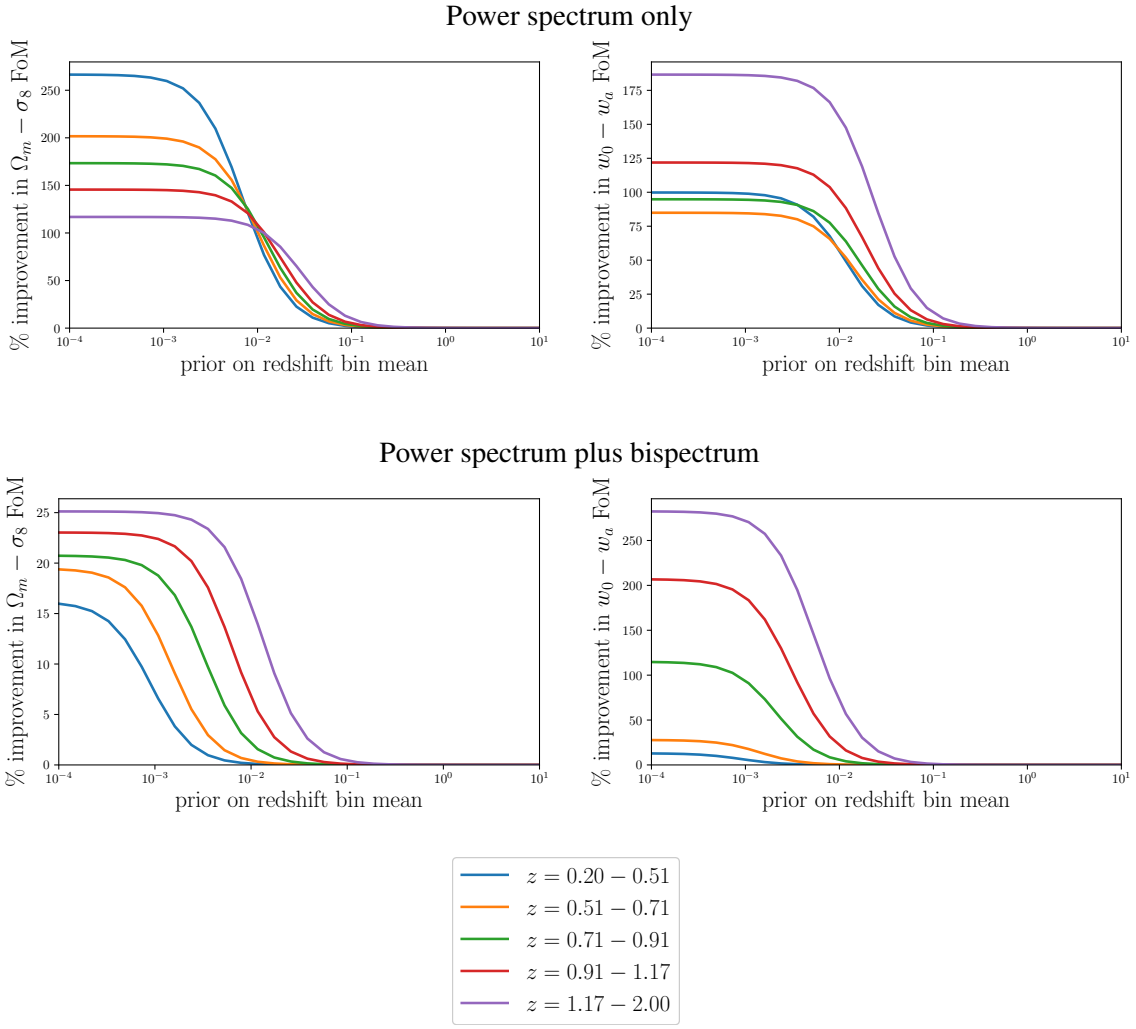

Figure 3.18: Percentage increase in figures of merit when the priors on the means of the five redshift bins are tightened, compared to a wide prior of 10. *Top*: Power spectrum. *Bottom*: Power spectrum and bispectrum combined *Left*: Figure of merit $\Omega_{\rm m}$–σ_8. *Right*: Figure of merit w_0–w_a. Only redshift nuisance parameters are considered. Each redshift bin prior is varied individually. Note different scales in each panel.

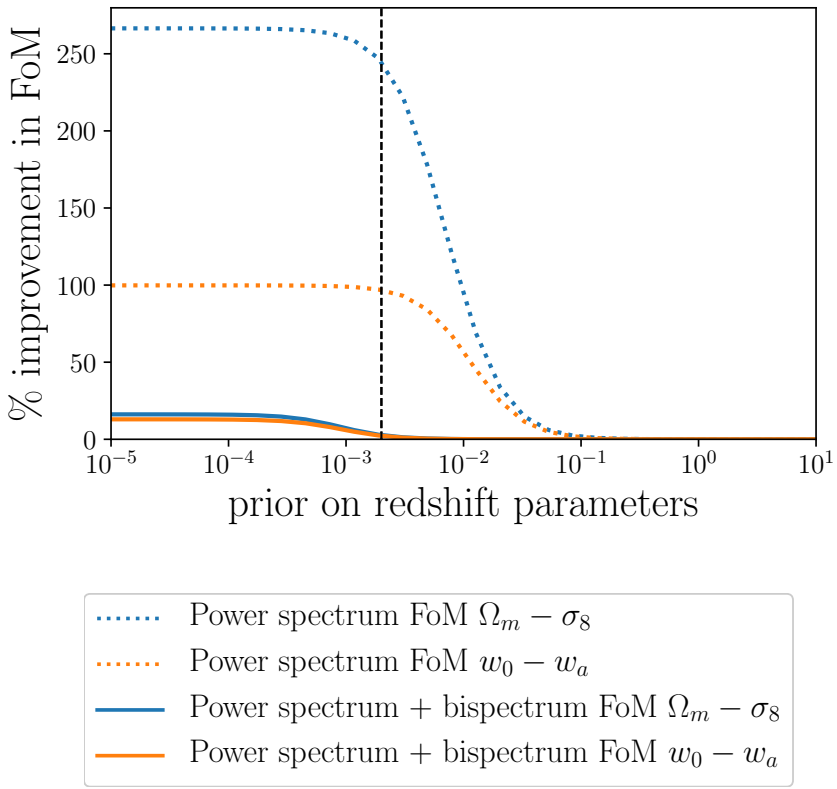

Figure 3.19: Percentage increase in figures of merit when the priors on all redshift parameters are tightened simultaneously. The vertical dashed line indicates the requirement for redshift accuracy from the Euclid Definition Study Report (Laureijs et al. 2011).

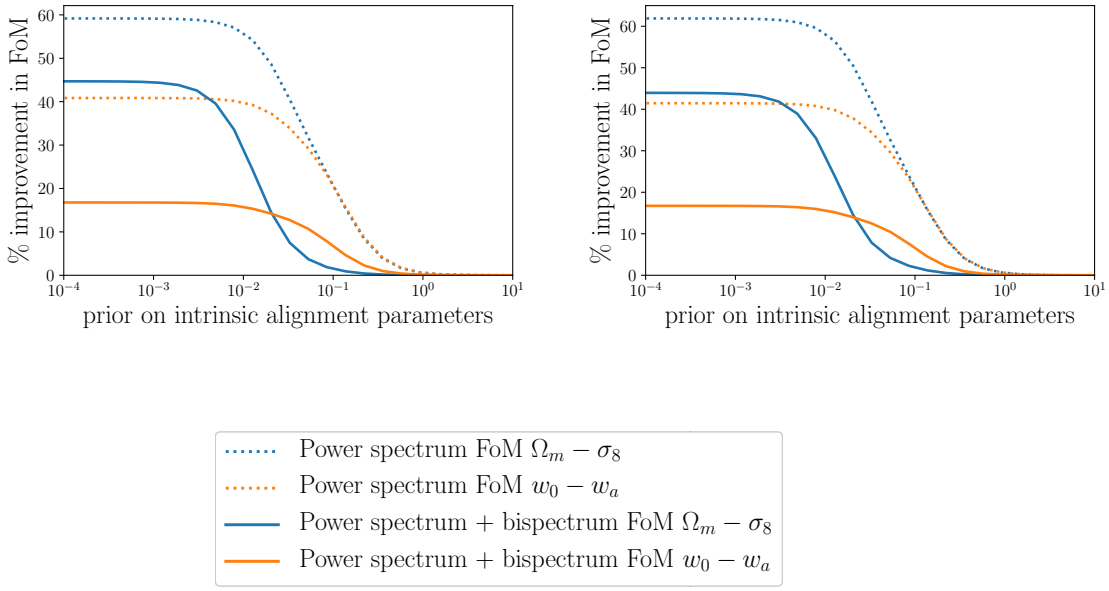

Figure 3.20: Percentage increase in figures of merit from varying the prior on the intrinsic alignment parameters when a fixed prior is imposed on every redshift parameter.
Left: Prior of 0.1. *Right*: Prior of 0.002.

Spectrum type	FoM Ω_m–σ_8	FoM w_0–w_a
PS 0.1 prior on IA parameters	254	0.4
PS 0.01 prior on redshift parameters	211	0.3
PS 0.002 prior on redshift parameters	215	0.3
PS + BS 0.1 prior on IA parameters	14927	5.1
PS + BS 0.01 prior on redshift parameters	14793	4.8
PS + BS 0.002 prior on redshift parameters	15429	4.9

Table 3.4: The table shows figures of merit obtained with the power spectrum only and with the power spectrum and bispectrum together, when priors are imposed on the nuisance parameters.

	Ratio of figures of merit	
Spectrum type	Ω_m–σ_8	w_0–w_a
(PS + BS)/PS 0.1 prior on IA parameters	58.8	12.8
(PS + BS)/PS 0.01 prior on redshift parameters	70.1	15.5
(PS + BS)/PS 0.002 prior on redshift parameters	71.8	15.0

Table 3.5: Based on the values in Table 3.4, we show the ratios of the figures of merit using the combined power spectrum and bispectrum to the figures of merit using the power spectrum only, when priors are imposed on the nuisance parameters.

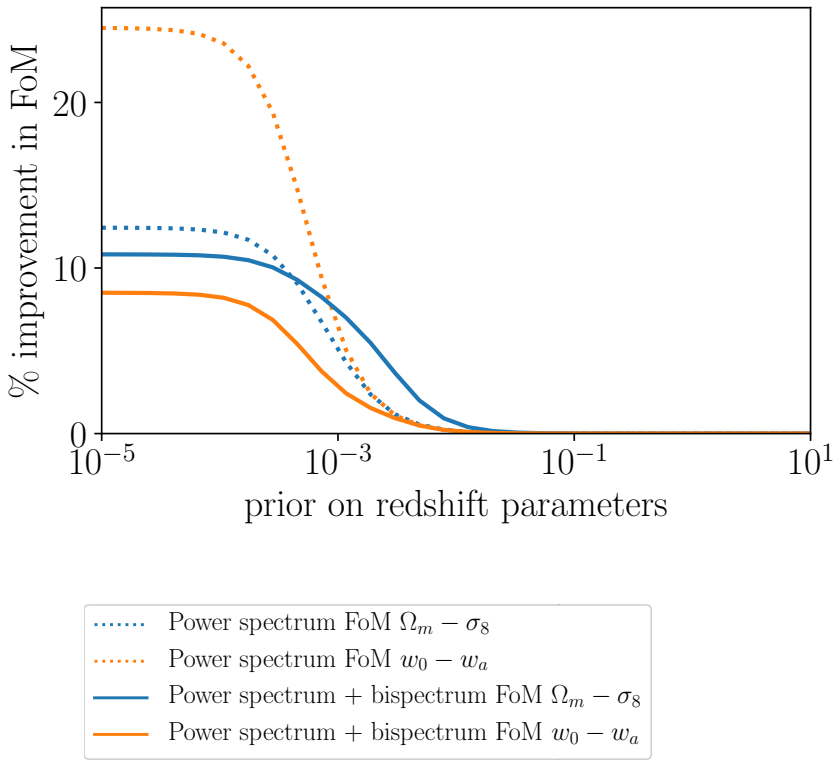

Figure 3.21: Percentage increase in figures of merit from varying the prior on the redshift parameters when a fixed prior of 0.1 is imposed on both intrinsic alignment parameters.

Bringing this all together, Figure 3.22 summarises our results, highlighting the 'worst case' where we use the power spectrum only and know nothing about the systematics and also the case where the power spectrum and bispectrum are combined and the *Euclid* requirement for redshift accuracy is met. Two key conclusions are firstly that using the bispectrum as well as the power spectrum goes a long way towards mitigating the effect of systematics and secondly that lack of knowledge of the nuisance parameters causes little loss of information so that self-calibration is possible. This is particularly true for the redshift parameters where the self-calibration regime coincides with the *Euclid* accuracy requirement.

To further highlight the main results from this section, in Figure 3.23 we recast the information presented in Figure 3.22 to plot the percentage change in the figures of merit compared with the figure of merit obtained from the power spectrum only and a tight redshift prior of 0.002. The improvement from the bispectrum is large and effectively independent of the prior.

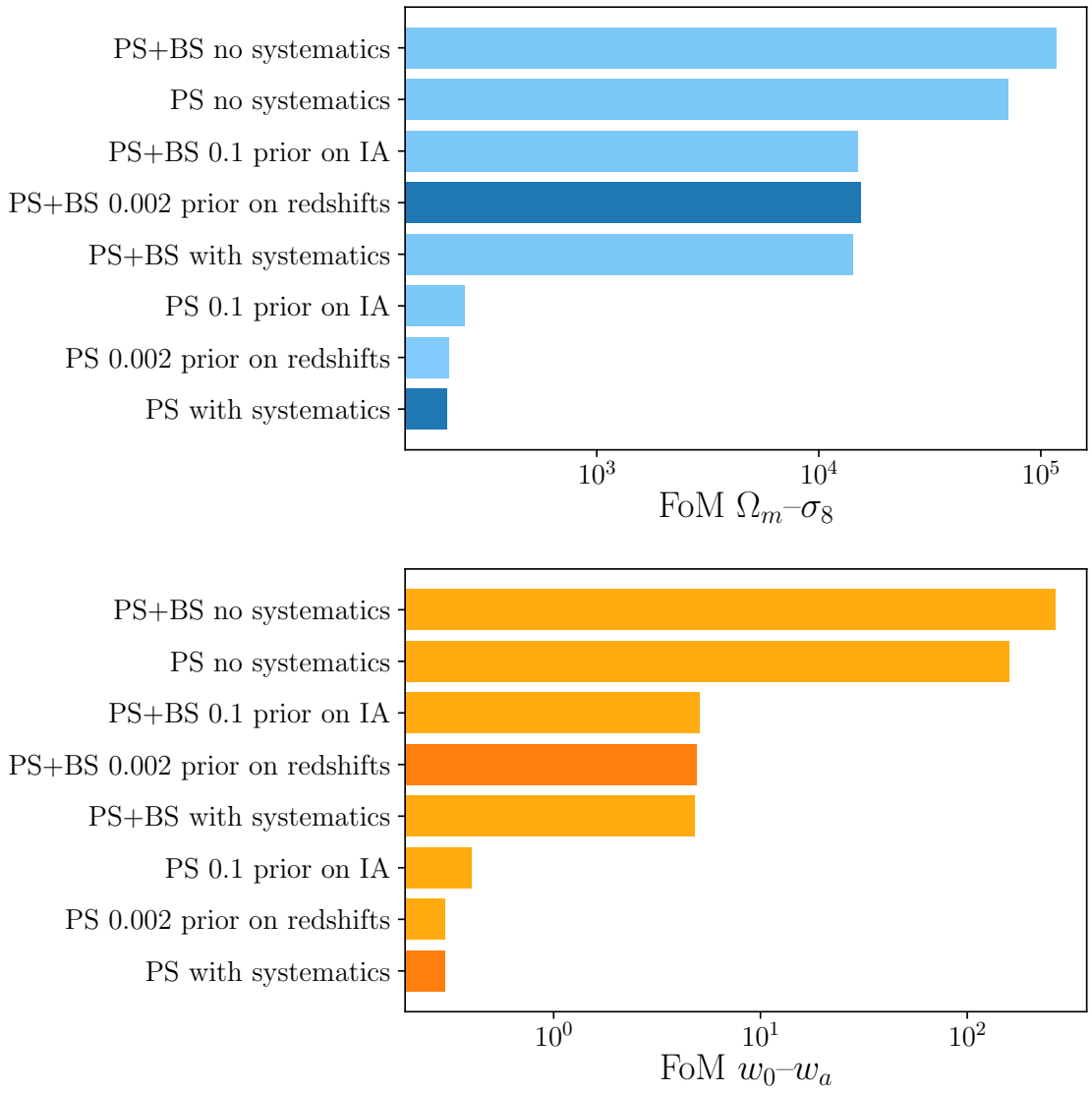

Figure 3.22: Summary of the findings in this section showing the effect of using the bispectrum as well as the power spectrum to constrain parameters and the further effect of priors on the nuisance parameters. *Top*: The $\Omega_{\rm m}$–σ_8 figure of merit. *Bottom*: The w_0–w_a figure of merit. We highlight two cases: power spectrum only with no knowledge of systematics, and power spectrum and bispectrum combined with 0.002 priors on the redshift parameters.

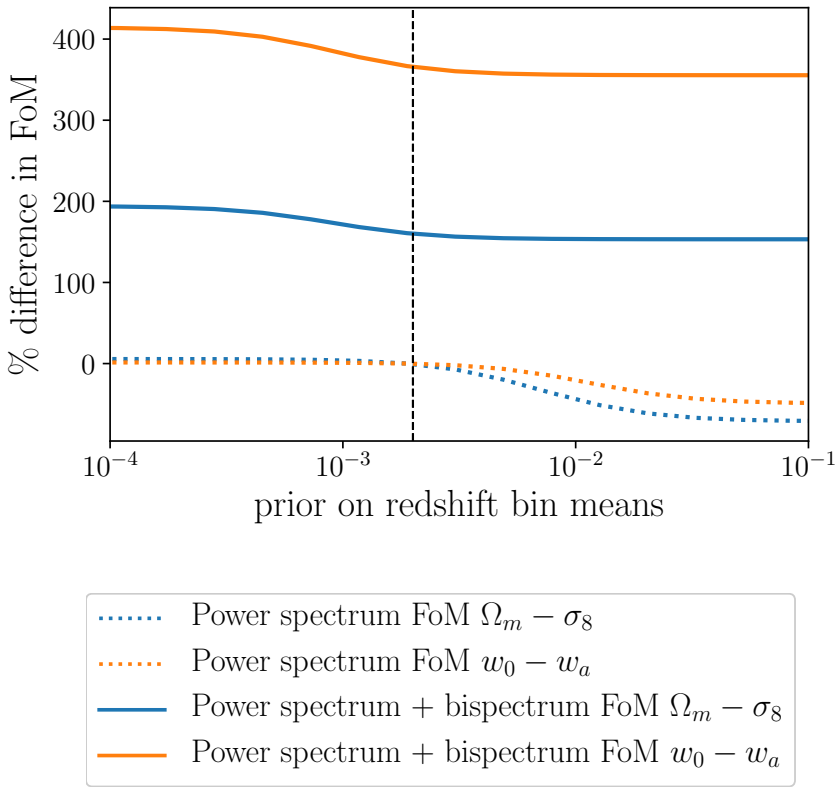

Figure 3.23: Based on the same information as Figure 3.22, we compare each figure of merit with the figure of merit obtained with the power spectrum only and priors of 0.002 on the redshift parameters, the requirement for redshift accuracy from the Euclid Definition Study Report (Laureijs et al. 2011). The vertical dashed line indicates the *Euclid* requirement.

3.4.4 Conclusions - systematics

In the context of a *Euclid*-like tomographic weak lensing survey we have considered two major sources of systematic uncertainty: contamination by intrinsic alignments which adds additional terms to the cosmic shear power spectrum and bispectrum, and uncertainty in the mean redshifts of the tomographic bins. We model the intrinsic alignment bispectrum using a variant of the linear alignment model developed for the power spectrum, with two free parameters: the intrinsic alignment amplitude and a power law dependence on redshift. We show that with our modelling assumptions intrinsic alignments affect the power spectrum much more than the bispectrum and thus the bispectrum can be expected to help alleviate the effect of this systematic error. We find similar results when we consider uncertainty in the means of redshift bins.

Whether we consider the power spectrum only or the combined power spectrum and bispectrum, the presence of these systematic uncertainties causes an order of magnitude decrease in both the

$\Omega_{\rm m}$–σ_8 and w_0–w_a figures of merit. However, compared to using the power spectrum only, combining the bispectrum and power spectrum has a very beneficial effect on the figures of merit. Much more can be gained from using the bispectrum than from using tighter priors with the power spectrum only. This is true even though we used a 'cut down' bispectrum which depends only on equilateral triangles. Using more triangle configurations could be expected to produce even greater gains.

The self-calibration regime extends to around 0.001 for redshift bin means. Since the *Euclid* requirement for redshift accuracy is 0.002, this suggests that self-calibration is a realistic, and helpful, possibility. There is no similar specification for intrinsic alignment accuracy, so it is hard to judge whether the self-calibration value of around 0.1 is consistent with the accuracy which can be achieved in practice. However our results do not rule out the possibility of self-calibration.

3.5 Discussion and further work

The results in this chapter suggest that in principle using three-point statistics in a *Euclid*-like survey could be very beneficial, especially for controlling systematic uncertainties where we have shown there is considerable potential for self-calibration. We discuss here possible further work to extend our results, to resolve some discrepancies with the literature, and to address practical challenges before our findings can be applied to a real survey.

Complexity of the bispectrum covariance

Any potential gain from the bispectrum must be balanced against the extra cost of measuring three-point statistics. It is common to tackle this by simplifying the bispectrum and its covariance matrix in ways which are expected to preserve most of the information content. For example we ignored the 'standard' non-Gaussian terms of the covariance, and showed that is safe to use only the one-halo term of the supersample covariance. We also showed that the bispectrum covariance need only be based on two or three tomographic bins, so long as the power spectrum is based on many bins. On the other hand we considered only equilateral triangles which certainly means we threw away some useful information, for example from squeezed triangles (Barreira 2019).

An alternative to ad hoc simplification is data compression. A standard approach is principal component analysis which was employed by Kayo et al. (2012) and Rizzato et al. (2019). Both concluded that a small fraction of eigenmodes carry nearly all the weak lensing bispectrum information. A promising alternative is the Karhunen-Loève algorithm (Tegmark et al. 1997).

Gualdi et al. (2018, 2019) showed that for the galaxy bispectrum this algorithm can reduce the dimension of the data vector to the number of cosmological parameters without significant loss of information. This has not been investigated for the weak lensing bispectrum and could be the best way forward in practice.

We discount other methods of accessing non-Gaussian weak-lensing information, for example peak counts (Kratochvil et al. 2010; Liu et al. 2015; Kacprzak et al. 2016) or Minkowski functionals (Kratochvil et al. 2012; Petri et al. 2013). Although they may bypass some of the complexities of the bispectrum these have other limitations. For example it is not clear how the modelling of systematics could be addressed.

Intrinsic alignment bispectra

Our model of the intrinsic alignment bispectra, based on the linear alignment model, is in disagreement with results obtained from simulations by Semboloni et al. (2008) in several respects. Most conspicuously, these authors found that intrinsic alignments affect three-point weak lensing statistics more strongly than two-point statistics, while we find the opposite. They also report that the III signal is negative whereas our model suggests it is positive (see also Figure 3.11). Nevertheless both studies find that intrinsic alignments affect the bispectrum differently from the power spectrum so support the idea of using the bispectrum to control systematics.

The discrepancies do not invalidate our method or general conclusions about self-calibration. The linear alignment model works well for two-point statistics. In principle there is no reason why our straightforward extension should not be valid for three-point statistics at the large, linear scales we are interested in where the halo model and perturbation theory work well. Semboloni et al. (2008) is based on simulations which are now quite old and assumes a much smaller survey than *Euclid*. The best way to settle the question would be to repeat their analysis using data from more modern simulations such as the Euclid Flagship Mock Galaxy Catalogue[6], or survey data such as the Dark Energy Survey Instrument Bright Galaxy Survey[7].

It would also be worthwhile to explore other approaches to modelling intrinsic alignments and where necessary to extend these to three-point statistics. For example Blazek et al. (2015) developed a model which goes beyond the linear alignment model to include nonlinear tidal effects. This approach was further developed by Blazek et al. (2019) to include tidal torquing as well as tidal alignment thus catering for spiral as well as elliptical galaxies. Vlah et al. (2020) recently developed

[6]https://sci.esa.int/web/euclid/-/59348-euclid-flagship-mock-galaxy-catalogue
[7]https://www.desi.lbl.gov/the-desi-survey/

an intrinsic alignment model based on perturbative effective field theory for both two-point and three-point statistics.

Theoretical modelling of the bispectrum

In this chapter we model the matter bispectrum using the fitting formula developed by Gil-Marín et al. (2012). This model has been used recently for example to constrain the sum of the neutrino mass (Coulton et al. 2019), and to investigate the bispectrum in modified gravity at non-linear scales (Bose et al. 2019). However the formula is known to have deficiencies, due partly to the limited size and scope of available simulations at the time it was developed. In particular it is known to overestimate the squeezed bispectrum (Namikawa 2016).

More recently Takahashi et al. (2019) have developed a new more accurate formula, named BiHalofit, based on high-resolution simulations. This is specifically designed to provide predictions at small scales and takes account of baryonic effects such as radiative transfer and active galactic nuclei feedback. Use of this formula would reduce modelling uncertainties, allow us to extend our results to smaller scales with more confidence and allow further investigation of the robustness of the results to changes in $\ell_{\rm max}$.

Multiplicative shear bias

It would be useful to explore whether the bispectrum can help with self-calibration of the bias due to the presence of noise in galaxy shape measurements, another major source of systematic uncertainty in weak lensing. This bias has a dominant multiplicative factor (Heymans et al. 2006), as well as a lesser additive term, and is usually calibrated using image simulations. Future *Euclid*-like surveys will require the accuracy of calibrated shapes to be around 0.1% (Cropper et al. 2013). To date many analyses of survey data have calibrated multiplicative bias from catalogues of image simulations (Heymans et al. 2012; Hildebrandt et al. 2016; Hikage et al. 2019) but an alternative method, which lends itself to the self-calibration approach, is to treat the multiplicative factors as nuisance parameters (Abbott et al. 2018).

Baryonic effects

Recently Foreman et al. (2019) reported a baryonic feature in the matter bispectrum which is not present in the power spectrum and attributed this to decreased feedback from active galactic nuclei at late times. If the result could be extended to weak lensing (using the BiHalofit model) this

would provide another justification for combining the bispectrum and power spectrum - to control systematic uncertainties from baryonic effects.

Choice of three-point statistic

In this chapter we have worked exclusively in Fourier space but in a real survey it can be easier to work in real space. It would be worth investigating statistics which are derived from three-point correlation functions, analogous to commonly used derived two-point real-space statistics such as aperture mass statistics (Schneider 1996) or complete orthogonal sets of E/B-mode integrals (COSEBIs) (Schneider et al. 2010). COSEBIs were designed to cleanly separate the 'gradient component' E-modes from the 'curl component' B-modes of the shear signal. This is desirable since the detection of B-modes strongly indicates the presence of systematics. However Shi et al. (2014) found that three-point aperture mass statistics separate E- and B-modes adequately, so it is not clear that three-point COSEBIs have any advantage.

Overall, despite the many unresolved issues, our results establish that weak lensing three-point statistics should be a useful tool for future surveys. We have shown that using the bispectrum as well as the power spectrum can help improve cosmological parameter constraints and that there is considerable scope for self-calibration of some of the main systematic uncertainties which affect weak lensing.

Nevertheless many challenges need to be overcome if three-point statistics are to be used successfully. In most cases this is not as daunting as it first seems. It should often be possible to build on solid work which has been done using two-point statistics - all the standard methods developed and implemented for current surveys. In short, the prospects for using three-point weak lensing statistics in next-generation surveys are very encouraging.

Appendices

3A Weak lensing covariance

In this appendix we give expressions for the components of the convergence power spectrum and bispectrum covariance for a single redshift bin. We assume a survey with area Ω_s in steradians and consider angular bins of width $\Delta\ell_i$ centred on the values ℓ_i. Thus $l_i - \Delta\ell/2 \leq |\boldsymbol{\ell}_i| \leq l_i + \Delta\ell/2$. Throughout we assume that the Limber and flat-sky approximations are valid. Further details and derivations can be found in Takada and Jain (2009), Kayo and Takada (2013) and Sato and Nishimichi (2013). Rizzato et al. (2019) gives all permutations of terms in these covariances for a tomographic survey.

Gaussian covariance

The Gaussian part of the convergence power spectrum covariance is

$$\mathrm{Cov}[P^\kappa(\ell_1), P^\kappa(\ell_2)]_\mathrm{G} = \frac{2\delta^\mathrm{K}_{\ell_1\ell_2}}{N_\mathrm{pairs}(\ell_1)} P^\kappa(\ell_1) P^\kappa(\ell_2) \,, \tag{3.105}$$

where $\delta^\mathrm{K}_{\ell_1\ell_2}$ is the Kronecker delta which is 1 if $\ell_1 = \ell_2$ and ℓ_1 is within the bin width $\Delta\ell_1$, and zero otherwise. $N_\mathrm{pairs}(\ell_1)$ is the number of independent pairs of modes within the bin width.
The Gaussian part of the convergence bispectrum covariance is

$$\begin{aligned}\mathrm{Cov}[B^\kappa(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \boldsymbol{\ell}_1), B^\kappa(\boldsymbol{\ell}_4, \boldsymbol{\ell}_5, \boldsymbol{\ell}_6)]_\mathrm{G} = \frac{\Omega_\mathrm{s}}{N_\mathrm{trip}(\ell_1, \ell_2, \ell_3)} P^\kappa(\ell_1) P^\kappa(\ell_2) P^\kappa(\ell_3) \\ \times [\delta^\mathrm{K}_{\ell_1\ell_4}\delta^\mathrm{K}_{\ell_2\ell_5}\delta^\mathrm{K}_{\ell_3\ell_6} + \delta^\mathrm{K}_{\ell_1\ell_4}\delta^\mathrm{K}_{\ell_2\ell_6}\delta^\mathrm{K}_{\ell_3\ell_5} + \delta^\mathrm{K}_{\ell_1\ell_5}\delta^\mathrm{K}_{\ell_2\ell_4}\delta^\mathrm{K}_{\ell_3\ell_6} \\ + 3 \text{ perms}] \,,\end{aligned} \tag{3.106}$$

where $N_\mathrm{trip}(\ell_1, \ell_2, \ell_3)$ is the number of triplets of modes which form triangles of side lengths ℓ_1, ℓ_2 and ℓ_3 within the specified bin widths $\Delta\ell_i$.

N_pairs and N_trip can be approximated by (Takada and Bridle 2007; Joachimi et al. 2009)

$$N_\mathrm{pairs}(\ell) = \frac{\Omega_\mathrm{s}\ell\Delta\ell}{2\pi} \,, \tag{3.107}$$

$$N_\mathrm{trip}(\ell_1, \ell_2, \ell_3) = \frac{\Omega_\mathrm{s}^2 \ell_1\ell_2\ell_3\Delta\ell_1\Delta\ell_2\Delta\ell_3}{2\pi^2\sqrt{2\ell_1^2\ell_2^2 + 2\ell_2^2\ell_3^2 + 2\ell_3^2\ell_1^2 - \ell_1^4 - \ell_2^4 - \ell_3^4}} \,. \tag{3.108}$$

'Standard' non-Gaussian covariance

The in-survey non-Gaussian part of the convergence power spectrum covariance is

$$\mathrm{Cov}[P^\kappa(\ell_1), P^\kappa(\ell_2)]_{\mathrm{NG}} = \frac{2\pi}{\Omega_\mathrm{s}} \int \frac{\mathrm{d}^2\boldsymbol{\ell}}{\ell_1 \Delta\ell_1} \int \frac{\mathrm{d}^2\boldsymbol{\ell}'}{\ell_2 \Delta\ell_2} T^\kappa(\boldsymbol{\ell}, -\boldsymbol{\ell}, \boldsymbol{\ell}', -\boldsymbol{\ell}') \;, \tag{3.109}$$

where T^κ is the convergence trispectrum. The integrals are over all wavevectors which are within the bin width $\Delta\ell$ around $\boldsymbol{\ell}$ or $\boldsymbol{\ell}'$. The in-survey non-Gaussian part of the convergence bispectrum covariance is

$$\begin{aligned}
&\mathrm{Cov}[B^\kappa(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \boldsymbol{\ell}_3), B^\kappa(\boldsymbol{\ell}_4, \boldsymbol{\ell}_5, \boldsymbol{\ell}_6)]_{\mathrm{NG}} = \qquad (3.110)\\
&\quad \frac{2\pi}{\Omega_\mathrm{s}} B^\kappa(\ell_1, \ell_2, \ell_3) B^\kappa(\ell_4, \ell_5, \ell_6) \left[\frac{\delta^\mathrm{K}_{\ell_1\ell_4}}{\ell_1 \Delta\ell_1} + \frac{\delta^\mathrm{K}_{\ell_1\ell_5}}{\ell_1 \Delta\ell_1} + 7 \text{ perms} \right] \\
&\quad + \frac{2\pi}{\Omega_\mathrm{s}} \left[\frac{\delta^\mathrm{K}_{\ell_1\ell_4}}{\ell_1 \Delta\ell_1} T^\kappa(\boldsymbol{\ell}_2, \boldsymbol{\ell}_3, \boldsymbol{\ell}_5, \boldsymbol{\ell}_6) P^\kappa(\ell_1) + \frac{\delta^\mathrm{K}_{\ell_1\ell_5}}{\ell_1 \Delta\ell_1} T^\kappa(\boldsymbol{\ell}_2, \boldsymbol{\ell}_3, \boldsymbol{\ell}_4, \boldsymbol{\ell}_6) P^\kappa(\ell_1) + 7 \text{ perms} \right] \\
&\quad + \frac{1}{\Omega_\mathrm{s}} \int \frac{\mathrm{d}\psi}{2\pi} P_6^\kappa(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \boldsymbol{\ell}_3, \boldsymbol{\ell}_4, \boldsymbol{\ell}_5, \boldsymbol{\ell}_6; \psi) \;,
\end{aligned}$$

where P_6^κ is the pentaspectrum. Triangle conditions mean that the P_6^κ term depends on two triangles; ψ is the angle between these triangles.

Supersample covariance

The weak lensing power spectrum and bispectrum supersample covariance are

$$\begin{aligned}
&\mathrm{Cov}[P^\kappa(\ell_1), P^\kappa(\ell_2)]_{\mathrm{SSC}} = \qquad (3.111)\\
&\quad \frac{1}{\Omega_\mathrm{s}^2} \int \mathrm{d}\chi \; q^4(\chi) \chi^{-6} \frac{\partial P_\delta(\ell_1/\chi; \chi)}{\partial \delta_b} \frac{\partial P_\delta(\ell_2/\chi; \chi)}{\partial \delta_b} \int \frac{\mathrm{d}^2\boldsymbol{\ell}}{(2\pi)^2} |\tilde{W}(\boldsymbol{\ell})|^2 P_\mathrm{L}(\ell/\chi; \chi) \;, \\
&\mathrm{Cov}[B^\kappa(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \boldsymbol{\ell}_3), B^\kappa(\boldsymbol{\ell}_4, \boldsymbol{\ell}_5, \boldsymbol{\ell}_6)]_{\mathrm{SSC}} = \qquad (3.112)\\
&\quad \frac{1}{\Omega_\mathrm{s}^2} \int \mathrm{d}\chi \; q^6(\chi) \chi^{-10} \frac{\partial B_\delta(, \boldsymbol{\ell}_1/\chi, \boldsymbol{\ell}/\chi_2, \boldsymbol{\ell}_3/\chi; \chi)}{\partial \delta_b} \frac{\partial B_\delta(, \boldsymbol{\ell}_4/\chi, \boldsymbol{\ell}_5/\chi, \boldsymbol{\ell}_6/\chi; \chi)}{\partial \delta_b} \\
&\qquad\qquad \times \int \frac{\mathrm{d}^2\boldsymbol{\ell}}{(2\pi)^2} |\tilde{W}(\boldsymbol{\ell})|^2 P_\mathrm{L}(\ell/\chi; \chi) \;,
\end{aligned}$$

where $q(\chi)$ is the lensing weight function (Equation 3.9), $\tilde{W}$ is the Fourier transform of the survey mask and P_L is the linear matter power spectrum.

3B Perturbation theory squeezed 'equilateral' trispectrum

To derive an approximation for the perturbation theory trispectrum with three equal short modes and one long mode $\mathbf{q}$ we start from Equation 3.69 for the general trispectrum $T_{\rm PT}$ and Equations 3.70 and 3.71 for its components T_a and T_b in the squeezed configuration. Three terms of Equation 3.70 are zero in the limit $\mathbf{q} \to 0$ because $F_2(\mathbf{k}, -\mathbf{k}) = 0$ and one term of Equation 3.71 is zero because $F_3(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = 0$ if $\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 = 0$. We set $|\mathbf{k_i}| = k$ for $i = 1, 2, 3$ and $\mathbf{k}_i \cdot \mathbf{k}_j = k^2/2$ for $i \neq j$. Then Taylor-expanding, for example, $P(|\mathbf{k} + \mathbf{q}|)$ to first order in q/k gives

$$P(|\mathbf{k} + \mathbf{q}|) \approx P(k) + \frac{\mathbf{k} \cdot \mathbf{q}}{k} \frac{\partial P(k)}{\partial k}. \tag{3.113}$$

Similarly, as examples, expanding $F_2(-\mathbf{q}, \mathbf{k} - \mathbf{q})$ gives

$$F_2(-\mathbf{q}, \mathbf{k} - \mathbf{q}) = \frac{4(\mathbf{k} \cdot \mathbf{q})^3}{7k^4q^2} + \frac{2(\mathbf{k} \cdot \mathbf{q})^2}{7k^2q^2} - \frac{15(\mathbf{k} \cdot \mathbf{q})}{14k^2} - \frac{\mathbf{k} \cdot \mathbf{q}}{2q^2} + \frac{17}{14} \tag{3.114}$$

$$\approx \frac{2(\mathbf{k} \cdot \mathbf{q})^2}{7k^2q^2} - \frac{(\mathbf{k} \cdot \mathbf{q})}{2q^2} + \frac{17}{14}\,, \tag{3.115}$$

and expanding $F_3(\mathbf{k}, \mathbf{k}, -\mathbf{q})$ gives

$$F_3(\mathbf{k}, \mathbf{k}, -\mathbf{q}) = -\frac{(\mathbf{k} \cdot \mathbf{q})^3}{3k^4q^2} + \frac{16(\mathbf{k} \cdot \mathbf{q})^2}{21k^2q^2} - \frac{2(\mathbf{k} \cdot \mathbf{q})}{3k^2} - \frac{4(\mathbf{k} \cdot \mathbf{q})}{3q^2} + \frac{26}{21} \tag{3.116}$$

$$\approx \frac{16(\mathbf{k} \cdot \mathbf{q})^2}{21k^2q^2} - \frac{4(\mathbf{k} \cdot \mathbf{q})}{3q^2} + \frac{26}{21}\,. \tag{3.117}$$

Expanding all terms similarly, combining them and keeping only leading order terms results in

$$T_a^{\rm equi} = \frac{9P(k)^2P(q)}{98k^2q^2}[17(\mathrm{k} \cdot \mathrm{q})^2 + 42k^2\mathrm{k} \cdot \mathrm{q} + 32k^2q^2] \tag{3.118}$$

$$- \frac{9P(k)P'(k)P(q)}{98k^3q^2}[(\mathrm{k} \cdot \mathrm{q})^3 + 14k^2(\mathrm{k} \cdot \mathrm{q})^2 + 6k^2q^2]$$

$$T_b^{\rm equi} = \frac{P(k)^2P(q)}{84k^2q^2}[106(\mathrm{k} \cdot \mathrm{q})^2 - 216k^2\mathrm{k} \cdot \mathrm{q} + 53k^2q^2] \tag{3.119}$$

$$+ \frac{P(k)P'(k)P(q)}{126k^3q^2}[70(\mathrm{k} \cdot \mathrm{q})^3 - 216k^2(\mathrm{k} \cdot \mathrm{q})^2 + 93k^2q^2]$$

$$T_{\rm PT}^{\rm equi} = \frac{P(k)^2P(q)}{98k^2q^2}[1354(\mathrm{k} \cdot \mathrm{q})^2 + 2055k^2q^2] \tag{3.120}$$

$$- \frac{P(k)P'(k)P(q)}{147k^3q^2}[436(\mathrm{k} \cdot \mathrm{q})^3 - 2268k^2(\mathrm{k} \cdot \mathrm{q})^2 + 327k^2q^2]\,,$$

where $P'(k) = \partial P(k)/\partial k$. Although T_a and T_b include terms in $\mathrm{k} \cdot \mathrm{q}/q^2$ which are divergent as $\mathbf{q} \to 0$, these cancel out in the final expression for the squeezed trispectrum.

3C Matter response function for equilateral bispectra

In this appendix we derive the perturbation theory bispectrum response for equilateral triangles from the general result given by Chan et al. (2018) which is

$$\left.\frac{\partial B(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3|\delta_b)}{\partial\delta_b}\right|_{\delta_b=0} = \frac{433}{126}B_{\rm m}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3) + \frac{5}{126}B_{\rm G}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3) - \frac{1}{3}\sum_{i=1}^{3}\frac{\partial B_{\rm m}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)}{\partial\ln k_i}\,, \tag{3.121}$$

where $B_{\rm PT}$ is the tree-level matter bispectrum and $B_{\rm G}$ is identical to $B_{\rm PT}$ but with the F_2 kernels replaced by G_2. This can be rewritten in terms of the cosines, μ_{ij}, between $\mathbf{k}_i$ and $\mathbf{k}_j$ as

$$\begin{aligned}\frac{\partial B_{\rm PT}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)}{\partial\delta_b} = \\ \left[\frac{2180k_1^2k_2 + (441k_1^2 + 1533k_1^3 - 441k_2^2 + 1533k1k_2^2)\mu_{12} + 886k_1^2k_2\mu_{12}^2}{441k_1^2k_2}\right] \\ \times P(k_1)P(k_2) + 2\text{ perms.} \\ - \left[\frac{10k_1k_2 + 7(k_1^2+k_2^2)\mu_{12} + 4k_1k_2\mu_{12}^2}{7k_1^2k_2}\right]\left[\frac{\partial P(k_1)}{\partial\ln k_1}P(k_2)P(k_3) + 2\text{ perms.}\right]\end{aligned} \tag{3.122}$$

For equilateral triangles we have $k_1 = k_2 = k_3 = k$ and $\mu_{ij} = 1/2$, leading to

$$\frac{\partial B_{\rm PT}^{\rm equi}}{\partial\delta_b} = \left[\frac{2623}{98} - \frac{36}{7}\frac{\partial\ln P(k)}{\partial\ln k}\right]P(k)^2 \tag{3.123}$$

and the halo model bispectrum response given by Equation 3.50 becomes:

$$\begin{aligned}\frac{\partial B}{\partial\delta_b} = I_3^1(k,k,k) + \frac{47}{7}\left[I_1^1(k)I_2^1(k,k)P(k)\right] \\ + (I_1^1(k))^3\left[\frac{2623}{98} - \frac{36}{7}\frac{\partial\ln P(k)}{\partial\ln k}\right]P(k)^2\,.\end{aligned} \tag{3.124}$$

3D Fisher matrix analysis of systematic uncertainties

We show here Fisher matrix plots corresponding to the analysis of systematic uncertainties in Section 3.4.

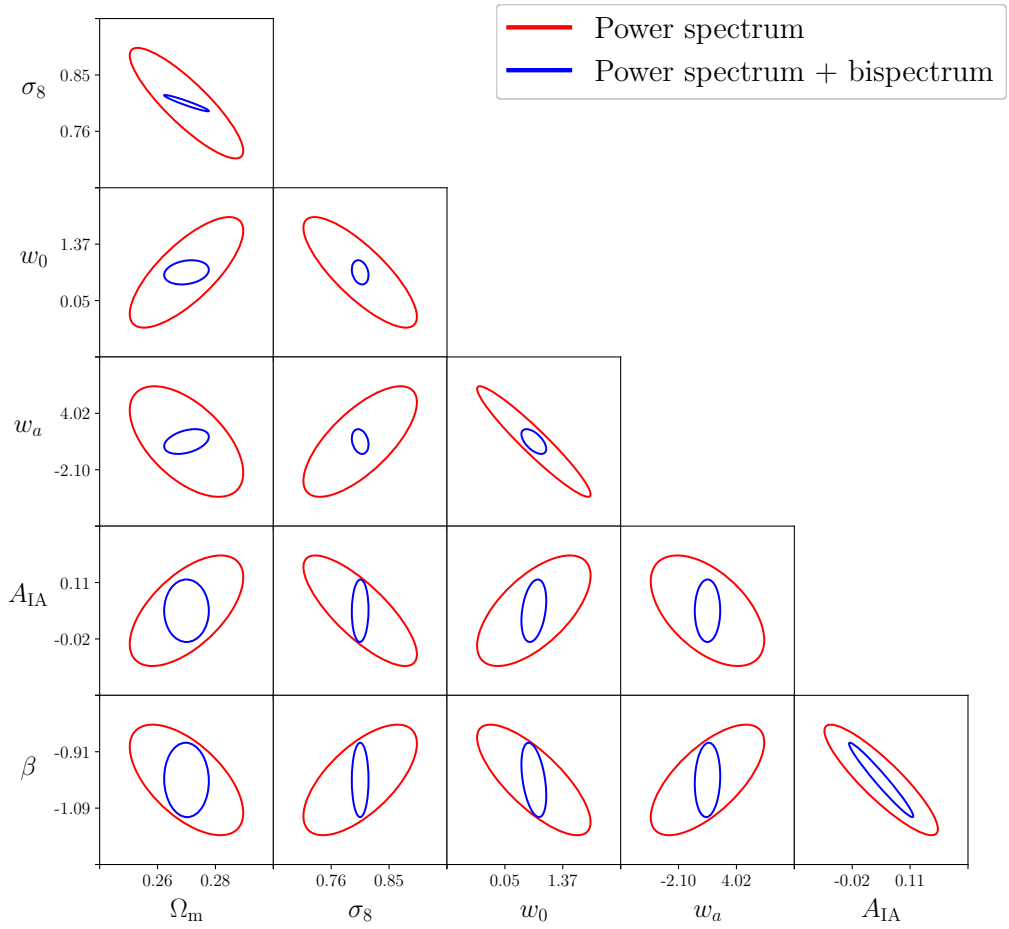

Figure 3.24: Fisher matrix analysis including intrinsic alignment nuisance parameters. Parameter constraints from weak lensing power spectrum and power spectrum plus bispectrum. Constraints using only the power spectrum covariance are shown in red and using the power spectrum plus the bispectrum are shown in blue. Five tomographic bins between $z = 0.2$ and $z = 2.0$ are used. Gaussian shape noise is included.

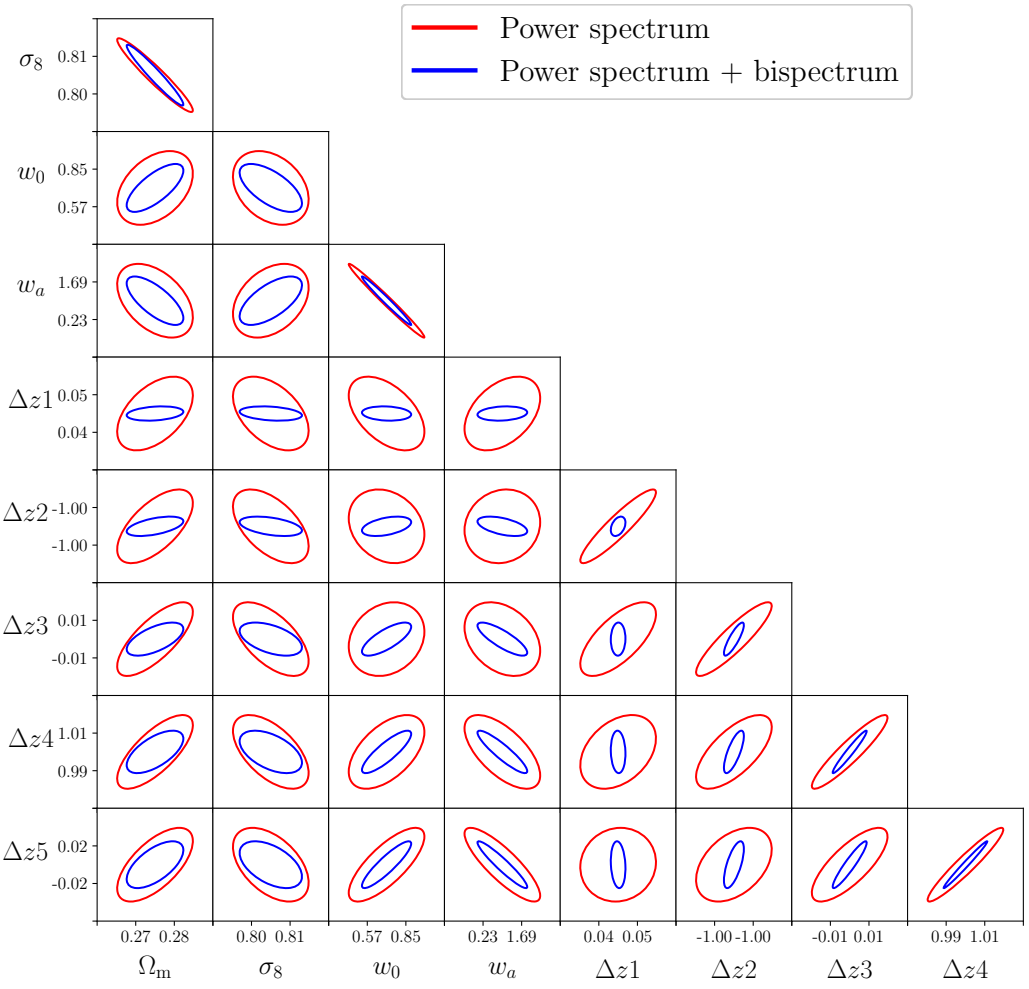

Figure 3.25: Fisher matrix analysis including redshift nuisance parameters. Parameter constraints from weak lensing power spectrum and power spectrum plus bispectrum. Constraints using only the power spectrum covariance are shown in red and using the power spectrum plus the bispectrum are shown in blue. Five tomographic bins between $z = 0.2$ and $z = 2.0$ are used. Gaussian shape noise is included.

4 Likelihoods for weak lensing

4.1 Introduction

This chapter discusses two issues which need to be considered when estimating cosmological parameters in a Bayesian framework: is it valid to use a Gaussian power spectrum likelihood, and if a Gaussian likelihood is used, does the parameter-dependence of the covariance matrix matter? The discussion is mainly framed around weak lensing but also includes results from the CMB and large-scale structure literature where the issues have often been investigated more thoroughly.

Section 4.2 focuses on the theoretical background. We discuss the true form of the power spectrum likelihood and show that it is not Gaussian even if the underlying density field is Gaussian. We recap the customary approach to Bayesian inference in cosmology and then discuss how to choose an appropriate likelihood. We show that in most cases the exact likelihood has a gamma distribution, then describe some alternative non-Gaussian likelihoods which approximate this. We also outline other strategies which allow a Gaussian likelihood to be used or which avoid explicitly evaluating the likelihood at all. We then identify and discuss variants of Gaussian likelihoods which are commonly used for real surveys.

In Section 4.3 we turn to likelihoods for weak lensing correlation functions, building on recent modelling work by Sellentin et al. (2018). We consider the joint distributions of correlation functions in two angular bins and confirm the conclusion that likelihoods for weak lensing correlation functions are also non-Gaussian.

In Section 4.4 we consider the parameter-dependence of the covariance matrix in a Gaussian likelihood. We first summarise an argument in Carron (2013) which shows why it is formally incorrect to recalculate the covariance matrix at each sampled point when using a Gaussian likelihood. We then review how the parameter-dependence of the covariance matrix has been treated in practice, and the conclusions which have been drawn (correctly or incorrectly). In the subsequent discussion we start from the assumption that for real surveys the only practicable approach may be to use a Gaussian likelihood, and therefore a fixed covariance matrix. We develop a simple iterative method to derive a suitable fiducial cosmology at which to calculate the covariance, based on an emulator for the power spectrum covariance matrix.

4.2 Choice of likelihood

4.2.1 Bayesian inference

As discussed in Chapter 1, the starting point for all Bayesian inference is Bayes' theorem.

$$P(\boldsymbol{\theta}|\mathbf{d}) = \frac{P(\mathbf{d}|\boldsymbol{\theta})P(\boldsymbol{\theta})}{P(\mathbf{d})} \,, \tag{4.1}$$

where $\mathbf{d} = (d_1, d_2, \cdots, d_n)$ is the data vector and $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_m)$ is the vector of model parameters. $P(\boldsymbol{\theta}|\mathbf{d})$ is the posterior. The denominator $P(\mathbf{d})$, the Bayesian evidence, is not important for the discussion in this chapter since it does not depend on the parameters.

The inputs to the inference are firstly the likelihood of the experimentally observed data (given a chosen set of parameters) and secondly priors on the parameters, determined from pre-existing knowledge. Then Bayes' theorem is used to infer the posterior distribution of the parameters and hence the bestfit parameter values (typically given by either the maximum or the mean of the posterior in parameter space). Often the priors are assumed to be uniform or 'flat' to reflect a lack of knowledge, on the assumption that flat priors are uninformative (which may not in fact be true, as discussed for example in Trotta (2017)). It follows that the likelihood carries a great deal of weight in the inference process and needs to be specified appropriately and accurately.

To estimate the values of the parameters which maximise the posterior, the likelihood and priors are typically fed into a Monte Carlo Markov Chain (MCMC) code or another sampling method such as nested sampling. These samplers can estimate not only the maximum likelihood values, but also the full shape of the posterior distribution. Crucially they require the likelihood to be estimated many thousands of times, so it is necessary to balance the correctness of the likelihood against the computational cost of calculating it.

4.2.2 Exact power spectrum likelihood

We want to estimate cosmological parameters from data which describes some density field, for example CMB temperature or polarisation anisotropies, cosmic shear, or galaxy clustering measurements. In the case of the CMB the underlying fields are Gaussian, but in most cosmological cases the field is not Gaussian. In particular the development of non-linear structure at small scales and late times makes both the matter and weak lensing fields non-Gaussian.

Following Carron (2013) we distinguish between 'field level' analysis based on the underlying density fluctuations, which we assume can be expressed in terms of spherical harmonics $a_{\ell m}$, and 'estimator level' analysis based on a summary statistic.

In principle we can carry out the likelihood analysis at the field level (for example Hinshaw et al. (2007)). If the field is Gaussian the $a_{\ell m}$ have zero mean and variance equal to the power spectrum of the fluctuations, C_ℓ. The likelihood based on the $a_{\ell m}$ is also precisely Gaussian. However if the field is not Gaussian then we would expect the likelihood (at the field level) to be non-Gaussian as well.

Even if the field is Gaussian, working at the field level is often intractable because of the very large size of the data vector and the need to invert large matrices. Instead it is usual to compress the field data into summary statistics, normally the power spectrum or two-point correlation function. In the remainder of this section we discuss the power spectrum, C_ℓ. For one multipole ℓ this is defined by

$$\langle a_{\ell m} a_{\ell' m'} \rangle = \delta_{\ell \ell'} \delta_{m m'} C_\ell \, . \tag{4.2}$$

If the underlying field is Gaussian then the power spectrum contains exactly the same information as the field so this data compression should not in itself reduce the accuracy of the parameter estimation. This is not the case for non-Gaussian fields.

For simplicity we ignore complications such as sky coverage and measurement noise and define an estimator for the power spectrum by

$$\hat{C}_\ell = \frac{1}{2\ell + 1} \sum_{m=-\ell}^{\ell} |a_{\ell m}^2| \, . \tag{4.3}$$

Now we need a likelihood for the power spectrum rather than the field. The most common way to proceed, especially in the weak lensing literature, is to assume that the likelihood of the power spectrum is Gaussian. For a single multipole ℓ this has the form

$$L_{\rm G}(\hat{C}_\ell | C_\ell) = \frac{1}{\sqrt{2\pi {\rm det} \Sigma_\ell}} \exp\left(- \frac{(C_\ell - \hat{C}_\ell)^2}{2 {\rm det} \Sigma_\ell} \right) , \tag{4.4}$$

where Σ_ℓ is the power spectrum covariance.

Strictly speaking the assumption that the power spectrum likelihood is Gaussian is incorrect even if the underlying field is Gaussian. $\hat{C}_\ell$ is the sum of squares of identically distributed random Gaussian variables. Such a sum has a gamma distribution, normally characterised by a shape

parameter k and a scale parameter θ. In terms of these parameters the probability density function of this distribution is

$$P(\hat{C}_\ell|C_\ell) = \exp\left(\frac{-\hat{C}_\ell}{\theta}\right)\frac{\hat{C}_\ell^{k-1}}{\theta^k\Gamma(k)}\,, \tag{4.5}$$

where $\Gamma(k)$ is the gamma function:

$$\Gamma(k) = \int_0^\infty x^{k-1}\exp(-x)\mathrm{d}x\,. \tag{4.6}$$

The gamma distribution has mean $k\theta$, variance $k\theta^2$ and skewness $2/\sqrt{k}$. In our case we have

$$k = \frac{2\ell+1}{2} \tag{4.7}$$

$$\theta = \frac{2C_\ell}{2\ell+1}\,, \tag{4.8}$$

so the mean of the distribution is C_ℓ, the variance is $2C_\ell^2/(2\ell+1)$ and the skewness is $2\sqrt{2/(2\ell+1)}$.

Thus using a Gaussian likelihood is formally incorrect and in principle could produce biased parameter estimates. The true likelihood is positively skewed. In weak lensing this skewness is most likely to affect estimates of Ω_m and σ_8 because these both contribute to the power spectrum amplitude.

4.2.3 Strategies for taking account of non-Gaussian likelihood

Given the difficulty of modelling the exact likelihood even in an ideal situation, various other strategies have been implemented or proposed in the literature. These can be broadly categorised as

- Do the analysis wholly or partly at the field level. This can be feasible for the CMB where the field is Gaussian (Hinshaw et al. 2007). Recent *Planck* analysis adopts a hybrid approach where large scales are analysed at field level but an estimator with Gaussian likelihood is used at small scales, for example Ade et al. (2014). However for weak lensing field-level analysis presents many practical difficulties, and in any case the likelihood is still non-Gaussian.

- Transform the underlying field so it is more Gaussian, for example using the transformations discussed in Box and Cox (1964). A more Gaussian field leads to a more Gaussian covariance and likelihood. For example Joachimi, Taylor and Kiessling (2011) applied such transformations to the convergence field.

- Devise a likelihood distribution which approximates the exact gamma distribution but is more computationally tractable. Several examples, mainly from CMB and galaxy clustering analyses, are discussed in Section 4.2.4.

- Make the likelihood more Gaussian by removing non-Gaussian contributions to the data vector. For example Lin et al. (2019) found that principal component analysis of their data vector removed much of the non-Gaussianity, and Sellentin et al. (2018) suggest a three-step process to remove B-mode contributions from cosmic shear data, resulting in a more symmetric likelihood.

- Use a Gaussian likelihood anyway and assume that it does not unduly bias parameter estimates. This is the approach normally adopted in weak lensing, for example Kilbinger et al. (2013); Hildebrandt et al. (2016); Köhlinger et al. (2017); Hikage et al. (2019); Troxel et al. (2018). It is a pragmatic choice based on computational considerations, but is also defended by appealing to the central limit theorem which ensures that even if the likelihood is not Gaussian it will approach Gaussianity as the number of modes increases. In Section 4.2.4 we discuss alternative formulations of Gaussian likelihoods which are commonly used.

- Take account of the non-Gaussianity of the field, at least partly, by including the full non-Gaussian terms in the covariance matrix Smith et al. (2006). As discussed in Chapter 3, these terms are generated by in-survey and super-survey mode-coupling (Takada and Hu 2013). They depend on four-point functions and are a direct consequence of the fact that the underlying field is not Gaussian. In current weak lensing analyses these terms are normally included as a matter of course if the covariance matrix is calculated analytically (Hildebrandt et al. 2016; Köhlinger et al. 2017; Hikage et al. 2019; Troxel et al. 2018; van Uitert et al. 2018). Alternatively if the covariance matrix is estimated from simulations then the non-Gaussianity is included implicitly, for example in Kilbinger et al. (2013).

For a real survey with complicated geometry the actual likelihood is very complex and may not be approximated well by any of these choices. This has prompted the development of so-called likelihood-free methods, for example approximate Bayesian computation, in which the likelihood is not specified explicitly but instead is estimated from simulations (Hartlap et al. 2009; Alsing et al. 2018; Leclercq 2018; Taylor et al. 2019).

4.2.4 Alternative likelihoods

In this section we discuss several likelihood functions which have been considered in the literature, including variants of the Gaussian distribution. A useful overview of CMB likelihood methods, published after this chapter was written, is Gerbino et al. (2019).

We assume flat priors and focus on the log posterior (rather than the likelihood) of a single multipole, $P(C_\ell|\hat{C}_\ell)$. In this view C_ℓ (the model) is a random variable which is a function of $\hat{C}_\ell$ (the data).

Ideally we would like both the likelihood and posterior to be good approximations of the true distributions. However to obtain unbiased estimates of cosmological parameters it is more important to match the true posterior probability.

Gamma distribution

A gamma-distributed log posterior with shape parameter $(2\ell+1)/2$ and scale parameter $2C_\ell/(2\ell+1)$ takes the form (up to a constant which can be ignored)

$$-2\ln P_\Gamma(C_\ell|\hat{C}_\ell) = (2\ell+1)\frac{\hat{C}_\ell}{C_\ell} + (2\ell+1)\ln C_\ell - (2\ell-1)\ln \hat{C}_\ell \tag{4.9}$$

$$-2\frac{\partial \ln P_\Gamma(C_\ell|\hat{C}_\ell)}{\partial C_\ell} = \left[-(2\ell+1)\frac{\hat{C}_\ell}{C_\ell^2} + (2\ell+1)\frac{1}{C_\ell}\right] \tag{4.10}$$

$$= (2\ell+1)\frac{C_\ell - \hat{C}_\ell}{C_\ell^2}\,, \tag{4.11}$$

so this peaks at the mean value $C_\ell = \hat{C}_\ell$.

The gamma distribution does not lend itself to the complications of real surveys and so is almost never used. Interestingly Sellentin et al. (2018) have argued that a likelihood based on the gamma distribution can and should be used with shear two-point correlation functions $\xi_\pm(\theta)$. If the power spectrum posterior is gamma-distributed, then transforming from Fourier to real space gives for the correlation function posterior

$$P_\Gamma(\xi_\pm(\theta)|\hat{\xi}_\pm(\theta)) = \sum_\ell \frac{\ell F(\ell\theta)}{2\pi} P_\Gamma(C_\ell|\hat{C}_\ell)\,, \tag{4.12}$$

where $F(\ell\theta)$ is a filter function (normally a Bessel function) corresponding to $\xi_\pm(\theta)$. We return to this in Section 4.3.

As a slight digression, the frequently-encountered chi-squared distribution is a special case of the gamma distribution where all the summed squared variables are drawn from a standard normal

distribution $\mathcal{N}(0,1)$. In multiple dimensions the gamma distribution generalises into the Wishart distribution which is the analogous likelihood distribution for more than one Gaussian field.

Likelihoods related to the lognormal distribution

This section summarises some salient points from comprehensive discussions of lognormal-based likelihoods in Percival and Brown (2006), Hamimeche and Lewis (2008) and Sun et al. (2013).

The lognormal distribution is a natural choice for an approximation which is related to the tractable Gaussian distribution but has non-zero skewness. It was first proposed by Bond et al. (2000) in the context of early CMB analysis. It has log posterior

$$-2\ln P_{\rm LN}(C_\ell|\hat{C}_\ell) = \frac{(2\ell+1)}{2}\left[\ln\left(\frac{\hat{C}_\ell}{C_\ell}\right)\right]^2 . \tag{4.13}$$

Figure 4.1, motivated by Figure 1 in Sun et al. (2013), compares the likelihood and posterior distributions for the exact gamma distribution, the lognormal approximation and a Gaussian distribution. The top row is for $\ell = 2$ and the bottom row has $\ell = 50$. The left panel shows the likelihood functions $L(\hat{C}_\ell|C_\ell)$ and the right panel shows the posterior distributions $P(C_\ell|\hat{C}_\ell)$. The likelihoods and posteriors are normalised so that their maxima are all equal to 1 and their means are also equal to 1. The quadrupole shows that the Gaussian posterior has a biased maximum-likelihood estimate and completely fails to match the shape of the exact likelihood and posterior. This mismatch reduces when more modes are included and the distributions become more Gaussian. In contrast, the lognormal approximation matches the gamma posterior quite well and produces the correct maximum-likelihood estimate.

Verde et al. (2003) found that the lognormal distribution was slightly biased for CMB measurements from the First Year Wilkinson Microwave Anisotropy Probe (WMAP)[1]. Instead these authors sought a more accurate approximation by writing $\hat{C}_\ell = (1+\epsilon)C_\ell$ and expanding around C_ℓ. Then, since $\hat{C}_\ell$ is a constant in the posterior, the log posterior from Equation 4.9 can be approximated as

$$-2\ln P_\Gamma(C_\ell|\hat{C}_\ell) \approx (2\ell+1)\left(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3}\right) . \tag{4.14}$$

[1] https://map.gsfc.nasa.gov

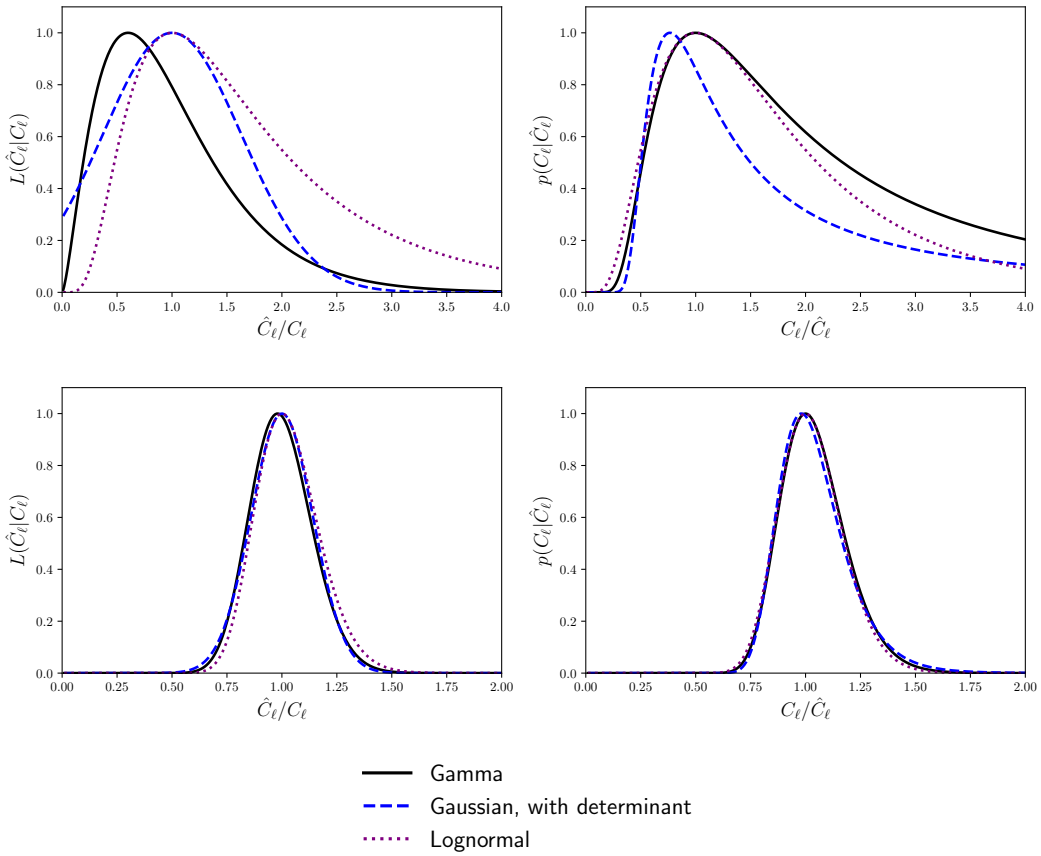

Figure 4.1: *Upper left*: The true gamma likelihood for $\ell = 2$, together with two approximations: lognormal and Gaussian. *Lower left*: As for upper left but for $\ell = 50$. *Upper right*: The posterior distributions corresponding to the likelihoods in the left panel for $\ell = 2$. *Lower right*: As for upper right but for $\ell = 50$. All likelihoods and posteriors are normalised to have a maximum-likelihood estimate of 1.0 and mean 1.0.

With the same expansion the Gaussian and lognormal posteriors, Equations 4.4 (ignoring the determinant term) and 4.13, become

$$-2\ln P_{\rm G}(C_\ell|\hat{C}_\ell) \approx (2\ell+1)\frac{\epsilon^2}{2} \tag{4.15}$$

$$-2\ln P_{\rm LN}(C_\ell|\hat{C}_\ell) \approx (2\ell+1)\left(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{2}\right). \tag{4.16}$$

This underlines the fact that the Gaussian distribution approximates the exact distribution up to $\mathcal{O}(\epsilon^2)$. Motivated by Equations 4.15 and 4.16, Verde et al. (2003) suggested the Gaussian plus lognormal approximation

$$-2\ln P_{\rm GLN}(C_\ell|\hat{C}_\ell) = \frac{1}{3}\ln P_{\rm G}(C_\ell|\hat{C}_\ell) + \frac{2}{3}\ln P_{\rm LN}(C_\ell|\hat{C}_\ell) \tag{4.17}$$

and found that this version estimated the amplitude of temperature fluctuations to better than 0.1% accuracy. Figure 4.2 compares this distribution with the gamma distribution.

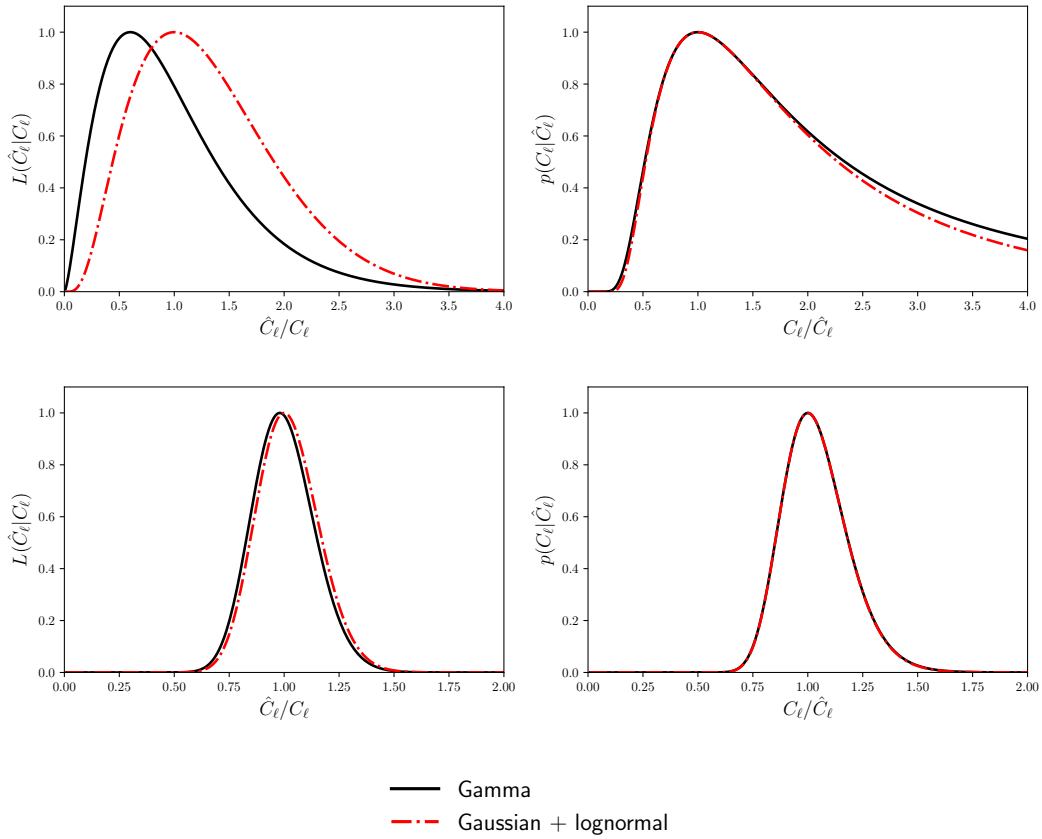

Figure 4.2: *Upper left*: The gamma likelihood for $\ell = 2$, together with the Gaussian plus lognormal distribution (Verde et al. 2003). *Lower left*: As for upper left but for $\ell = 50$. *Upper right*: The posterior distributions corresponding to the likelihoods in the left panel for $\ell = 2$. *Lower right*: As for upper right but for $\ell = 50$. All likelihoods and posteriors are normalised to have a maximum-likelihood estimate of 1.0 and mean 1.0.

Other candidate distributions have been suggested which also approximate the expansion in Equation 4.14 (Percival and Brown 2006). For example the offset lognormal distribution gives:

$$-2\ln P_{\rm OLN}(C_\ell|\hat{C}_\ell) = \frac{2\ell+1}{2\hat{C}_\ell^2}\left[\hat{C}_\ell(1+a)\ln\left(\frac{C_\ell+a\hat{C}_\ell}{\hat{C}_\ell+a\hat{C}_\ell}\right)\right]^2. \tag{4.18}$$

The free parameter a can be varied to fit a particular situation. Practical applications using this distribution include Sievers et al. (2003) (CMB) and Kalus et al. (2015) (galaxy clustering).

Another approximation, also related to the expansion in Equation 4.14, was derived by Smith et al. (2006) for B-mode polarisation data. By considering the third derivative of the likelihood

around its maximum value these authors found new variables, transformations of C_ℓ or $\hat{C}_\ell$, which had Gaussian likelihoods. This led to a pair of related likelihoods

$$-2\ln P_{\rm ICN}(C_\ell|\hat{C}_\ell) = \frac{9(2\ell+1)}{2\hat{C}_\ell^2}\left(\frac{2\ell+\alpha}{2\ell+1}\right)^{1/3}\left[\left(\frac{2\ell+\alpha}{2\ell+1}\right)^{1/3}\hat{C}_\ell - \hat{C}_\ell^{4/3}C_\ell^{-1/3}\right]^2 , \quad (4.19)$$

where α can take the values 1 or -1 (Hamimeche and Lewis 2008). The variant with $\alpha = 1$ is sometimes referred to as the inverse cubic normal distribution. In this case the distribution of $\hat{C}_\ell^{1/3}$ is Gaussian.

Kalus et al. (2015) found that both Equations 4.18 and 4.19 performed better than Gaussian likelihoods when predicting the amplitude of the primordial non-Gaussianity parameter $f_{\rm NL}$ from galaxy clustering data.

Although not strictly a lognormal approach, Hamimeche and Lewis (2008) developed a likelihood for correlated Gaussian fields which is exact in the full-sky limit. This has since been used in particular for likelihoods involving polarisation measurements, for example Mangilli et al. (2015).

One notable advantage of all the lognormal-based distributions is that they do not require a separate evaluation of the power spectrum covariance matrix, thus reducing computational complexity.

Gaussian-like likelihoods

In this section we clarify various likelihood functions which are each often loosely termed 'Gaussian'. We draw in particular on the discussion in Sun et al. (2013).

The 'standard' Gaussian likelihood for a single ℓ leads to

$$-2\ln P_{\rm Gd}(C_\ell|\hat{C}_\ell) = \frac{(2\ell+1)}{2}\frac{(C_\ell-\hat{C}_\ell)^2}{C_\ell^2} + \ln C_\ell \, . \quad (4.20)$$

This version has a parameter-dependent covariance matrix and also a parameter-dependent determinant term, hence the label Gd. Its maximum is not at $C_\ell = \hat{C}_\ell$. To see this we have

$$-2\frac{\partial \ln P_{\rm Gd}(C_\ell|\hat{C}_\ell)}{\partial C_\ell} = \frac{(2\ell+1)}{2}\left[\frac{2(C_\ell-\hat{C}_\ell)}{C_\ell^2} - \frac{2(C_\ell-\hat{C}_\ell)^2}{C_\ell^3}\right] - \frac{1}{C_\ell} \quad (4.21)$$

$$= \frac{(2\ell+1)(\hat{C}_\ell^2 - \hat{C}_\ell C_\ell) - 2C_\ell^2}{C_\ell^3} \, . \quad (4.22)$$

This peaks at $C_\ell = (2\ell - 1)\hat{C}_\ell/(2\ell + 1)$, so the maximum likelihood is biased low compared to the mean.

Two variants of Equation 4.20 are also used in the literature. Firstly the determinant term is sometimes ignored, giving

$$-2 \ln P_{\rm Gnd}(C_\ell|\hat{C}_\ell) = \frac{(2\ell+1)}{2} \frac{(C_\ell - \hat{C}_\ell)^2}{C_\ell^2} \ . \tag{4.23}$$

One justification for this choice for large surveys is that the determinant becomes relatively less important as the survey size increases (Eifler et al. 2009).

More commonly, a constant covariance matrix is assumed, say $\Sigma = 2(C_\ell^f)^2/(2\ell + 1)$ for some fiducial value C_ℓ^f. In this case the determinant is constant so we have

$$-2 \ln P_{\rm Gcc}(C_\ell|\hat{C}_\ell) = \frac{(2\ell+1)}{2} \frac{(C_\ell - \hat{C}_\ell)^2}{(C_\ell^f)^2} \ . \tag{4.24}$$

Both these versions have maxima at $C_\ell = \hat{C}_\ell$.

Although both $P_{\rm Gnd}$ and $P_{\rm Gcc}$ produce unbiased maximum posterior estimates, none of the Gaussian approximations reproduces the shape of the exact gamma distribution, especially for low values of ℓ. This is illustrated in Figure 4.3 which shows the likelihood $P(\hat{C}_\ell|C_\ell)$ and posterior $P(C_\ell|\hat{C}_\ell)$ for these three common variants of the Gaussian distribution, with the 'true' gamma distribution also shown for comparison. The top panels show the distributions for $\ell = 2$ where the differences in shape are very pronounced. The lower panels are for $\ell = 50$ where it can be seen that all the likelihoods and posteriors tend towards Gaussian. As shown by Equation 4.22, the Gaussian likelihood with parameter-dependent covariance matrix and determinant produces a biased maximum likelihood estimate.

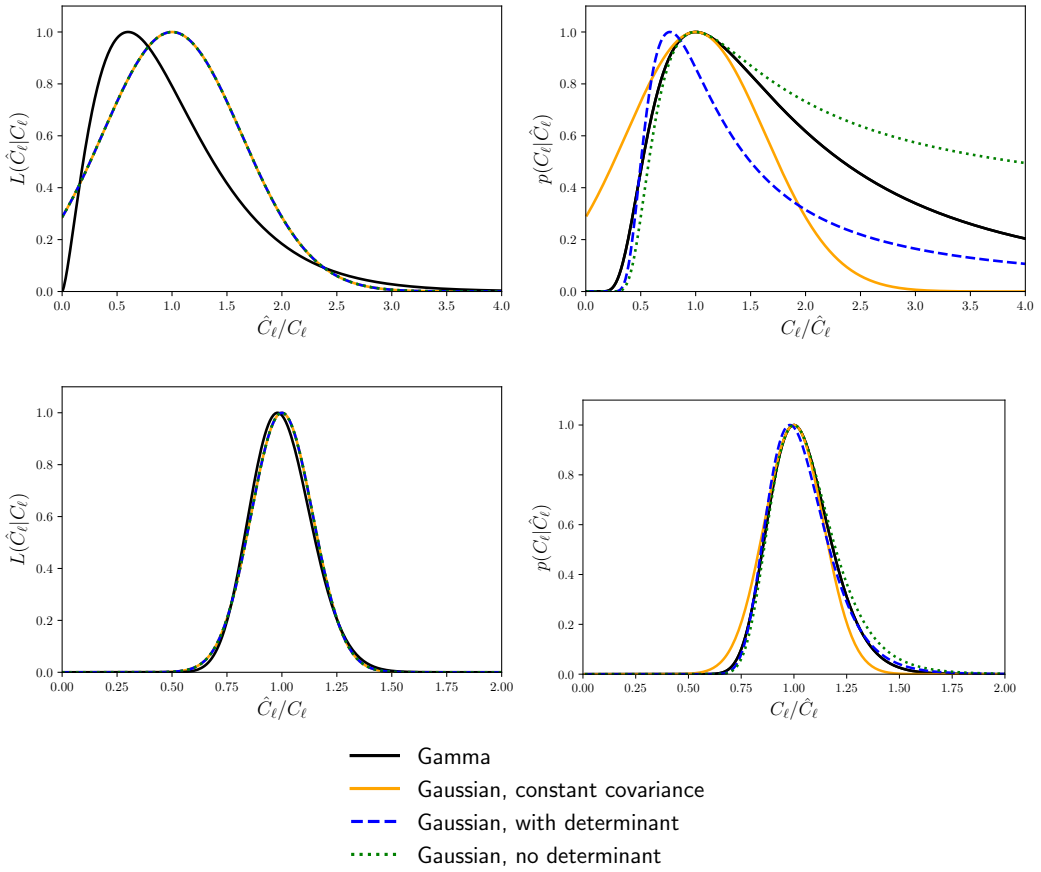

Figure 4.3: *Upper left*: Three different versions of Gaussian likelihoods for $\ell = 2$: with a parameter-dependent covariance matrix and determinant, with a parameter-dependent covariance matrix but no determinant, and with a constant covariance matrix. Also shown is the 'true' gamma distribution. *Lower left*: As for upper left but for $\ell = 50$. *Upper right*: The posterior distributions corresponding to the likelihoods in the left panel for $\ell = 2$. *Lower right*: As for upper right but for $\ell = 50$. All likelihoods and posteriors are normalised to have a maximum-likelihood estimate of 1.0 and mean 1.0.

4.2.5 Parameter constraints with Gaussian and non-Gaussian likelihoods

We have established that in principle the choice of likelihood matters and have considered various alternative likelihoods for the power spectrum and correlation functions. However this does not necessarily mean that in a practical context likelihood choice adversely affects parameter estimates; the form of the likelihood may be less important than the multiplicity of other factors which influence parameter constraints. We now briefly consider what evidence there is that the likelihood choice matters in practice.

Two studies involving weak lensing correlation functions reached conflicting conclusions. Hartlap et al. (2009) used simulations to estimate the non-Gaussian cosmic shear correlation function likelihood and compare its impact to that of a Gaussian likelihood. They found that their non-Gaussian likelihood made the posterior distribution more sharply peaked and skewed and applied their theoretical methods to cosmic shear data from the *Chandra* Deep Field South[2]. Against this, Lin et al. (2019) found no evidence that a Gaussian likelihood would produce biased parameter constraints from a future LSST-like survey, despite detecting non-Gaussianity in their simulated cosmic shear two-point correlation functions. They cautioned, however, that this result might not hold for different redshift distributions, survey parameters or cosmologies.

Outside weak lensing, Hahn et al. (2019) re-analysed two large-scale structure studies which had used Gaussian likelihoods and found that non-Gaussian likelihoods improved parameter constraints. They were equivocal about whether the non-Gaussianity of the likelihood will matter for next-generation surveys, arguing that the central limit theorem would make the likelihood more Gaussian, but also suggesting that the inclusion of extra small- and large-scale modes would enhance non-Gaussianity. Consequently they recommended that the Gaussian assumption should always be tested before use in future analyses.

4.3 Likelihoods for weak lensing correlation functions

The discussion above has focused on the power spectrum, C_ℓ, as the two-point statistic of choice. However for weak lensing this is an idealised approach since it is not possible to apply the theoretical C_ℓ to real data. In a practical situation it is often easier to work in real space and use two-point correlation functions or functions derived from them (for example Hildebrandt et al. (2016); Troxel et al. (2018)). However if the power spectrum is not Gaussian then nor are the correlation functions. It is again not obvious that assuming a Gaussian likelihood is legitimate.

[2]htttp://www.nasa.gov/mission_pages/chandra/main/index.html

As mentioned in Section 4.2.4, Sellentin et al. (2018) (hereafter SHH18) questioned the assumption that the shear two-point correlation functions have Gaussian likelihoods. They developed a hierarchical model of the correlation functions based on the premise that the power spectrum estimator $\hat{C}_\ell$ has a gamma distribution. This work was motivated partly by the detection of significant non-Gaussianity in CFHTLenS shear correlation functions (Sellentin and Heavens 2017) but also by theoretical considerations. We now build on this to explore the joint distribution of the correlation functions in two angular bins and hence to clarify the correct form of likelihood.

To summarise, the SHH18 method involves sampling from the distribution of $\hat{C}_\ell$ and Hankel-transforming to real space to obtain expressions for the likelihoods of the correlation functions. Using the flat sky approximation, their model for the correlation functions is

$$P(\hat{C}_\ell|C_\ell) = \text{Gamma}\left[\frac{\nu(\ell)}{2}, \frac{2C_\ell}{\nu(\ell)}\right], \tag{4.25}$$

$$\hat{\xi}_+(\theta) = \int \frac{\mathrm{d}\ell}{2\pi} \ell J_0(\ell\theta)\hat{C}_\ell \,, \tag{4.26}$$

$$\hat{\xi}_-(\theta) = \int \frac{\mathrm{d}\ell}{2\pi} \ell J_4(\ell\theta)\hat{C}_\ell \,, \tag{4.27}$$

where $\nu(\ell)$ is the degrees of freedom on the masked sky, and $J_0(\ell\theta)$ and $J_4(\ell\theta)$ are Bessel functions of the first kind. For the full sky there are $\nu = (2\ell + 1)/2$ degrees of freedom. In a real survey which does not cover the full sky and has complex geometry fewer ℓ modes are available so ν is reduced.

SHH18 validated this model against results from simulations, and found that the likelihoods of the correlation functions are not Gaussian. They then applied the model to more realistic survey volumes than possible with simulations which are necessarily restricted to small angular scales. They showed that the likelihood becomes more Gaussian as the survey area increases but found that even for *Euclid*-like surveys there is some residual non-Gaussianity at larger scales.

To explore the joint distributions between angular bins we follow a similar procedure. We first calculate the correlation functions from simulations. As far as possible we mirror the set-up in SHH18. We use the same weak lensing simulations, the KiDS-450 mock source catalogue from Scinet LIght Cone Simulations (SLICS) (Harnois-Déraps and van Waerbeke 2015; Harnois-Déraps et al. 2018), although we use only 800 simulations rather than the 930 in SHH18. The angular range of the simulations is 0.5-300 arcmin, similar to that considered in recent KiDS analyses (Hildebrandt et al. 2016, 2020). For consistency with the presentation in SHH18, we do not add

shape noise to the correlation functions. This means we are considering a worst case since shape noise would make the distributions more Gaussian. We use a slightly higher redshift than presented in SHH18 but this does not alter the main conclusions.

We use the TreeCorr package (Jarvis et al. 2004) to calculate the correlation functions from the galaxy positions provide by SLICS. Our results for ξ_+ and ξ_- are shown in Figure 4.4 for similar (small) separation angles to those plotted in Figure 4 of SHH18, and also for larger angles. The skewed distributions for ξ_+ at small angles agree qualitatively with SHH18. We note that although we have only plotted positive values of the correlation functions, they can in fact become negative. For ξ_- we do not show results for angular separations below 1 arcmin. Resolution limits of the SLICS simulations mean that the simulations lack power at low angular scales, which particularly affects ξ_-. This is discussed in detail in Harnois-Déraps and van Waerbeke (2015) and we return to the issue later when we compare our modelling with the simulations. SHH18 adjust for known deficiencies in the SLICS simulations by aligning the simulation results to their model, but we do not make this change.

The relative positions of the ξ_- distributions at small angular separations arise because at small scales ξ_- is an increasing function of θ, unlike ξ_+ which decreases monotonically across the whole range of angular scales.

In Figure 4.5 we illustrate the joint distribution of ξ_+, calculated from the simulations, between the second and third angular bins of Figure 4.4. The figure shows a strong relationship between the two bins but we do not have enough simulations to explore this in any detail. Instead we need to turn to the analytical model. To visually demonstrate the effect of non-Gaussianity we also show the joint distribution of two Gaussian distributions with the same means and standard deviations. This is purely to show the difference between the shapes of the distributions.

We now use Equations 4.25 to 4.27 to generate samples of ξ_+ and ξ_-. To calculate the correlation functions we use the cosmological parameter estimation code Cosmosis[3] (Zuntz et al. 2015). We calculate the non-linear matter power spectrum using the Halofit model (Smith et al. 2003). As far as possible we align our cosmological parameter values and survey characteristics such as redshift, area and galaxy number density with the simulation set up. Thus we model a 100 deg^2 survey. We consider a single tomographic bin with mean redshift $z = 0.33$ convolved with a Gaussian filter of width $0.02(1 + z)$. We use 32 logarithmically equally spaced angular bins.

[3]https://bitbucket.org/joezuntz/cosmosis/wiki/Home

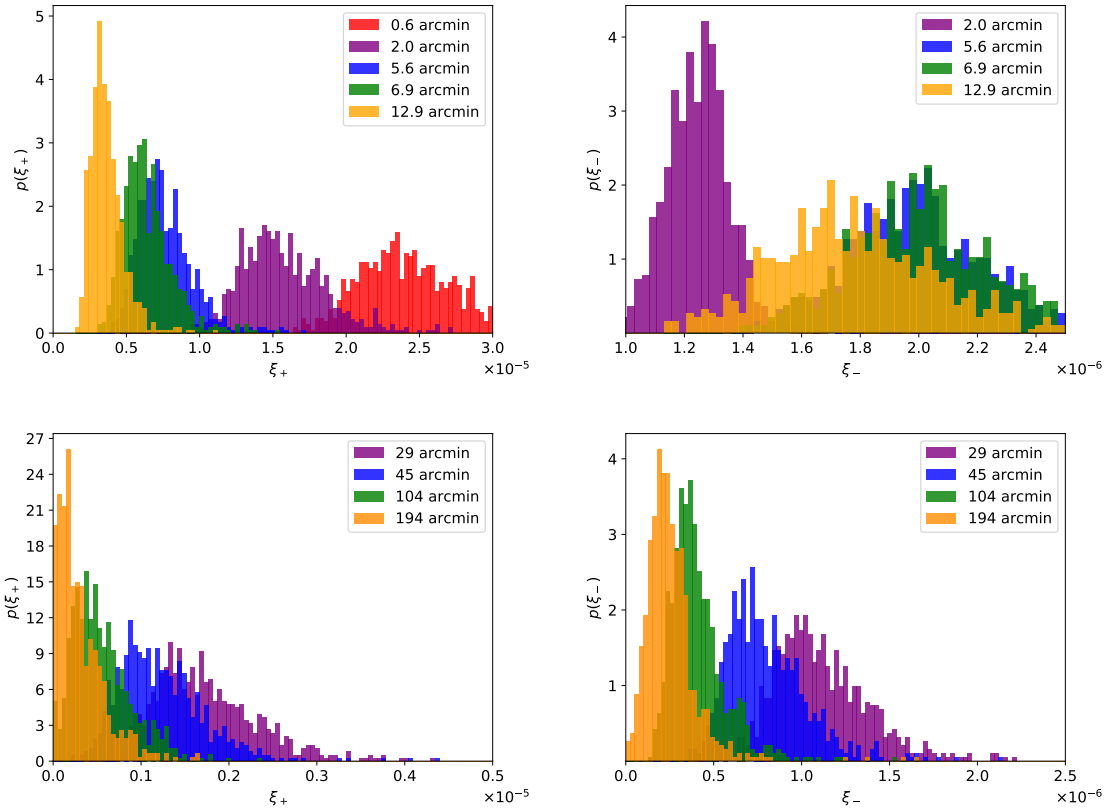

Figure 4.4: Results from 800 SLICS simulations of 100 deg^2: marginal distributions of ξ_+ and ξ_- at redshift ~ 0.33. The area of each histogram is normalised to 1.
Left: ξ_+, *Right*: ξ_- , *Top*: Small angular separations, *Bottom*: Large angular separations.

We determine the degrees of freedom ν using a similar method to SHH18. They derive the formula

$$\nu \approx \frac{f_{\rm sky}(2\ell+1)g_{\rm eff}(\ell)}{\ell_{\rm pix}}\,, \tag{4.28}$$

where $f_{\rm sky} = A_{\rm sky}[\rm sterad]/4\pi$ is the sky fraction. The factor $g_{\rm eff}$ quantifies the effect of the survey mask. For a real survey with complicated geometry $g_{\rm eff}$ will not generally be easy to estimate. SHH18 approximated it from the SLICS simulations by minimising the distance between histograms for ξ_+ derived from the simulations and from the model. We adopt their estimate which is 2.29, independent of ℓ. The quantity $\ell_{\rm pix}$ expresses the loss of degrees of freedom caused by the discrete nature of the shear field. It represents the size of a 'pixel' which contains sufficient galaxies to be considered as a smooth field. SHH18 derived a value for $\ell_{\rm pix}$ by considering a hypothetical pixel containing $N_{\rm gal}$ galaxies and asking how large $N_{\rm gal}$ must be to allow the galaxy distribution

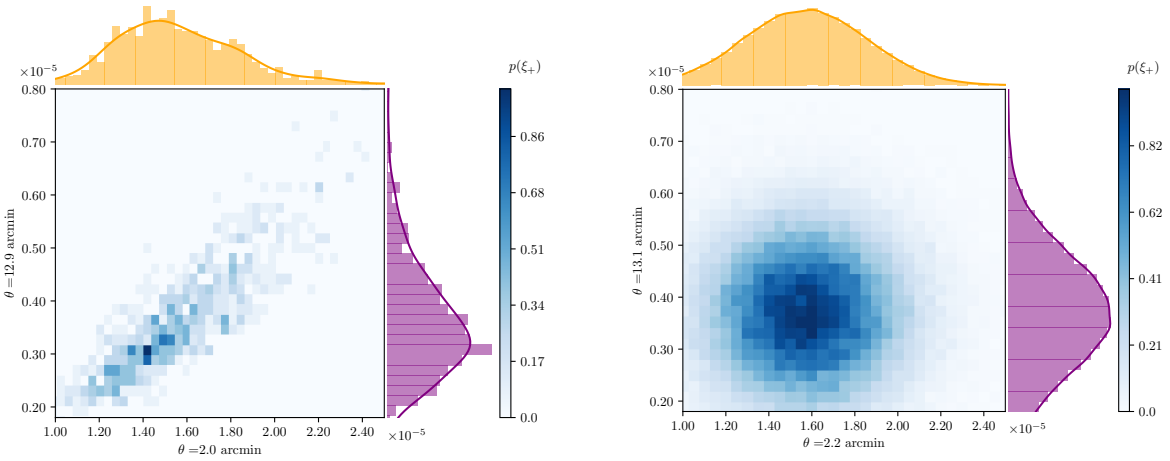

Figure 4.5: *Left*: Joint distribution of ξ_+ between two angular bins at approximately 2 arcmin and 13 arcmin, based on 800 SLICS simulations. *Right*: Joint distribution of two Gaussian distributions with the same means and variance.

within the pixel to be treated as continuous. This is related to the number density of galaxies, $\bar{n}$, by

$$\ell_{\rm pix} \approx \sqrt{\frac{\pi^2 \bar{n}}{N_{\rm gal}}} \,. \tag{4.29}$$

We take $\bar{n}$ to be 2.6 arcmin^{-2} per tomographic bin, as in SHH18, and estimate $N_{\rm gal}$ by fitting model results for ξ_+ to results from the simulations, in a similar way to the estimation of $g_{\rm eff}$. There is very little room for manoeuvre in this fitting process as we are aiming to fit several histograms for both ξ_+ and ξ_- by changing a single variable, $N_{\rm gal}$. Since $g_{\rm eff}$ is already estimated, the estimation of degrees of freedom is the weakest part of the model. We found it difficult to obtain a value of $N_{\rm gal}$ which worked well for both ξ_+ and ξ_-, and so settled for one which provided a good fit for the distribution of ξ_+. The poorer fit for ξ_- is evident in the results presented below. There is no obvious reason why the same estimate of ν should not work for both ξ_+ and ξ_-; a more sophisticated model of the degrees of freedom may be needed.

As basic validation of our modelling, Figure 4.6 compares the mean values of the correlation functions from our model (red) and the simulations (blue) to the analytical results from Cosmosis (orange). The model and Cosmosis results agree very well and are barely distinguishable in this figure. However the simulations show large deviations at small and large scales. As discussed before, the small-scale deviations are due to the limited resolution of the simulations. To partly compensate for this we set $k_{\rm max} = 40h^{-1}{\rm Mpc}$ in the halofit model. This artificially reduces

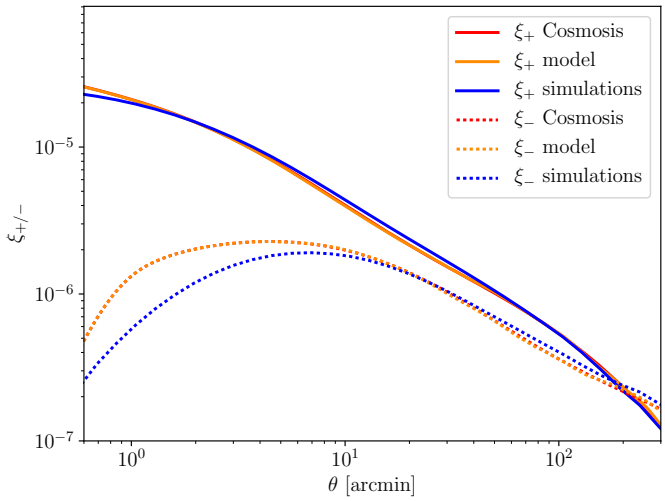

Figure 4.6: Mean two-point correlation functions ξ_+ and ξ_- estimated from SLICS simulations (blue) compared with the analytical values from Cosmosis using the halofit model (red) and the mean values from the hierarchical model given by Equations 4.25 and 4.27 (orange). The analytical and modelled results agree well and are visually almost indistinguishable.

the theoretical and modelled correlation functions at small angles, especially ξ_- which probes small scales more than ξ_+ does. Similar high-k cut-offs were explored in Harnois-Déraps and van Waerbeke (2015). This reduces the small-scale discrepancy but at the expense of distorting the model slightly. At large scales the simulations lack modes larger than the simulation box size – the so-called finite support effect, which particularly affects ξ_+. This too is discussed by Harnois-Déraps and van Waerbeke (2015) who propose a simple rescaling to compensate for the effect. We have not made any adjustment for the finite box size and simply note that at both the smallest and largest scales we cannot expect our modelled correlation functions to exactly match the simulations.

Figure 4.7 shows our modelled results for ξ_+ and ξ_-. To more explicitly show how the modelled distributions match the simulations, in Figure 4.8 we compare model and simulation results for some illustrative angular bins. In keeping with Figure 4.6, the mean values from the model agree better with the simulations for ξ_+ than ξ_-. Moreover because we estimated $N_{\rm gal}$ to agree with ξ_+ and not ξ_- the distributions for ξ_+ match the simulations better than the ξ_- distributions do. Nevertheless there is broad qualitative agreement with the simulation results. In particular the skewness of the distributions is apparent and it is clear that the skewness increases with separation angle.

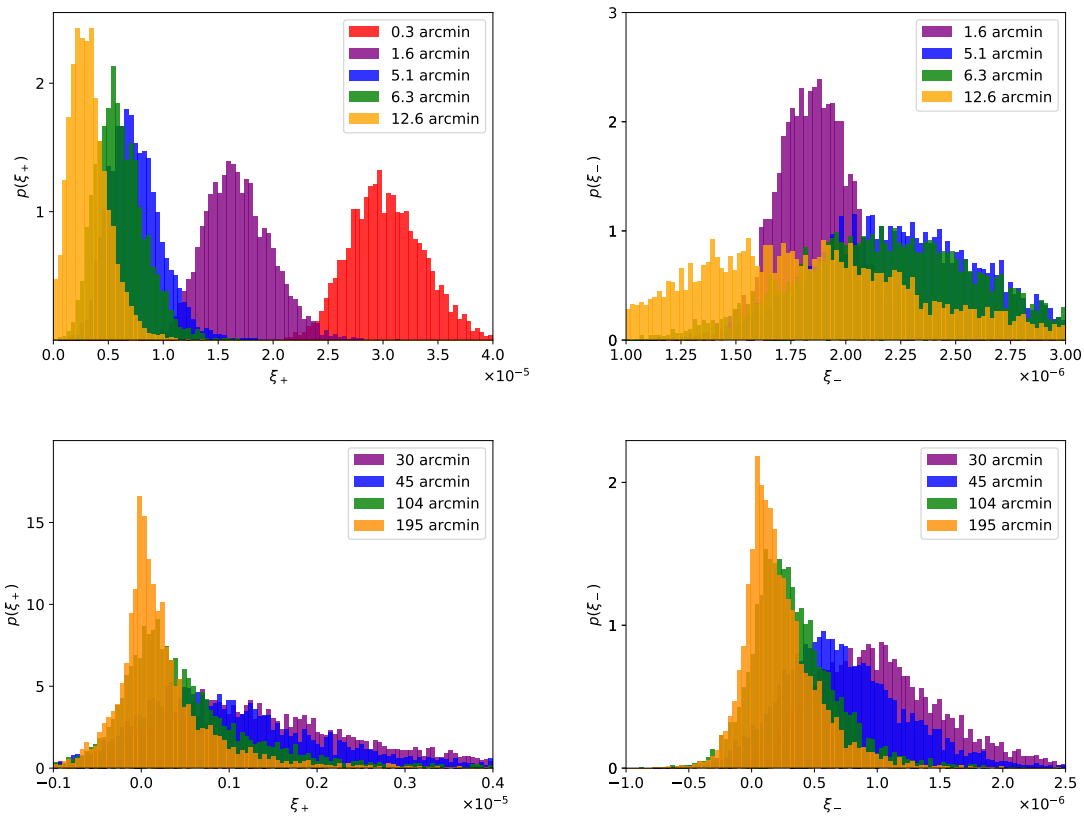

Figure 4.7: Marginal distributions of ξ_+ and ξ_-. The distributions are derived from the hierarchical model given by Equations 4.25 and 4.27 assuming a survey area of 100 deg^2. The area of each histogram is normalised to 1. *Left*: ξ_+, *Right*: ξ_- , *Top*: Small angular separations, *Bottom*: Large angular separations.

One advantage of using the model rather than the simulations is that we can consider a larger survey area. Figure 4.9 is similar to Figure 4.7 but for a KiDS-like survey of 450 deg^2. The main point to take from this is the decreased skewness of the distributions for a larger survey.

We now explore the joint distributions of pairs of angular bins. Figures 4.10 and 4.11 show joint distributions of ξ_+ between two angular bins for surveys with area 100 deg^2 and 450 deg^2 respectively. Figure 4.12 shows results for ξ_- for a 450 deg^2 survey. In each case we show results for a pair of relatively small angles and for two larger angles. Alongside each plot we also show comparable results for Gaussian distributions with the same means and standard deviations.

These results confirm that the correlation function likelihoods are not Gaussian. This is shown clearly by the SLICS simulations but also applies to larger KiDS-like survey areas. The skewness of the distributions shows clearly in the bivariate plots which confirm that the non-Gaussianity decreases for larger survey areas but increases with angular separation. The SHH18 model can fit the simulations well, although we found it difficult to fit ξ_+ and ξ_- simultaneously, and provides a

simple recipe for a non-Gaussian likelihood which is not much more complicated than the normal Gaussian likelihood. Once again we conclude that in this idealised situation a Gaussian likelihood may not be valid, except for small scales and large survey areas.

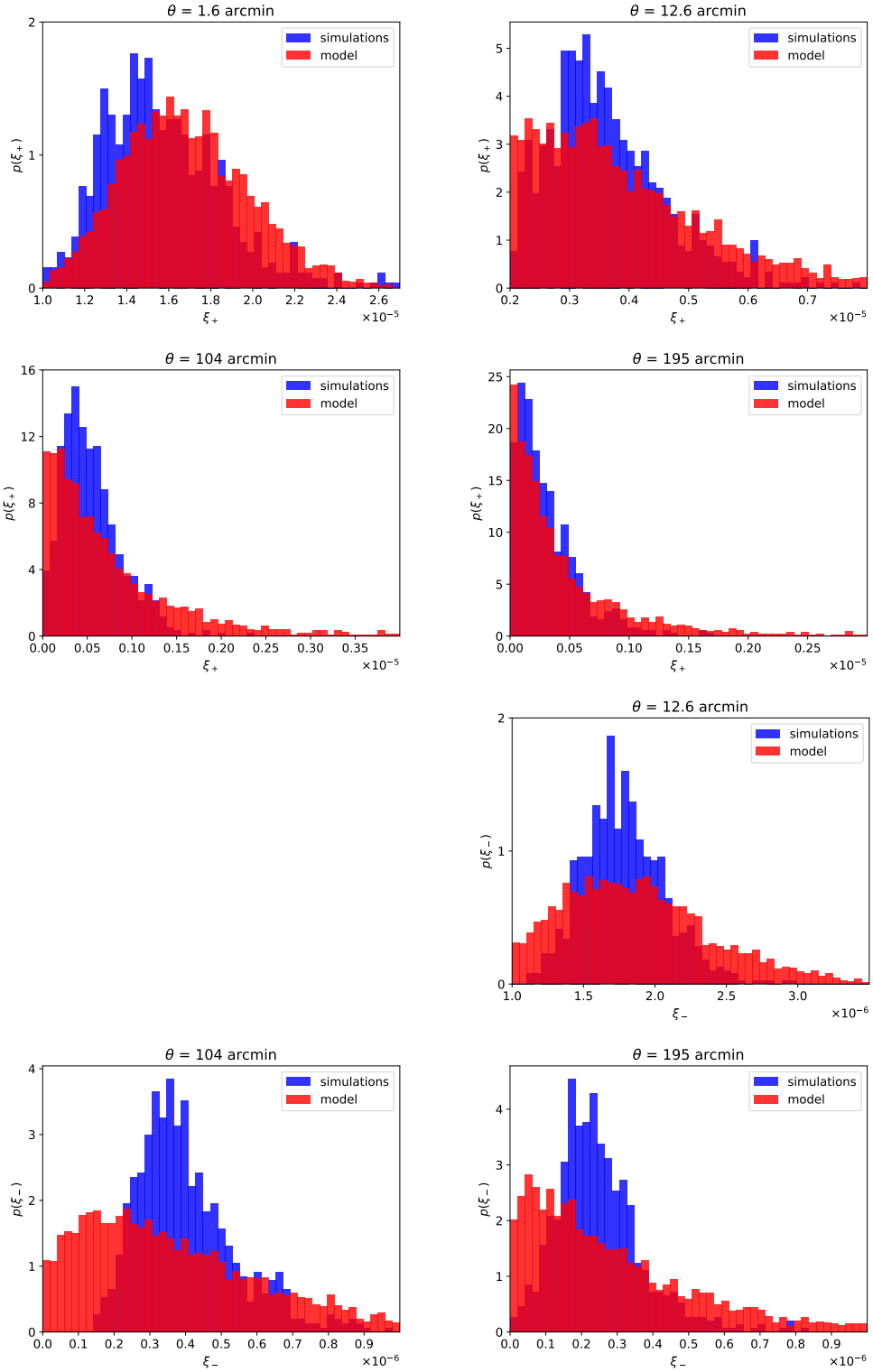

Figure 4.8: Comparison between distributions of ξ_+ and ξ_- from simulations and the hierarchical model given by Equations 4.25 and 4.27 for illustrative angular bins, assuming a survey area of 100 deg^2 and $z \sim 0.33$. The area of each histogram is normalised to 1.

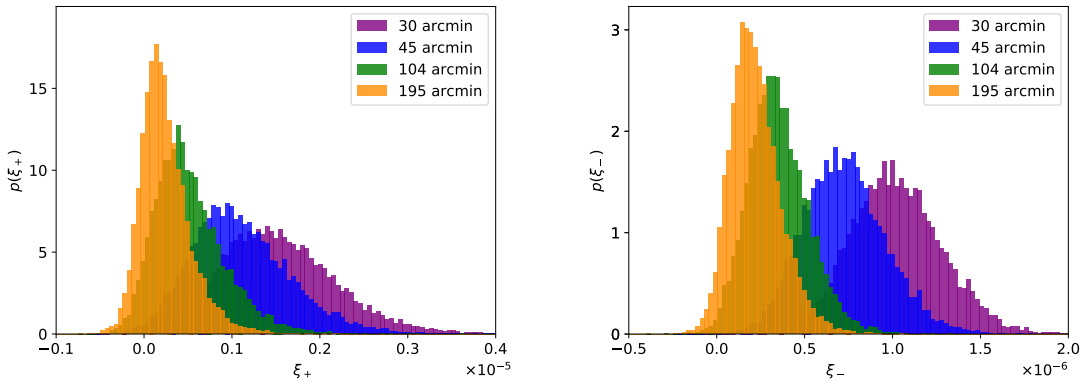

Figure 4.9: Marginal distributions of ξ_+ and ξ_-, for four illustrative angular bins for a KiDS-like survey. The distributions are derived from the hierarchical model given by Equations 4.25 and 4.27 assuming a survey area of 450 deg^2. The area of each histogram is normalised to 1.

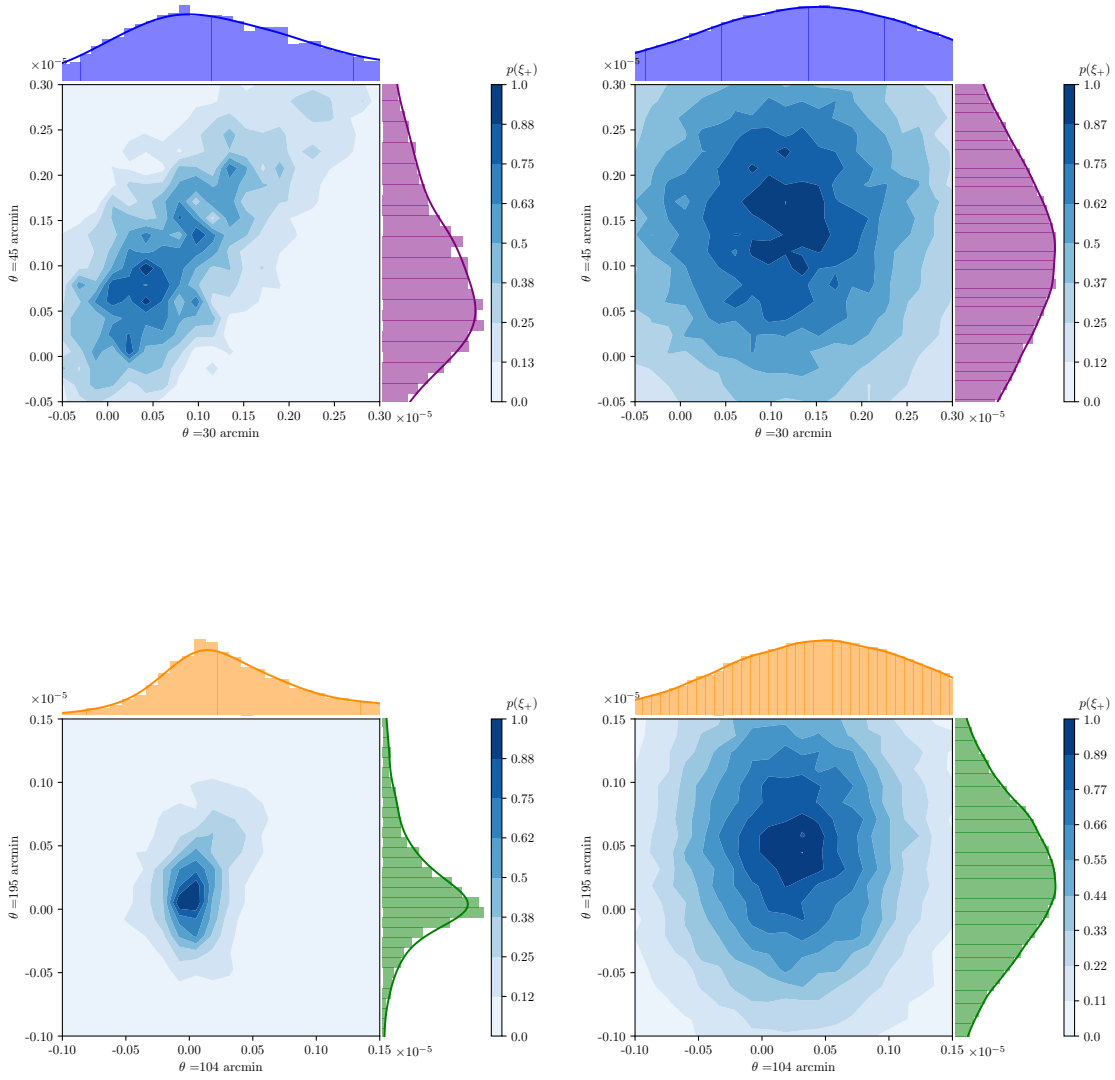

Figure 4.10: *Left*: Joint distributions of ξ_+ in pairs of angular bins for a survey area of 100 deg^2, based on the hierarchical model given by Equations 4.25 to 4.27. *Right*: Joint distributions of two Gaussian distributions with the same means and standard distributions, for comparison.

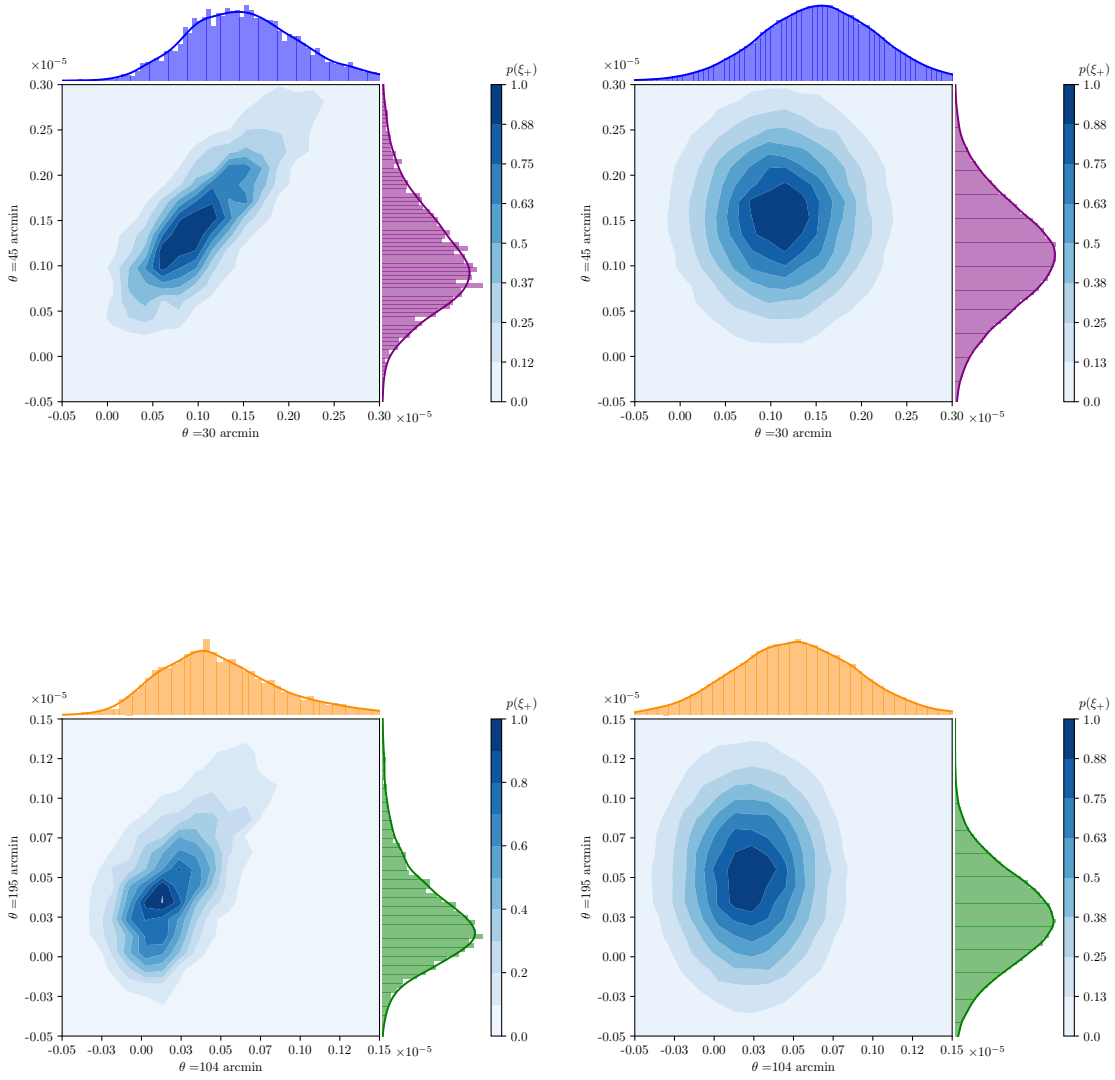

Figure 4.11: *Left*: Joint distributions of ξ_+ in pairs of angular bins for a survey area of 450 deg^2, based on the hierarchical model given by Equations 4.25 to 4.27. *Right*: Joint distributions of two Gaussian distributions with the same means and standard distributions, for comparison.

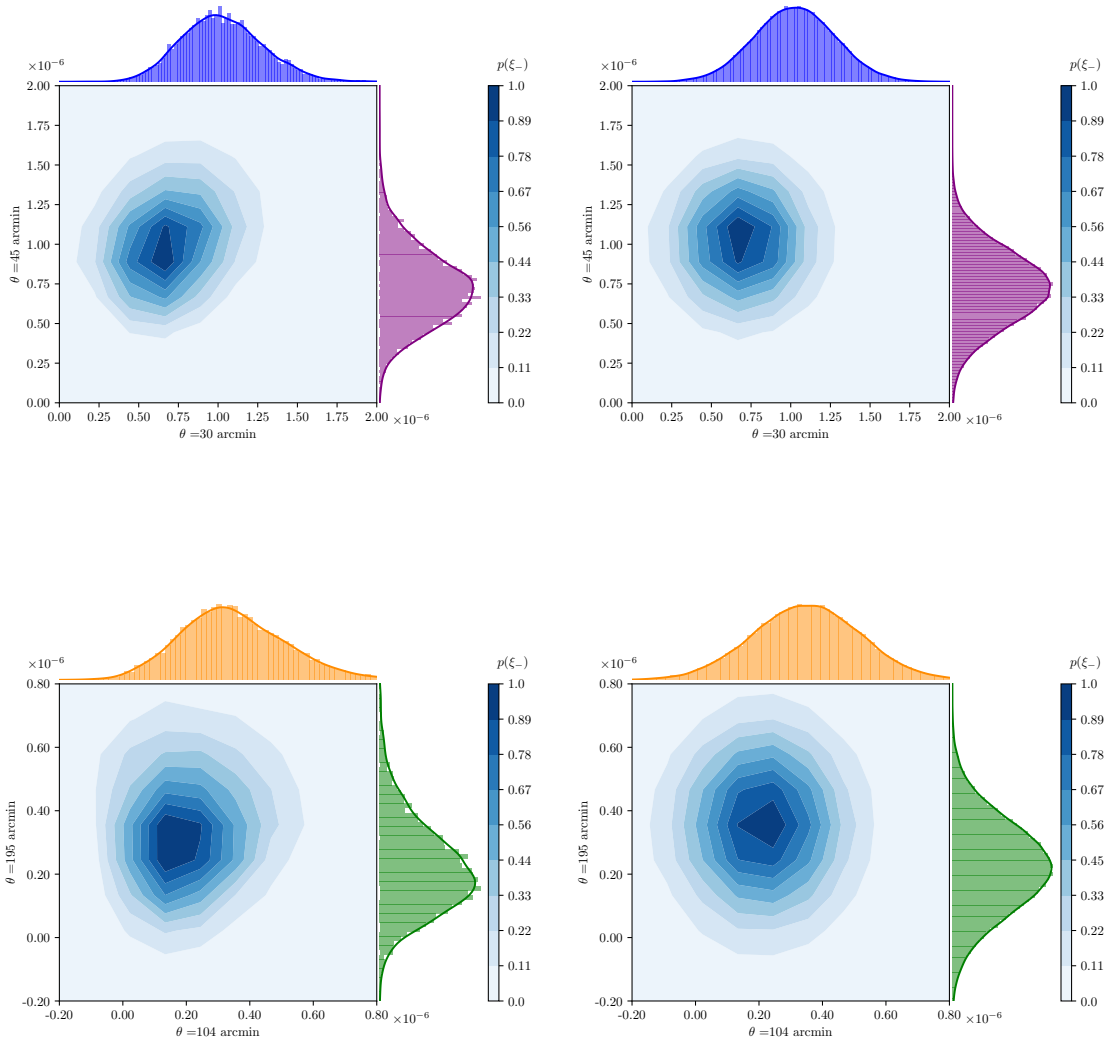

Figure 4.12: *Left*: Joint distributions of ξ_- in pairs of angular bins for a survey area of 450 deg^2, based on the hierarchical model given by Equations 4.25 to 4.27. *Right*: Joint distributions of two Gaussian distributions with the same means and standard distributions, for comparison.

4.4 Parameter-dependence of the covariance matrix

This section picks up from the discussion of Gaussian power spectrum likelihoods in Section 4.2.4 and discusses whether the covariance matrix should be treated as parameter-dependent.

4.4.1 A Fisher matrix viewpoint

Section 4.2.4 identified pitfalls from using a Gaussian likelihood with a varying covariance matrix, because the maximum likelihood is biased low compared to the mean. To underline this, the same problems can also be seen from an information-content perspective. In this section we closely follow the arguments in Carron (2013) based on Fisher matrix considerations.

The Fisher matrix quantifies the amount of information contained within a field. Consider a Gaussian distribution P with mean μ and variance Σ. For simplicity assume this depends on only two parameters α and β. The Fisher matrix is then defined by

$$F_{\alpha\beta} = \left\langle \frac{\partial P}{\partial \alpha} \frac{\partial P}{\partial \beta} \right\rangle , \tag{4.30}$$

evaluated at the maximum-likelihood point. This can be written as (Tegmark et al. 1997)

$$F_{\alpha\beta} = \sum_{i,j} \frac{\partial \mu_i}{\partial \alpha} \Sigma^{-1}_{ij} \frac{\partial \mu_j}{\partial \beta} + \frac{1}{2} \mathrm{Tr} \left[\Sigma^{-1} \frac{\partial \Sigma}{\partial \alpha} \Sigma^{-1} \frac{\partial \Sigma}{\partial \beta} \right] . \tag{4.31}$$

If we work in the field perspective as defined in Section 4.2.2 then the Fisher matrix is

$$F_{\alpha\beta} = \frac{1}{2} \sum_{l=0}^{\infty} (2\ell + 1) \frac{1}{C_\ell} \frac{\partial C_\ell}{\partial \alpha} \frac{1}{C_\ell} \frac{\partial C_\ell}{\partial \beta} , \tag{4.32}$$

since the field has zero mean and covariance C_ℓ.

If instead we adopt the estimator approach then the mean of the estimator is C_ℓ and its covariance is $2C_\ell^2/(2\ell + 1)$. So in this case the Fisher matrix is

$$F_{\alpha\beta} = \frac{1}{2} \sum (2\ell + 1) \frac{1}{C_\ell} \frac{\partial C_\ell}{\partial \alpha} \frac{1}{C_\ell} \frac{\partial C_\ell}{\partial \beta} + \frac{1}{2} \sum 4 \frac{1}{C_\ell} \frac{\partial C_\ell}{\partial \alpha} \frac{1}{C_\ell} \frac{\partial C_\ell}{\partial \beta} . \tag{4.33}$$

As remarked by Carron (2013), this result raises questions because it seems to imply that the estimator contains more information than the field itself. This contradicts the Cramér–Rao inequality which sets a theoretical limit on the information content: transforming the data by constructing a power spectrum estimator can at best conserve information and certainly cannot increase it.

Carron (2013) proceeds to show that this paradox is a consequence of using a Gaussian power spectrum likelihood. If instead we use the exact gamma likelihood we have from Equation 4.30

$$F_{\alpha\beta} = \frac{\partial\theta}{\partial\alpha}\frac{\partial\theta}{\partial\beta}\left\langle\left(\frac{\partial \ln P(\hat{C}_\ell|C_\ell)}{\partial\theta}\right)^2\right\rangle\Bigg|_{\rm ML} \tag{4.34}$$

$$= \left(\frac{2}{2\ell+1}\right)^2\frac{\partial C_\ell}{\partial\alpha}\frac{\partial C_\ell}{\partial\beta}\left\langle\left(\frac{\partial \ln P(\hat{C}_\ell|C_\ell)}{\partial\theta}\right)^2\right\rangle\Bigg|_{\rm ML}. \tag{4.35}$$

From Equation 4.9

$$-2\frac{\partial \ln P(\hat{C}_\ell|C_\ell)}{\partial\hat{C}_\ell} = \frac{2\ell+1}{C_\ell} - \frac{2\ell-1}{\hat{C}_\ell}\,, \tag{4.36}$$

so the maximum likelihood is at $\hat{C}_\ell = (2\ell-1)C_\ell/(2\ell+1)$. Substituting this into Equation 4.35 we get

$$F_{\alpha\beta} = \left(\frac{2}{2\ell+1}\right)^2\frac{\partial C_\ell}{\partial\alpha}\frac{\partial C_\ell}{\partial\beta}\frac{1}{C_\ell^2}\left(\frac{2}{(2\ell+1)}\right)^2, \tag{4.37}$$

and summing over all ℓ retrieves Equation 4.32. So in this idealised example the exact 'estimator' likelihood contains the same information as the field.

The extra anomalous term in Equation 4.33 arises because the covariance matrix has been allowed to vary with the parameters. If instead the covariance is taken as fixed then this term disappears and Equation 4.33 becomes the same as Equation 4.32. This is precisely what is done in most real weak lensing likelihood analyses, but for the 'wrong' reasons. However with the availability of more computing power some recent analysis has instead used a parameter-dependent covariance matrix, for example Hikage et al. (2019). This strategy is considered further in Section 4.4.2.

The main conclusion from this discussion is that using a Gaussian likelihood and a varying covariance matrix introduces spurious information and is formally incorrect. It will lead to biased parameter estimates and underestimates of parameter uncertainties. However, fortuitously, a Gaussian likelihood and fixed covariance matrix produce unbiased maximum-likelihood estimates (though uncertainties are still mis-estimated as discussed in Section 4.2.4). For real surveys with complications such as systematics, sky cuts and tomography this validates the use of a Gaussian likelihood and parameter-independent covariance matrix. In practice this is what is commonly done, but mainly for computational ease rather than as a matter of principle.

4.4.2 Effect of using a parameter-dependent covariance matrix

Notwithstanding the discussion above, there is no real consensus about whether a parameter-dependent covariance matrix should be used. The conclusions in Eifler et al. (2009) have been particularly influential. This work, based on cosmic shear, demonstrated convincingly that parameter estimates depend on the parameter values at which the covariance matrix is calculated. The authors recommended that the covariance should ideally be recalculated at every point in parameter space, and devised a fitting function for this purpose.

Following this several authors investigated the effect of using (or not using) a parameter-dependent covariance matrix with a Gaussian likelihood, mostly to justify a fixed covariance. However the results of these investigations have been inconsistent and inconclusive. Some studies have suggested that a varying covariance matrix makes little difference. In a weak lensing analysis using data from CFHTlens, Kilbinger et al. (2013) used the fitting function from Eifler et al. (2009) to continuously update their covariance matrix and concluded that the impact was small. Hikage et al. (2019) used the halo model to re-estimate the power spectrum, and hence the covariance matrix, at each step of their nested-sampling likelihood analysis. They too concluded that the cosmology-dependence of the covariance matrix had only a small effect: estimates of S_8 and $\Omega_{\rm m}$ agreed with fiducial results to within 20% of the statistical uncertainty, with errors also agreeing within 10%.

Beyond weak lensing, two analyses of SDSS[4] BAO data, Labatie et al. (2012) using BOSS DR7 and Loureiro et al. (2019) using BOSS DR12, investigated the use of parameter-dependent covariance matrices and found little difference in parameter constraints. More recently Kodwani et al. (2019) used Fisher matrix analysis to forecast the effects of using a fixed rather than varying covariance matrix. They concluded that for future large surveys such as *Euclid* the parameter-dependence of the covariance matrix will be immaterial, so long as the fixed covariance is calculated at the true bestfit cosmology (although in practice this may not be known).

In contrast others have found significant effects. Jee et al. (2013) in their analysis of data from the Deep Lens Survey[5] reported that using a (partly) parameter dependent covariance matrix significantly shifted the parameter contours in the $\Omega_{\rm m} - \sigma_8$ plane. Kalus et al. (2015) using BOSS DR9 galaxy clustering data concluded that a varying rather than fixed covariance matrix improved the estimation of some parameters, though not of others.

[4] https://www.sdss.org/surveys/boss/
[5] https://www.noao.edu/survey-archives/deeplens/index.html

Thus a common view seems to be that the covariance matrix should ideally be treated as parameter-dependent but that this does not make much difference in practice. For example, both Joudaki et al. (2016) and Hildebrandt et al. (2016) cite Eifler et al. (2009) and Kilbinger et al. (2013) to justify using a fixed covariance matrix. As we have seen in Section 4.4.1 this view is not correct. Differences in parameter constraints obtained with varying, rather than fixed, covariance matrices may be spurious. It is fortunate that these differences seem to be small so that when the covariance is incorrectly allowed to vary this does not have much impact. The most accurate procedure if using a Gaussian likelihood is to use a fixed covariance matrix.

4.4.3 Choice of parameter values at which to calculate the covariance

Having decided to use a fixed, parameter-independent covariance matrix, a choice must be made about the fiducial cosmology at which to calculate it. This choice is rarely discussed explicitly even though it makes a difference to parameter constraints, as demonstrated by Eifler et al. (2009). Indeed Harnois-Deraps et al. (2019) concluded from Fisher matrix analysis that choosing the best cosmology at which to calculate the covariance matrix might be an order of magnitude more important than optimising the covariance estimation method. However they used some quite extreme cosmologies in their forecasts, which may have affected their conclusions. In practice recent weak lensing analyses have used current accurately-known sets of parameter values, for example most KiDS-450 analysis has used WMAP9 values (Hildebrandt et al. 2016; Köhlinger et al. 2017).

Eifler et al. (2009) and Harnois-Deraps et al. (2019) both suggest that the best way to choose where to evaluate the covariance matrix is to use an iterative approach: guess a cosmology, use this to calculate the covariance matrix and estimate parameters, recalculate the covariance at the new bestfit values, and continue until convergence is reached, that is until further iteration produces no further change in the bestfit values. The disadvantage of this is that the covariance and bestfit parameters must be estimated several times. This is computationally demanding whether the covariance is calculated analytically or with simulations. Eifler et al. (2009), working in real space, demonstrated that an iterative process starting from an arbitrary cosmology could converge within four iterations, though none of their five trials converged to their fiducial parameter values. In a similar spirit van Uitert et al. (2018) started from both Planck15 and WMAP9 parameter values and iteratively re-estimated the covariance matrix and bestfit parameter values, reaching convergence after a single iteration in both cases. This approach is becoming more common: for example both

Krause et al. (2017) and Hildebrandt et al. (2020) used a similar process, starting from existing bestfit values and iterating once (without explicitly testing for convergence).

Thus an iterative method seems to offer a systematic way of choosing where to evaluate the covariance matrix. However it is not obvious how quickly such a process will converge in general, or indeed whether it will necessarily converge in a small number of steps if the starting cosmology is not ideal. Nor is it clear whether the converged values depend on the chosen starting point. Finally, and most importantly, it must be demonstrated that calculating the covariance at a 'converged' point does not bias the final parameter estimates obtained from a maximum-likelihood analysis.

4.4.4 Iterative process

In the light of the previous discussion, in this section we develop a more rigorous iterative method for choosing where to evaluate the covariance matrix. Our iterative scheme has two main components: an emulator which can recalculate the covariance matrix quickly at different cosmologies and a simple routine which finds the maximum of a short MCMC chain generated by the Monte Python sampler (Brinckmann and Lesgourgues 2019; Audren et al. 2013). In combination these components can iterate many times to find converged parameter values which are then used to calculate a fixed covariance matrix. This is fed into a nested sampling analysis which explores the full posterior in the normal way. For this we use Monte Python together with the nested sampling code MultiNest (Feroz and Hobson 2008; Feroz et al. 2013).

Details of the emulator are given in Appendix 4A. Here we only note a few key points. First, we take the 'true' covariance which the emulator must reproduce to be that generated by the analytical code used in most KiDS analyses (Hildebrandt et al. 2016; Köhlinger et al. 2017; van Uitert et al. 2018; Hildebrandt et al. 2020). This approach differs from other covariance matrix emulators which obtain the truth from simulations (Morrison and Schneider 2013; White and Padmanabhan 2015). Secondly, we emulate the elements of the covariance matrix directly, rather than calculating the matrix from an emulated power spectrum. Thirdly, we are interested only in $\Omega_{\rm m}$ and σ_8 but the emulator uses $\Omega_{\rm c}h^2$ and $\ln 10^{10}A_{\rm s}$ because $\Omega_{\rm m}$ and σ_8 are derived parameters in the KiDS-450 analyses. We translate between the two sets of parameters using the CLASS Boltzmann code (Blas et al. 2011). Finally, we emulate only the Gaussian and supersample terms of the covariance, ignoring the other non-Gaussian terms which are sub-dominant (Barreira et al. 2018*a*).

Our analysis is based on a KiDS-like survey, using survey details from Hildebrandt et al. (2016). Following van Uitert et al. (2018) we use four redshift bins and five logarithmically-spaced angular bins with means ranging from $\ell = 200$ to $\ell = 1500$ (see Figure A1 in van Uitert et al. (2018)).

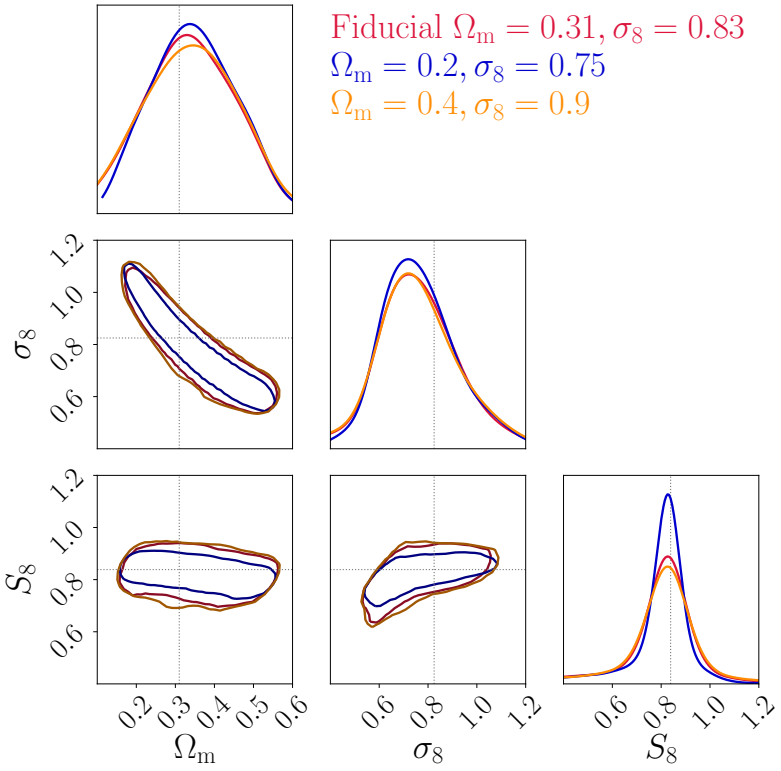

Figure 4.13: 68% credible regions and marginal distributions when the covariance matrix (including Gaussian and supersample terms) is calculated at the 'true' cosmology $\Omega_m = 0.31$, $\sigma_8 = 0.83$ (red), and at two arbitrary cosmologies $\Omega_m = 0.2$, $\sigma_8 = 0.75$, (blue) and $\Omega_m = 0.4$, $\sigma_8 = 0.9$ (orange). Black dotted lines indicate the fiducial parameter values. Calculating the covariance matrix at an inappropriate point can produce misleading values for the uncertainties on S_8, as shown by the blue contours.

We construct a mock data vector at a fiducial cosmology based on values from Ade et al. (2016). We use the fitting function of Takahashi et al. (2012) to calculate the full non-linear matter power spectrum.

Purely for illustration, Figure 4.13 confirms that biased estimates of $S_8 = \sigma_8\sqrt{\Omega_m/0.3}$ can indeed arise if an inappropriate cosmology is used to calculate the covariance matrix. This figure compares the 68% credible regions and marginal distributions obtained when the covariance matrix is evaluated first at our fiducial cosmology and then at two arbitrary cosmologies on either side of the Ω_m–σ_8 'banana'. Other parameters are kept fixed in this analysis. The blue contours in particular confirm that the choice of cosmology at which the covariance matrix is calculated does matter. Our iterative method is designed to avoid this sort of inadvertent bias by systematically choosing a cosmology which will produced unbiased estimates of S_8.

Iteration results

To test the iteration process we use the emulator to calculate the Gaussian and supersample terms of the covariance matrix. To mimic lack of knowledge of the true covariance we add to the data vector noise sampled from the Gaussian term of the 'true' analytical covariance matrix. We compare the results to a benchmark which uses only the Gaussian part of the covariance matrix and assumes that the data vector is known with certainty. In each case we generate 10 random starting cosmologies with $\Omega_{\rm c}h^2$ in the range $[0.11, 0.13]$ and $\ln 10^{10}A_{\rm s}$ in the range $[2.9, 3.1]$. These ranges include both the *Planck* and KiDS bestfit values. The starting points are generated separately for each case.

Results are shown in Figures 4.14 (Gaussian covariance with no noise in the data vector) and 4.15 (Gaussian plus supersample covariance with a noisy data vector). The top panels show the starting positions, marked by red dots, and the number of iterations required to reach convergence from each point. The black dashed line marks the line of constant fiducial S_8, which is a proxy for the direction of the $\Omega_{\rm m}$–σ_8 degeneracy line, and the blue cross indicates the fiducial cosmology at which the data vector was calculated. The bottom panels show convergence to the final value of S_8 as the number of iterations increases. For completeness in Appendix 4B we also show results for two other cases: Gaussian covariance with a noisy data vector, and Gaussian plus supersample covariance with no noise in the data vector.

For the benchmark case of a Gaussian covariance matrix with no noise in the data vector the iterative process reaches the constant-S_8 line within six iterations, apart from one starting point far from the degeneracy line which required eight iterations. As might be expected, the more complicated covariance in Figure 4.15 combined with a noisy data vector makes convergence slower. Nevertheless in all but one case, where the starting point was again far from the degeneracy line, the process converges within 10 iterations. The number of iterations required depends most strongly on the starting position and is relatively less affected by the complexity of the covariance matrix or uncertainty in the data vector.

These results give confidence that the iterative process works and generally converges within no more than about six iterations, provided that the starting point is reasonably close to the constant-S_8 line. The converged positions are all close to the lines of constant S_8, though none are exactly at the fiducial values of $\Omega_{\rm m}$ and σ_8, probably because the likelihood is very flat along the degeneracy line.

Having shown that the iterative process works, we also check that calculating the covariance matrix at a 'converged' position rather than the fiducial cosmology does not introduce bias into the parameter estimates. To test this we run a full nested-sampling analysis firstly with a covariance

matrix calculated at the fiducial cosmology and then at two of the 'worst' converged points. These were chosen slightly subjectively as those which were furthest from the constant-S_8 line or furthest along that line from the fiducial point. The results are shown in Figure 4.16, for the same combinations of covariance (Gaussian term only or Gaussian plus supersample) and data vector (without or with noise drawn from the Gaussian covariance). There is no indication that the uncertainties on S_8 are biased in any of the scenarios. The effects on parameter uncertainties are also summarised in Figure 4.17. Here the coloured lines show errors on the parameters calculated at the 'badly converged' points and the dotted lines and grey shaded regions indicated fiducial bestfit values and 1σ uncertainties.

These results suggest that in a real situation where we do not know, and choose not to guess, the true cosmology from which the data comes, the iterative procedure will converge to suitable fiducial values close to the constant-S_8 line, and evaluating the covariance matrix at a 'converged' point will generate unbiased estimates of S_8. For safety at least six iterations are advisable to ensure convergence; we have shown that if the starting point is far from the 'true' cosmology then convergence may not be achieved in less than this. However in practice the starting point may often be known to be a good estimate and fewer iterations will suffice.

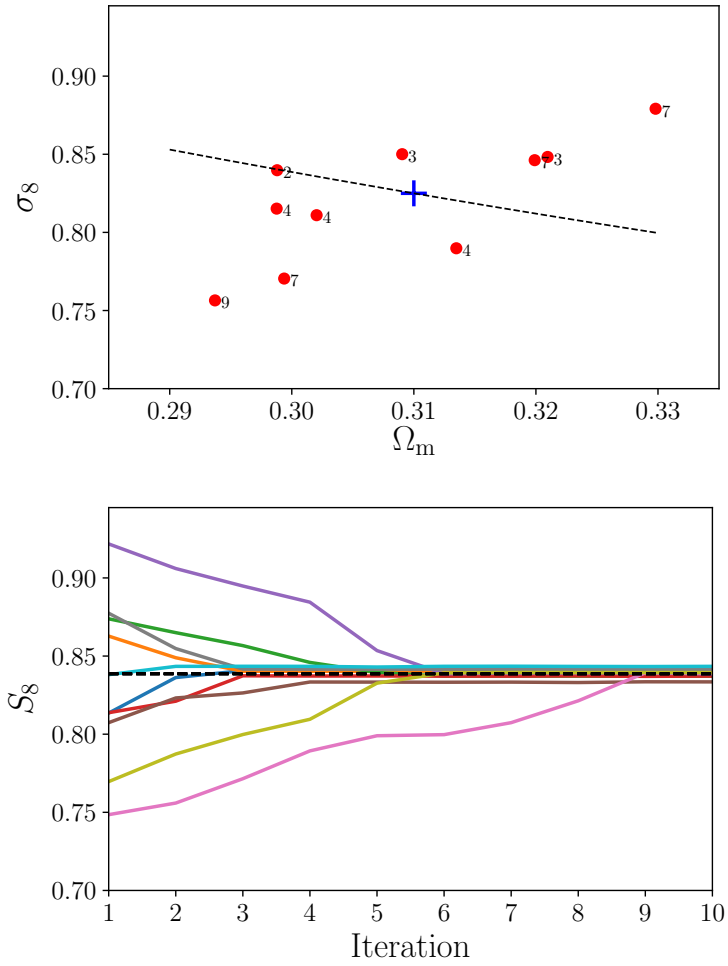

Figure 4.14: Results of iteratively recalculating the covariance matrix when only the Gaussian term of the covariance matrix is used and no noise is added to the data vector. *Top*: Red dots indicate starting positions, numbers indicate the number of iterations required for convergence.The black dashed line marks the line of constant fiducial S_8 and the blue cross indicates the fiducial cosmology at which the data vector was calculated. *Bottom*: Convergence to the final value of S_8 for each of the 10 paths. The black dashed line indicates the fiducial value of S_8.

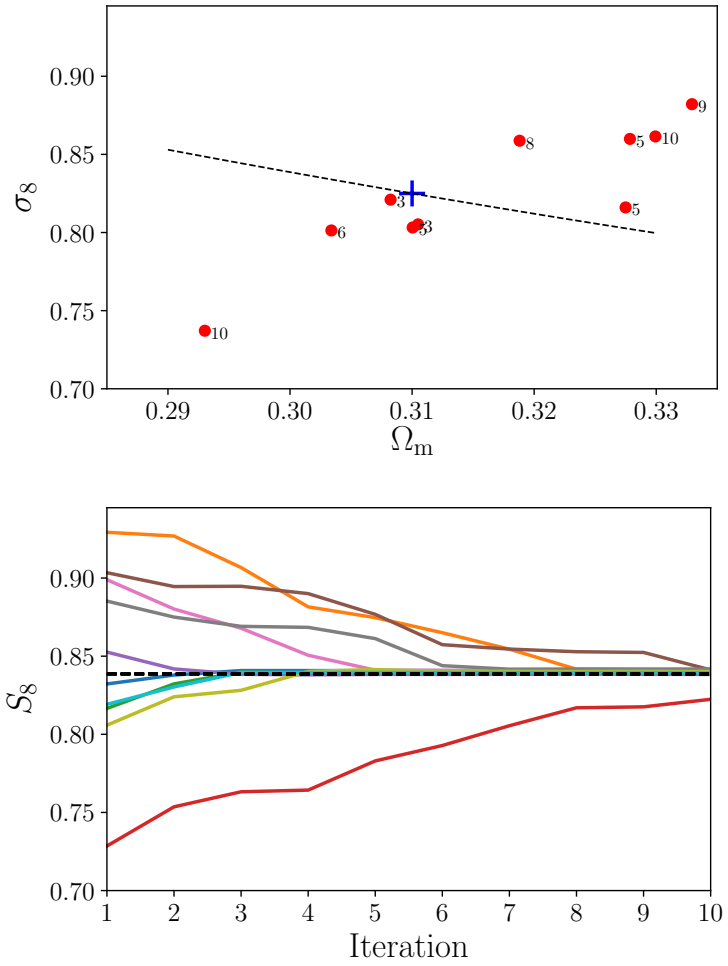

Figure 4.15: Results of iteratively recalculating the covariance matrix when both the Gaussian and supersample terms of the covariance matrix are used and the data vector contains noise sampled from the Gaussian covariance to test the effect of an incorrect covariance model. *Top*: Red dots indicate starting positions, numbers indicate the number of iterations required for convergence. The black dashed line marks the line of constant fiducial S_8 and the blue cross indicates the fiducial cosmology at which the data vector was calculated. *Bottom*: Convergence to the final value of S_8 for each of the 10 paths. The black dashed line indicates the fiducial value of S_8.

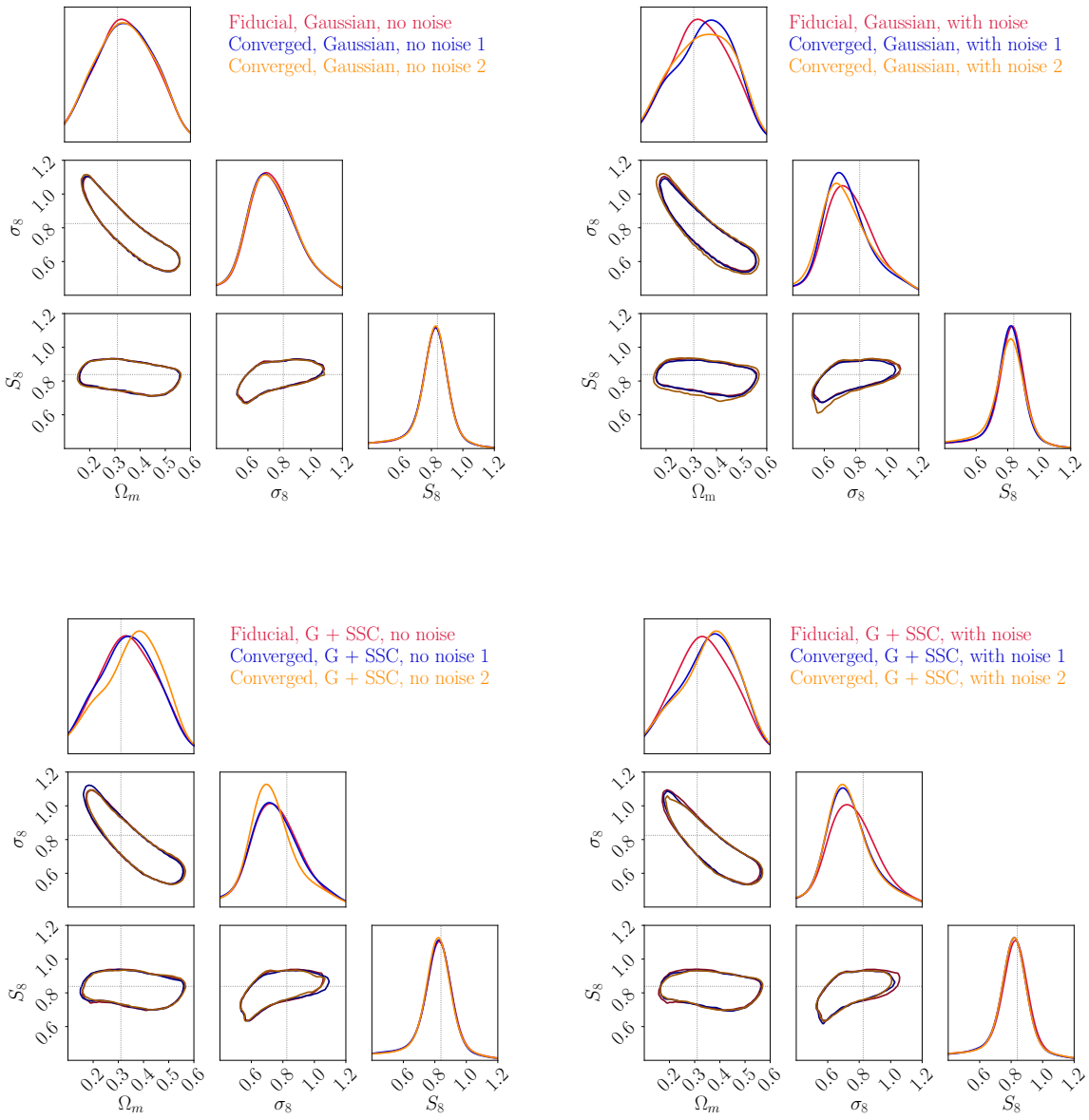

Figure 4.16: Parameter constraints obtained with fiducial covariance matrix (red) and with covariance matrices evaluated at two 'badly converged' points (blue and orange, labelled 1 and 2). Black dotted lines indicate the 'true' parameter values. *Top left*: Gaussian covariance with no noise in data vector. *Top right*: Gaussian covariance with noise in data vector. *Bottom left*: Gaussian plus supersample covariance with no noise in data vector. *Bottom right*: Gaussian plus supersample covariance with noisy data vector. There is no evidence of bias in the estimation of S_8 in any of these cases.

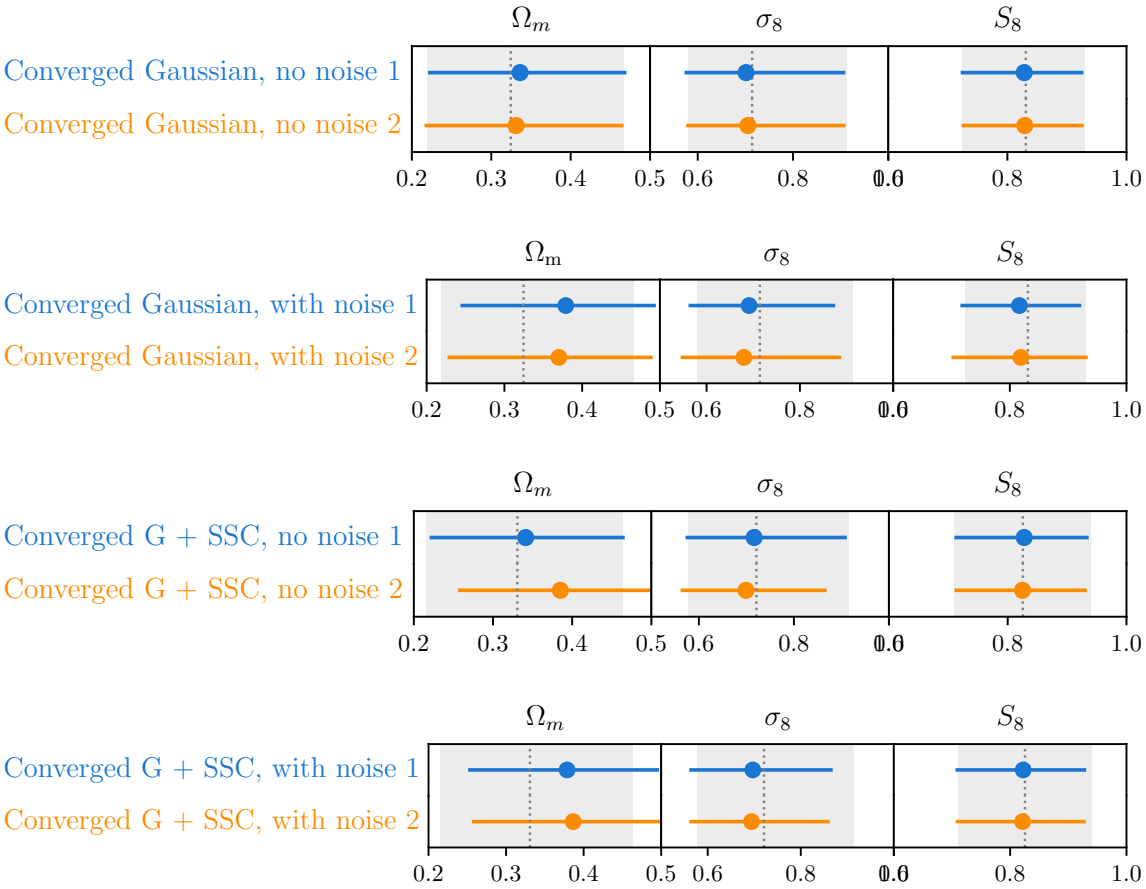

Figure 4.17: Uncertainties arising from covariance matrices evaluated at two 'badly converged' points (blue and orange, labelled 1 and 2). From top to bottom: Gaussian covariance with no noise in data vector, Gaussian covariance with noise in data vector, Gaussian plus supersample covariance with no noise in data vector, Gaussian plus supersample covariance with noise in data vector. Black dotted lines and grey bands indicate the fiducial bestfit values and 1σ uncertainties. Calculating the covariance matrix even at a badly converged point does not introduce any bias to S_8.

4.5 Discussion and conclusions

Form of the likelihood

In this chapter we have demonstrated that in theory the weak lensing power spectrum likelihood is not Gaussian and that assuming a Gaussian likelihood will produce biased parameter estimates, especially of $\Omega_{\rm m}$ and σ_8.

We have described approximations to the true power spectrum likelihood which in CMB and large-scale structure analyses have been found to be preferable to the Gaussian assumption. These are often not feasible for weak lensing because of complicated systematics and survey masks, so it is an open question whether these approximations could be worthwhile. Numerical studies of the impact of using a Gaussian likelihood have not conclusively shown that it is a problem, and in any case alternative methods may be more suitable for weak lensing.

An overall conclusion is that a Gaussian likelihood should never be adopted as a matter of course: the assumption should always be rigorously scrutinised. Then a principled decision can be made either to use a Gaussian likelihood, or to adopt one of the many available strategies which take account of non-Gaussianity.

We have also confirmed the recent conclusions of Sellentin et al. (2018): if the power spectrum has a gamma distribution then the shear two-point correlation function likelihoods are also skewed, so that once again it is not formally correct to use a Gaussian likelihood. We have shown that this model can agree well with simulations, although some aspects need further refinement if the model is to be used in future, particularly the estimation of degrees of freedom and the modelling of ξ_-.

This discussion comes from a very idealised perspective. For weak lensing we do not have access to the true power spectrum, and survey complexities make reality very different from the theoretical picture. In practice the raw correlation functions are not normally used and other options are available which ensure a more Gaussian likelihood. These include data compression, bandpower methods (van Uitert et al. 2018), quadratic estimators (Köhlinger et al. 2017) and combinations of correlation functions such as COSEBIs (Asgari et al. 2019). So in practice the non-Gaussianity of the likelihood is not a problem.

It is unclear whether this non-Gaussianity will become more or less important for next-generation surveys. The naive view might be that as surveys become larger more modes can be included in the likelihood and so by the central limit theorem it approaches Gaussianity. However this is not necessarily the case: larger surveys will include more small and large modes. Small-scale

modes are more affected by nonlinear structure development so can be expected to introduce non-Gaussianity, and it cannot be taken for granted that no additional non-Gaussianity will appear at large scales (Hahn et al. 2019). Our modelling of correlation function likelihoods suggests that increasing the survey size does not remove all non-Gaussianity. Furthermore, non-Gaussianity will become increasingly important if higher-order statistics are analysed.

The non-Gaussianity of the likelihood raises some interesting questions which are discussed in detail in Sellentin et al. (2018). How can we make sense of the fact that the most likely parameter values are not equal to their mean values? Can we accept that our analysis is likely to produce values which are lower than the average? Sellentin et al. (2018) suggest that the skewed likelihood is not a problem for a Bayesian analyst who will accept the maximum likelihood values, conditional on the priors and likelihood. It would seem slightly inconsistent in practice to declare a Bayesian stance but then insist on a Gaussian likelihood to ensure that the maximum likelihood and mean values are the same. However it is not inconsistent to deliberately choose a Gaussian likelihood for practical reasons, provided the consequences have been considered properly.

Parameter-dependence of covariance matrices

It is likely that Gaussian likelihoods will continue to be used for real weak lensing surveys because of the complexity of the true likelihood. We have clarified the correct form of Gaussian likelihood to use and shown that it is incorrect to re-estimate the covariance matrix at each step of a likelihood analysis. Instead a fixed covariance matrix evaluated at an appropriate fiducial cosmology should be used. There is considerable confusion about this in the literature and mixed findings about the impact of a using a parameter-dependent covariance matrix, mostly based on the misconception that a parameter-dependent covariance matrix improves parameter constraints.

Despite this, the cosmology used to estimate the covariance matrix does matter, but is rarely explicitly discussed. Evaluating the (parameter-independent) covariance matrix far from the cosmology preferred by the data will alter the likelihood contours and lead to incorrect inferences about parameters (Eifler et al. 2009). We have developed an emulator for the weak lensing power spectrum covariance matrix which can be used iteratively to find a suitable cosmology at which to evaluate the covariance matrix. We have shown that the iterative process converges to the constant-S_8 degeneracy line of the data in a small number of steps, from any reasonable starting cosmology. This provides a systematic way of ensuring that the covariance is calculated at an appropriate point. The computational cost of the iterative procedure is relatively small and is justified by the elimination of one small source of uncertainty in the likelihood analysis.

Appendices

4A Emulator

Background

Emulation in cosmology is the process of inferring the values of a quantity or model at multiple positions in parameter space from the values at a small set of 'training' positions. It is particularly suitable for parameter estimation where the likelihood must be estimated many thousands of times. Simulations are generally far too computationally expensive for this purpose and analytical methods such as perturbation theory, the halo model or fitting functions calibrated to simulations may have limited accuracy, be valid only over certain ranges, or still be demanding to calculate.

Emulation was introduced to cosmology by Heitmann et al. (2006) to estimate the matter power spectrum, but the method has very general application. For example, emulators have been constructed for the nonlinear matter power spectrum (Heitmann et al. 2010; Lawrence et al. 2010), the galaxy power spectrum (Kwan et al. 2015), weak lensing peak statistics (Martinet et al. 2017), the boost factor (the ratio between nonlinear and linear contributions to the matter power spectrum) (Knabenhans et al. 2019), the 1D Lyman-alpha forest flux power spectrum (Bird et al. 2019; Rogers et al. 2019) and likelihoods (McClintock and Rozo 2019; Pellejero-Ibañez et al. 2019).

Two key decisions to be made when developing an emulator are the experimental design - the selection of training points at which to obtain accurate estimates - and the method of inferring values at positions between the training points.

For the experimental design it is preferable to use some form of stratified sampling with good space-filling properties, especially if simulations are being used. A common approach is Latin hypercube sampling. A Latin hypercube design is an $n \times m$ matrix in which each column is a unique random permutation of the set $\{1, \ldots, n\}$. A refinement is to supplement a Latin hypercube with Bayesian optimisation methods (Brochu et al. 2010) which build up the training set iteratively by maximising a problem-specific acquisition function. This balances the choice of additional points between those in regions of high uncertainty and those most likely to improve the emulation.

Various methods have been used in the literature to infer intermediate points. Many recent cosmological emulators use Gaussian process regression (Schulz et al. 2018) which not only provides accurate emulation but also facilitates careful control of errors. Other methods which have been used include linear interpolation (White and Padmanabhan 2015), polynomial interpolation

(Schneider et al. 2011)), radial basis functions (Martinet et al. 2017) and sparse polynomial chaos expansion (Knabenhans et al. 2019).

Emulator for the weak lensing power spectrum covariance matrix

Our emulator is designed specifically to provide an efficient way to carry out the iterative process described in section 4.4 and to demonstrate proof of concept. This determines some features of its design. Firstly, it is intended to be used in a likelihood analysis based on the KiDS-450 analysis in Köhlinger et al. (2017) and van Uitert et al. (2018). In particular this determines the prior ranges over which the emulator needs to work. For simplicity we consider only $\Omega_{\rm m}$ and σ_8, keeping all other parameters fixed at fiducial values. In practice these two parameters are derived parameters in the KiDS-450 analysis, so the emulator instead uses $\Omega_{\rm c}h^2$ and $\ln 10^{10} A_{\rm s}$.

The emulator is trained on analytical calculations of the weak lensing power spectrum covariance matrix which are based on the halo model, as used in van Uitert et al. (2018). A major advantage compared with using simulations to generate the training data is that it is not unduly expensive to work with a large set of training points. It is also relatively easy to try out different sets of training points or, for example, to produce emulators for different components of the covariance.

Since we can be generous with the number of training points we use a regular grid covering the whole parameter space. However the KiDS-450 priors are very wide: [0.0, 0.99] for $\Omega_{\rm c}h^2$ and [1.7, 5.0] for $\ln 10^{10} A_{\rm s}$. To ensure that the (known) high-posterior region is covered by a large enough set of points our final grid is 9×9. A final consideration is that a rectangular grid in $\Omega_{\rm c}h^2$ and $\ln 10^{10} A_{\rm s}$ space is not rectangular in $\Omega_{\rm m}$ and σ_8 space (see Figure 4.18) and the grid needed to cover the $\Omega_{\rm m} - \sigma_8$ plane adequately. Adding extra points along the $\Omega_{\rm m} - \sigma_8$ degeneracy line did not produce a worthwhile increase in accuracy, but we confirmed that reducing the grid to 7×7 would still be satisfactory. Below this, accuracy in the high-posterior region falls.

We use the whole training set to train the emulator but when it is used to interpolate we impose narrower bounds: $\Omega_c h^2$ [0.09, 0.40] and $\ln 10^{10} A_{\rm s}$ [2.4, 4.5]. Outside these ranges the emulated value is set equal to the nearest boundary value to eliminate unphysical values. It can be particularly problematic to calculate the covariance at the lower extremes of the KiDS priors, especially of $\ln 10^{10} A_{\rm s}$, as these cosmologies represent essentially homogeneous, structureless universes. However these regions are also far from the regions where the posterior is large so it is not a problem if they are approximated by nearby values.

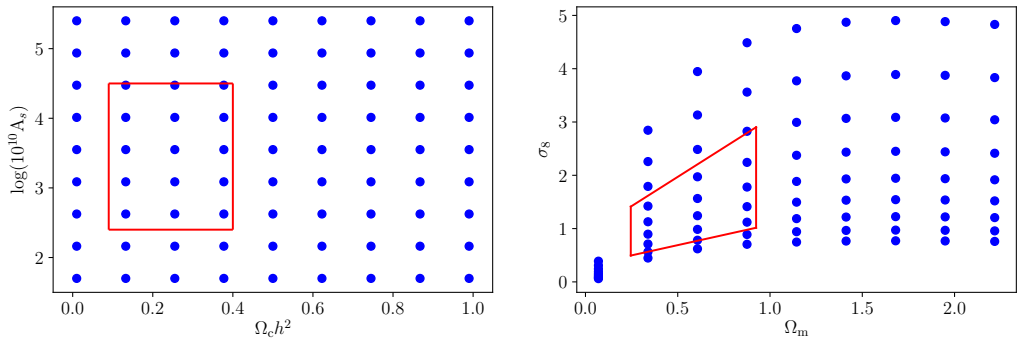

Figure 4.18: *Left*: Rectangular grid of training points in the $\Omega_{\rm c}h^2$–$\ln 10^{10}A_{\rm s}$ plane, covering the whole KiDS-450 prior ranges. Red lines indicate the boundaries of the smaller more realistic region used for interpolation. *Right*: Translation of the rectangular grid into the $\Omega_{\rm m}$–σ_8 plane. Red lines indicate the boundaries (approximated as straight lines) of the smaller more realistic region used for interpolation.

Clearly this simple grid configuration is not particularly efficient. Even a basic Latin hypercube would be more parsimonious and probably equally accurate. Certainly a more efficient design would be necessary if the emulator was expanded to more than two dimensions in parameter space.

Unlike other covariance emulators which derive the covariance from an emulated power spectrum (for example Harnois-Deraps et al. (2019)), every element of the covariance matrix is emulated separately. With 4 redshift bins and 5 ℓ-bins, 1275 unique elements are necessary to reproduce the whole covariance matrix. To ensure that the emulated covariance matrix is positive definite (essential since it must be inverted in the likelihood analysis), we perform a Cholesky decomposition in which the covariance matrix $\mathbf{C}$ is decomposed into a lower triangular matrix $\mathbf{L}$ and a diagonal matrix $\mathbf{D}$ so that

$$\mathbf{C} = \mathbf{L}\mathbf{D}\mathbf{L}^{\rm T}\,, \tag{4.38}$$

and then emulate the diagonal and lower triangular parts separately. Also we emulate the logarithm of the diagonal elements to ensure that their emulated values are positive. In theory the sign of the elements of $\mathbf{L}$ is not well-defined and could vary between adjacent points in parameter space. In practice this was not a problem since the covariance is well-behaved over the two-parameter space.

We use radial basis functions to perform the emulation. These are a special case of the more general Gaussian processes which are commonly used in cosmological emulators. If we have m points $x_1 \ldots x_m$ at which the true values $f(x_1) \ldots f(x_m)$ are known, then the value at another

point x is approximated by

$$g(x) = \sum_{j=1}^{m} \lambda_j \phi(||x - x_j||) \,, \tag{4.39}$$

where ϕ is a radial basis function and $||x - x_j||$ is a norm or distance which can be measured in different ways but is often the Euclidean norm. This is a system of linear equations for the weights λ_j. It can be shown that it can always be solved for values which make $g(x_j) = f(x_j)$ for all j. Many different forms can be used for the function ϕ. We use the multiquadric form

$$\phi(r) = \sqrt{r^2 + c^2} \,, \tag{4.40}$$

where r is the Euclidean distance and c is an adjustable hyper-parameter.

Radial basis functions are a common choice for interpolation and are known to perform well if the function being emulated varies smoothly, which is the case for any individual covariance matrix element. Unlike some other interpolation methods radial basis functions do not require the training set to be a grid, and they reproduce the training values exactly. They are quick to compute and are guaranteed to find a solution for all points. Their main disadvantage is that they do not extrapolate well so perform poorly at the edges of the training region. Since we cut off the more extreme parts of the plane in any case, we avoid this problem. We considered using more sophisticated Gaussian process models with more adjustable parameters. After some preliminary investigations we concluded that these were not necessary for our simple problem and that radial basis functions performed adequately, as discussed below.

Accuracy of the emulator

We are interested primarily in the accuracy of the covariance matrix as a whole, rather than individual elements, so choose to measure performance through the accuracy of the determinant, trace and Fröbenius norm (Reischke et al. 2016), which for a matrix $\mathbf{C}$ is defined as

$$\|\mathbf{C}\| = \sqrt{\sum_i \sum_j C_{ij}^2} \,. \tag{4.41}$$

In practice we focus on the inverse covariance matrix (the precision matrix) since this is what is used in the likelihood. We estimate the analytical covariance matrix (Gaussian terms only) at 48 random 'comparison' points roughly spaced around the maximum likelihood contours and compared these

with their emulated counterparts. Figure 4.19 shows the error in the determinant of the precision matrix at each of these points, where the error is defined as the absolute difference between the emulated and exact determinants relative to the exact value. Large blue dots show comparison points; small black dots show the training grid. This figure shows no particular pattern in the size of the errors apart from some indication that higher errors occur at low values of $\Omega_c h^2$. Repeating with different sets of randomly-chosen points confirmed this conclusion. Figure 4.20 shows percentage

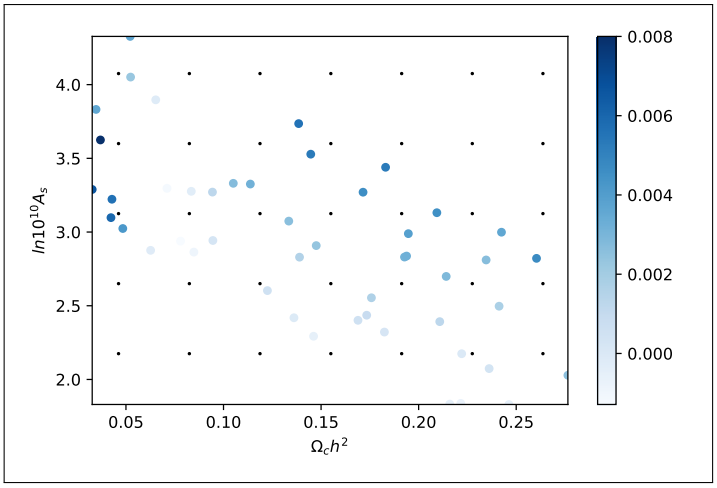

Figure 4.19: Absolute errors in the determinant of the emulated inverse covariance matrix, evaluated at 48 random points in the $\ln 10^{10} A_s$–$\Omega_c h^2$ plane. Large blue dots indicate points at which the covariance matrix was emulated. Small black dots indicate the grid used for emulation. Errors are the absolute value of the difference between the emulated and exact determinant relative to the exact value.

errors in the determinant, norm, and trace of the inverse covariance matrix. The errors in the determinant are all below 1%. Errors in the norm and trace are mainly less than 10%. The accuracy of the emulated determinant is also demonstrated in Figure 4.21 which compares the analytical and emulated determinants of the precision matrices, showing good agreement. (The unpopulated white regions arise because the emulator uses $\Omega_c h^2$ and $\ln 10^{10} A_s$ rather than Ω_m and σ_8).

Our ultimate aim is to use the emulated covariance matrix to calculate the likelihood. To demonstrate its accuracy for this purpose, Figure 4.22 shows the log likelihood at a fixed value of $\ln 10^{10} A_s$ and 10 equally spaced values of $\Omega_c h^2$. The left panel shows the results if the likelihood is calculated analytically at each point and the right panel shows the results of using an emulated covariance matrix calculated at each of the points. The error in the log likelihood is less than 0.5% at all points and the two plots are indistinguishable by eye. Similar results are obtained if $\Omega_c h^2$ is fixed and $\ln 10^{10} A_s$ varied. Although we recalculate the likelihood at multiple points, this is a test of the accuracy of the emulator, not of the effect of a parameter-dependent covariance matrix.

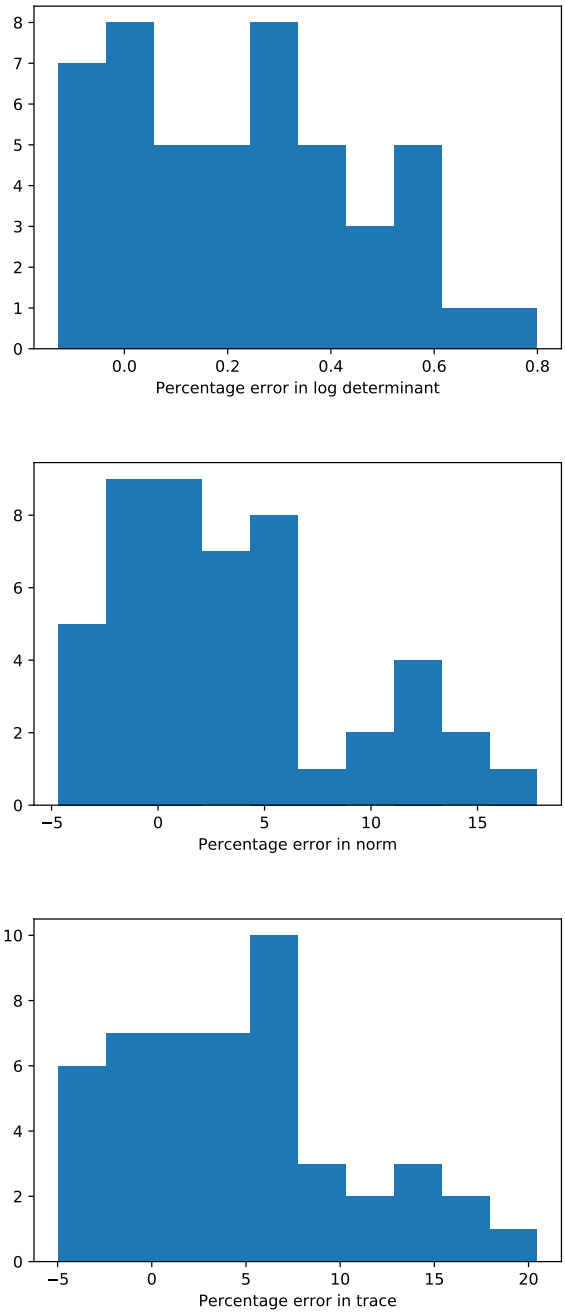

Figure 4.20: Percentage errors in the emulated inverse covariance matrix evaluated at 48 random points. *Top*: Determinant. *Centre*: Fröbenius norm. *Bottom*: Trace.

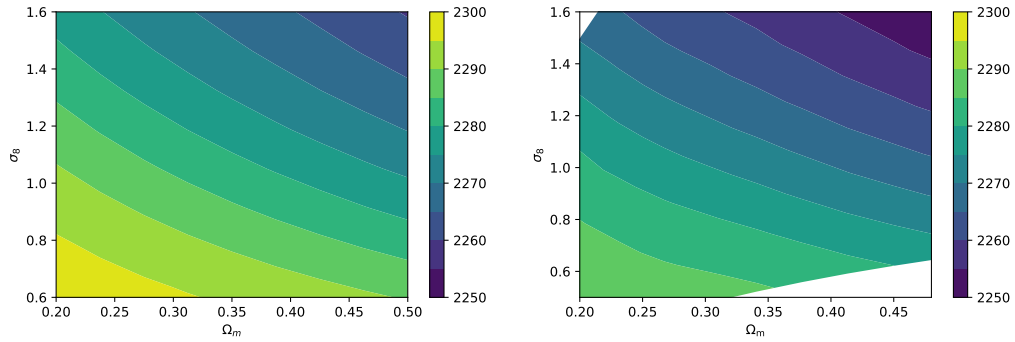

Figure 4.21: Comparison between analytical and emulated log determinant of inverse covariance matrix in the $\Omega_{\rm m}$–σ_8 plane. *Left*: Analytically calculated. *Right*: Emulated. (The white regions arise because the emulator uses $\Omega_{\rm c}h^2$ and $\ln 10^{10} A_{\rm s}$ rather than $\Omega_{\rm m}$ and σ_8).

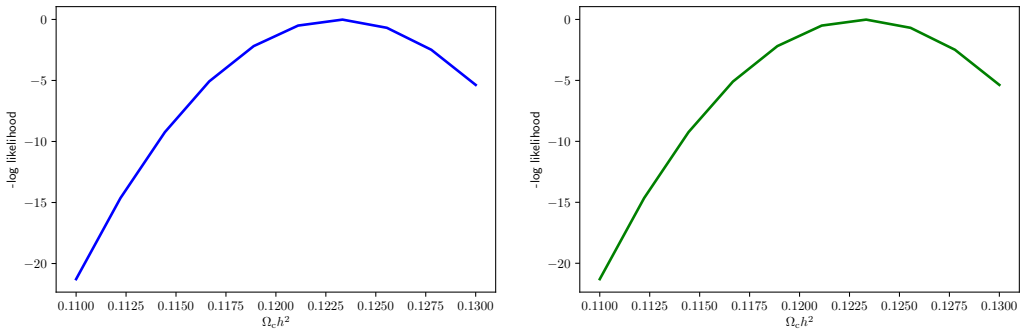

Figure 4.22: Comparison between likelihood estimated with analytical and emulated covariance matrices. Shown are the log likelihood calculated at the fixed value $\ln 10^{10} A_{\rm s} = 2.99$ for varying values of $\Omega_{\rm c}h^2$. *Left*: Using analytically calculated covariance matrix. *Right*: Using covariance matrix emulated at the same values of $\Omega_{\rm c}h^2$ and $\ln 10^{10} A_{\rm s}$.

4B Additional iteration results

This appendix provides additional results from the iterative process discussed in Section 4.4.4. Specifically we show the convergence of the process when we use only the Gaussian part of the covariance matrix but a noisy data vector and when we use the Gaussian and supersample terms with no noise in the data vector.

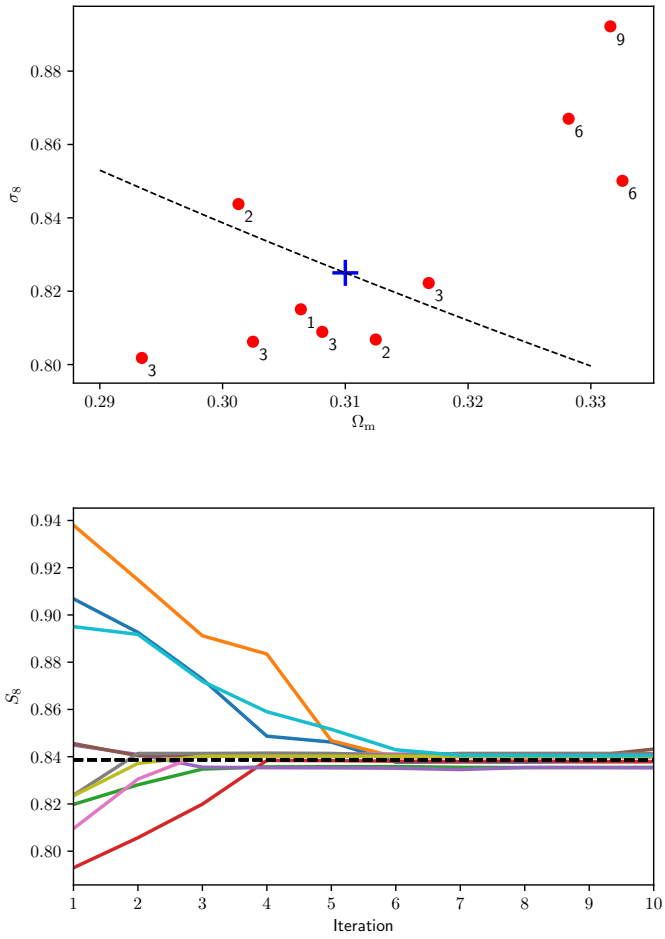

Figure 4.23: Results of iteratively recalculating the covariance matrix when only the Gaussian term of the covariance matrix is used and the data vector contains noise sampled from the Gaussian covariance. *Top*: Red dots indicate starting positions, numbers indicate the number of iterations required for convergence. The black dashed line is the line of constant S_8, and the blue cross indicates the fiducial cosmology. *Bottom*: Convergence to the final value of S_8 for each of the 10 paths. The black dashed line indicates the fiducial value of S_8.

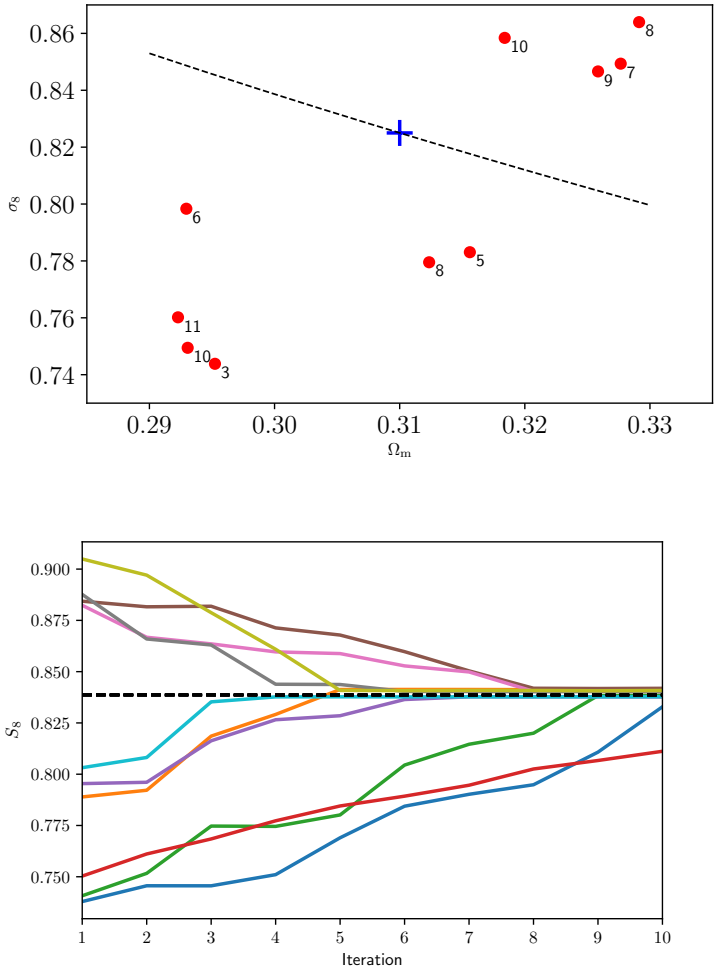

Figure 4.24: Results of iteratively recalculating the covariance matrix when both the Gaussian and supersample terms of the covariance matrix are used and the data vector is not noisy. *Top*: Red dots indicate starting positions, numbers indicate the number of iterations required for convergence. The black dashed line is the line of constant S_8, and the blue cross indicates the fiducial cosmology. *Bottom*: Convergence to the final value of S_8 for each of the 10 paths. The black dashed line indicates the fiducial value of S_8.

5 Conclusions and future work

Twenty years ago weak gravitational lensing was an almost speculative idea. The effect had been detected and measured but its analysis was far from being routine. Today lensing is one of a handful of primary cosmological probes with an impressive array of results from recent surveys such as CFHTLens, DES and KiDS. Over the intervening two-decade period many methods of measurement and analysis have been developed and honed and are now standard. We know how to 'do' weak lensing.

The next stage in the evolution of weak lensing is to prepare for the large surveys like *Euclid* which will shortly come on stream. This means refining current methodologies, but also exploring new approaches which were not previously contemplated, either because they learn from earlier experience or because they simply were not feasible with past data and technologies. In this spirit, this thesis mainly considers ideas which even for next-generation surveys are not mainstream, but which *Euclid*-like surveys could render viable and valuable.

In Chapter 2 our aim was to discover whether one generally overlooked consequence of weak gravitational lensing will be detectable in next-generation surveys. The point was to identify whether this signal, the displacement of observed source positions, could provide additional information from data which anyway is collected by the survey. Dodelson et al. (2008) had already investigated how this displacement affects the two-point correlation functions of various observables but found the effect to be small except for high redshift observations with rapidly varying correlation functions. We extended the analysis to three-point correlation functions but, not surprisingly, found that the effect was again small. Even a survey like *Euclid* will not observe enough bright high-redshift galaxies or quasars to allow us to detect the effect in the correlation function. This is not necessarily the end of the story, however. In the short term the methodology could be applied to other cosmological three-point functions. In the longer term future generations of galaxy surveys will surely probe even higher redshifts and capture many more bright objects so that the effect may become detectable and useful.

Chapter 3 focused on the weak lensing bispectrum. The idea of using the bispectrum alongside the power spectrum to improve cosmological parameter constraints has been around for at least a decade. In the era of relatively small galaxy surveys, which were not optimised for weak lensing, reducing statistical errors was a pressing issue and it was expected, for example, that the bispectrum

could help break parameter degeneracies. Using a variety of approaches and assumptions several previous authors concluded that in principle the bispectrum can add extra information, and we confirmed these findings. We also clarified some details. In particular we derived all the terms of the weak lensing supersample covariance and showed that the one-halo term is dominant. We also confirmed that the in-survey non-Gaussian covariance is sub-dominant and can be neglected. Finally we showed that most of the gain from the bispectrum can be achieved with a small number of tomographic bins, provided that the power spectrum covariance is based on many bins.

Even with these and other simplifications the bispectrum is more complicated than the power spectrum, and its covariance involves even more complexity. The trade-off between the cost of dealing with this complexity and the benefit of the extra information is implicit in recent observational weak lensing studies which almost without exception use only two-point statistics, and in the simplifications invariably introduced into three-point analyses and forecasts. Subjectively we concluded that the reduction in statistical errors may not be enough to justify the cost of measuring and analysing the bispectrum; squeezing out every possible improvement is less important for a *Euclid*-like survey with an abundance of high-quality observations. However the bispectrum may still have its uses and in the second part of Chapter 3 we looked at a different rationale for using three-point statistics: to control systematic uncertainties. Dealing with systematics, whether observational (for example due to characteristics of the survey instrument) or astrophysical, is one of the major concerns of current weak lensing research (Mandelbaum 2018) as the community prepares for next-generation surveys.

Before our work there was some existing evidence that systematics affect two-point and three-point statistics in different ways, but in practice there had been very little detailed analysis of the impact of systematics on three-point statistics. In Chapter 3 we made a start on remedying this situation and it is this part of the thesis which most strongly points to future research with real potential. Our analysis of two major systematic uncertainties, intrinsic alignments and photometric redshift biases, suggests that three-point statistics have much to offer. In particular, we have shown that using the bispectrum as well as the power spectrum can allow self-calibration without excessively degrading parameter constraints.

This is a promising finding but more work needs to be done to demonstrate that self-calibration with three-point statistics is a practical proposition for *Euclid*. At the end of Chapter 3 we list several immediate ideas for further research, including refining our intrinsic alignment model and testing it with simulations or data; investigating other systematics such as multiplicative shear bias and baryonic effects; and adapting the method to appropriate three-point correlation functions

which are better suited to observational studies. We also suggest that more work should be done to confirm the validity of current bispectrum models over the scales of interest and in particular to investigate the model recently published by Takahashi et al. (2019) which probes smaller scales. This programme of work could lead to firmer forecasts and concrete proposals for implementing the method when data from *Euclid* becomes available.

One final comment on the use of three-point statistics is that it is often possible to build on solid methods developed for two-point statistics. This is particularly true for systematics where modelling approaches which have worked well for two-point statistics can easily be extended – for example modelling of intrinsic alignments or baryonic effects. The underlying physics is the same and there is no need to re-invent the wheel.

Slightly changing tack, Chapter 4 addressed two assumptions that are almost universally made in analyses of current weak lensing surveys: that the power spectrum (or two-point correlation function) likelihood is Gaussian, and that the covariance matrix should be treated as parameter-dependent. There are often good pragmatic reasons for making these assumptions but they can lead to flawed conclusions. In particular, even if it is valid to use a Gaussian likelihood, combining this with a cosmology-dependent covariance matrix can result in biased parameter estimates. This matters for future surveys: given the greater statistical accuracy of *Euclid* it would be unfortunate to inadvertently introduce avoidable bias. The question of the non-Gaussianity of the likelihood will also not go away. We built on a hierarchical model developed by Sellentin et al. (2018) which suggests that, at least theoretically, non-Gaussianity of the likelihood will matter even for a *Euclid*-like survey. We found that this model, especially its relationship to simulations, would need to be refined if it was to be used for further research. Nevertheless we conclude that the consequences of assuming a Gaussian likelihood should always be carefully weighed up.

Although it is not necessary (and may even be wrong) to use a cosmology-dependent covariance matrix, the position in parameter space at which the matrix is evaluated does matter. In the final part of Chapter 4 we demonstrated a way to derive this fiducial cosmology robustly and checked that it did not produce biased parameter estimates. The method relies on a simple emulator of the weak lensing power spectrum covariance. This was essentially a prototype based on only two cosmological parameters, but potentially paves the way for a more sophisticated multidimensional emulator if a cosmology-dependent covariance matrix is required.

In summary, the work in this thesis provides pointers towards practical approaches for using three-point statistics in future weak lensing surveys. We have found that some potential avenues are not worth exploring further but have also identified others which could be very fruitful.

Bibliography

Abbott, B. P., Abbott, R., Abbott, T., Abernathy, M., Acernese, F., Ackley, K., Adams, C., Adams, T., Addesso, P., Adhikari, R. et al. (2016), 'Observation of gravitational waves from a binary black hole merger', *Physical Review Letters* **116**(6), 061102.

Abbott, T., Abdalla, F., Alarcon, A., Aleksić, J., Allam, S., Allen, S., Amara, A., Annis, J., Asorey, J., Avila, S. et al. (2018), 'Dark Energy Survey year 1 results: cosmological constraints from galaxy clustering and weak lensing', *Physical Review D* **98**(4), 043526.

Abell, P., J., A. and et al. (2009), 'LSST Science Book, version 2.0'.

Ade, P. A., Aghanim, N., Armitage-Caplan, C., Arnaud, M., Ashdown, M., Atrio-Barandela, F., Aumont, J., Baccigalupi, C., Banday, A. J., Barreiro, R. et al. (2014), 'Planck 2013 results. XV. CMB power spectra and likelihood', *Astronomy & Astrophysics* **571**, A15.

Ade, P. A., Aghanim, N., Arnaud, M., Ashdown, M., Aumont, J., Baccigalupi, C., Banday, A., Barreiro, R., Bartlett, J., Bartolo, N. et al. (2016), 'Planck 2015 results XIII. Cosmological parameters', *Astronomy & Astrophysics* **594**, A13.

Adhikari, S., Jeong, D. and Shandera, S. (2016), 'Constraining primordial and gravitational mode coupling with the position-dependent bispectrum of the large-scale structure', *Physical Review D* **94**(8), 083528.

Aghamousa, A., Aguilar, J. and et al. (2016), 'The DESI Experiment Part I: Science, Targeting, and Survey Design'.

Akrami, Y., Arroja, F., Ashdown, M., Aumont, J., Baccigalupi, C., Ballardini, M., Banday, A., Barreiro, R., Bartolo, N., Basak, S. et al. (2018), 'Planck 2018 results. I. Overview and the cosmological legacy of Planck', *arXiv preprint arXiv:1807.06205* .

Alam, S., Ata, M., Bailey, S., Beutler, F., Bizyaev, D., Blazek, J. A., Bolton, A. S., Brownstein, J. R., Burden, A., Chuang, C.-H. et al. (2017), 'The clustering of galaxies in the completed sdss-iii baryon oscillation spectroscopic survey: cosmological analysis of the dr12 galaxy sample', *Monthly Notices of the Royal Astronomical Society* **470**(3), 2617–2652.

Albrecht, A., Bernstein, G., Cahn, R., Freedman, W. L., Hewitt, J., Hu, W., Huth, J., Kamionkowski, M., Kolb, E. W., Knox, L. et al. (2006), 'Report of the dark energy task force', *arXiv preprint astro-ph/0609591* .

Alpher, R. A. and Herman, R. (1948), 'Evolution of the universe', *Nature* **162**(4124), 774.

Alsing, J., Wandelt, B. and Feeney, S. (2018), 'Massive optimal data compression and density estimation for scalable, likelihood-free inference in cosmology', *Monthly Notices of the Royal Astronomical Society* **477**(3), 2874–2885.

Amendola, L., Appleby, S. and et al. (2013), 'Cosmology and fundamental physics with the Euclid satellite', *Living Rev. Relativity* .

Anderson, L., Aubourg, E., Bailey, S., Bizyaev, D., Blanton, M., Bolton, A. S., Brinkmann, J., Brownstein, J. R., Burden, A., Cuesta, A. J. et al. (2012), 'The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Data Release 9 spectroscopic galaxy sample', *Monthly Notices of the Royal Astronomical Society* **427**(4), 3435–3467.

Asgari, M. and Schneider, P. (2015), 'A new data compression method and its application to cosmic shear analysis', *Astronomy & Astrophysics* **578**, A50.

Asgari, M., Tröster, T., Heymans, C., Hildebrandt, H., Busch, J. L. v. d., Wright, A. H., Choi, A., Erben, T., Joachimi, B., Joudaki, S. et al. (2019), 'KiDS+ VIKING-450 and DES-Y1 combined: Mitigating baryon feedback uncertainty with COSEBIs', *arXiv preprint arXiv:1910.05336* .

Audren, B., Lesgourgues, J., Benabed, K. and Prunet, S. (2013), 'Conservative Constraints on Early Cosmology: an illustration of the Monte Python cosmological parameter inference code', *JCAP* **1302**, 001.

Baldauf, T., Seljak, U., Senatore, L. and Zaldarriaga, M. (2016), 'Linear response to long wavelength fluctuations using curvature simulations', *Journal of Cosmology and Astroparticle Physics* **2016**(09), 007.

Bardeen, J. M., Szalay, A., Kaiser, N. and Bond, J. (1985), 'The statistics of peaks of Gaussian random fields', *The Astrophysical Journal* **304**, 15–61.

Barreira, A. (2019), 'The squeezed matter bispectrum covariance with responses', *Journal of Cosmology and Astroparticle Physics* **2019**(03), 008.

Barreira, A., Krause, E. and Schmidt, F. (2018*a*), 'Accurate cosmic shear errors: do we need ensembles of simulations?', *Journal of Cosmology and Astroparticle Physics* **2018**(10), 053.

Barreira, A., Krause, E. and Schmidt, F. (2018*b*), 'Complete super-sample lensing covariance in the response approach', *Journal of Cosmology and Astroparticle Physics* **2018**(06), 015.

Bartelmann, M. and Maturi, M. (2016), 'Weak gravitational lensing', *arXiv preprint arXiv:1612.06535* .

Bartelmann, M. and Schneider, P. (2001), 'Weak gravitational lensing', *Physics Reports* **340**(4-5), 291–472.

Bassett, B. and Hlozek, R. (2010), *Dark energy: observational and theoretical approaches. Baryon acoustic oscillations*, Cambridge University Press: Cambridge, UK.

Bayes, T. (1763), 'LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFRS', *Philosophical transactions of the Royal Society of London* .

Bernardeau, F., Colombi, S., Gaztanaga, E. and Scoccimarro, R. (2002), 'Large-scale structure of the universe and cosmological perturbation theory', *Physics reports* **367**(1-3), 1–248.

Bernardeau, F., Van Waerbeke, L. and Mellier, Y. (2003), 'Patterns in the weak shear 3-point correlation function', *Astronomy & Astrophysics* **397**(2), 405–414.

Bertacca, D., Maartens, R. and Clarkson, C. (2014), 'Observed galaxy number counts on the lightcone up to second order: I. Main result', *Journal of Cosmology and Astroparticle Physics* **9**.

Bertacca, D., Raccanelli, A., Bartolo, N., Liguori, M., Matarrese, S. and Verde, L. (2018), 'Relativistic wide-angle galaxy bispectrum on the light cone', *Physical Review D* **97**(2), 023531.

Bird, S., Rogers, K. K., Peiris, H. V., Verde, L., Font-Ribera, A. and Pontzen, A. (2019), 'An emulator for the Lyman-α forest', *Journal of Cosmology and Astroparticle Physics* **2019**(02), 050.

Blake, C. and Glazebrook, K. (2003), 'Probing dark energy using baryonic oscillations in the galaxy power spectrum as a cosmological ruler', *The Astrophysical Journal* **594**(2), 665.

Blanchard, A. and Schneider, J. (1987), 'Gravitational lensing effect on the fluctuations of the cosmic background radiation', *Astronomy and Astrophysics* **184**, 1–6.

Blas, D., Lesgourgues, J. and Tram, T. (2011), 'The cosmic linear anisotropy solving system (CLASS). Part II: approximation schemes', *Journal of Cosmology and Astroparticle Physics* **2011**(07), 034.

Blazek, J. A., MacCrann, N., Troxel, M. and Fang, X. (2019), 'Beyond linear galaxy alignments', *Physical Review D* **100**(10), 103506.

Blazek, J., Vlah, Z. and Seljak, U. (2015), 'Tidal alignment of galaxies', *Journal of Cosmology and Astroparticle Physics* **2015**(08), 015.

Bohm, V., Schmittfull, M. and Sherwin, B. D. (2016), 'Bias to CMB lensing measurements from the bispectrum of large-scale structure', *Physical Review D* **94**(4).

Bond, J. and Efstathiou, G. (1987), 'The statistics of cosmic background radiation fluctuations', *Monthly Notices of the Royal Astronomical Society* **226**(3), 655–687.

Bond, J., Jaffe, A. H. and Knox, L. (2000), 'Radical compression of cosmic microwave background data', *The Astrophysical Journal* **533**(1), 19.

Bonvin, C. and Durrer, R. (2011), 'What galaxy surveys really measure', *Physical Review D* **84**(6), 16.

Bose, B., Byun, J., Lacasa, F., Dizgah, A. M. and Lombriser, L. (2019), 'Modelling the non-linear bispectrum in modified gravity', *arXiv preprint arXiv:1909.02504* .

Box, G. E. and Cox, D. R. (1964), 'An analysis of transformations', *Journal of the Royal Statistical Society: Series B (Methodological)* **26**(2), 211–243.

Bridle, S. and King, L. (2007), 'Dark energy constraints from cosmic shear power spectra: impact of intrinsic alignments on photometric redshift requirements', *New Journal of Physics* **9**(12), 444.

Brinckmann, T. and Lesgourgues, J. (2019), 'MontePython 3: boosted MCMC sampler and other features', *Physics of the Dark Universe* **24**, 100260.

Brochu, E., Cora, V. M. and De Freitas, N. (2010), 'A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning', *arXiv preprint arXiv:1012.2599* .

Byun, J., Eggemeier, A., Regan, D., Seery, D. and Smith, R. E. (2017), 'Towards optimal cosmological parameter recovery from compressed bispectrum statistics', *Monthly Notices of the Royal Astronomical Society* **471**(2), 1581–1618.

Carron, J. (2013), 'On the assumption of Gaussianity for cosmological two-point statistics and parameter dependent covariance matrices', *Astronomy & Astrophysics* **551**, A88.

Chan, K. C. and Blot, L. (2017), 'Assessment of the information content of the power spectrum and bispectrum', *Physical Review D* **96**(2), 023528.

Chan, K. C., Dizgah, A. M. and Noreña, J. (2018), 'Bispectrum supersample covariance', *Physical Review D* **97**(4), 043532.

Chiang, C.-T., Wagner, C., Schmidt, F. and Komatsu, E. (2014), 'Position-dependent power spectrum of the large-scale structure: a novel method to measure the squeezed-limit bispectrum', *Journal of Cosmology and Astroparticle Physics* **2014**(05), 048.

Coe, D. (2009), 'Fisher matrices and confidence ellipses: a quick-start guide and software', *arXiv preprint arXiv:0906.4123* .

Cole, S., Percival, W., Peacock, J., Norberg, P., Baugh, C., Frenk, C., Baldry, I., Bland-Hawthorn, J., Bridges, T., Cannon, R. et al. (2005), 'The 2dF Galaxy Redshift Survey: power-spectrum analysis of the final data set and cosmological implications', *Monthly Notices of the Royal Astronomical Society* **362**, 505–534.

Cooray, A. (2001), 'Applications of halo approach to non-linear large scale structure clustering', *arXiv preprint astro-ph/0105440* .

Cooray, A. and Hu, W. (2001), 'Power spectrum covariance of weak gravitational lensing', *The Astrophysical Journal* **554**(1), 56.

Cooray, A. and Sheth, R. (2002), 'Halo models of large scale structure', *Physics reports* **372**(1), 1–129.

Coulton, W. R., Liu, J., Madhavacheril, M. S., Böhm, V. and Spergel, D. N. (2019), 'Constraining neutrino mass with the tomographic weak lensing bispectrum', *Journal of Cosmology and Astroparticle Physics* **2019**(05), 043.

Cropper, M., Hoekstra, H., Kitching, T., Massey, R., Amiaux, J., Miller, L., Mellier, Y., Rhodes, J., Rowe, B., Pires, S. et al. (2013), 'Defining a weak lensing experiment in space', *Monthly Notices of the Royal Astronomical Society* **431**(4), 3103–3126.

de Jong, J. T., Kleijn, G. A. V., Kuijken, K. H., Valentijn, E. A. et al. (2013), 'The Kilo-Degree Survey', *Experimental Astronomy* **35**(1-2), 25–44.

Desjacques, V., Jeong, D. and Schmidt, F. (2018), 'Large-scale galaxy bias', *Physics reports* **733**, 1–193.

Di Dio, E. (2017), 'Lensing smoothing of BAO wiggles', *Journal of Cosmology and Astroparticle Physics* (03), 016.

Di Dio, E., Durrer, R., Marozzi, G. and Montanari, F. (2016), 'The bispectrum of relativistic galaxy number counts', *Journal of Cosmology and Astroparticle Physics* (1).

Dicke, R. H., Peebles, P. J. E., Roll, P. G. and Wilkinson, D. T. (1965), 'Cosmic black-body radiation.', *The Astrophysical Journal* **142**, 414–419.

Diemer, B. and Joyce, M. (2019), 'An accurate physical model for halo concentrations', *The Astrophysical Journal* **871**(2), 168.

Dodelson, S., Schmidt, F. and Vallinotto, A. (2008), 'Universal weak lensing distortion of cosmological correlation functions', *Physical Review D* **78**(4), 12.

Duffy, A. R., Schaye, J., Kay, S. T. and Dalla Vecchia, C. (2008), 'Dark matter halo concentrations in the Wilkinson Microwave Anisotropy Probe year 5 cosmology', *Monthly Notices of the Royal Astronomical Society: Letters* **390**(1), L64–L68.

Eddington, A. (1919), 'The deflection of light during a solar eclipse', *Nature* **104**(2615), 372.

Edge, A., Sutherland, W., Kuijken, K., Driver, S., McMahon, R., Eales, S. and Emerson, J. P. (2013), 'The VISTA Kilo-degree Infrared Galaxy (VIKING) Survey: Bridging the Gap between Low and High Redshift', *The Messenger* **154**, 32–34.

Eifler, T., Schneider, P. and Hartlap, J. (2009), 'Dependence of cosmic shear covariances on cosmology – impact on parameter estimation', *Astronomy & Astrophysics* **502**(3), 721–731.

Einstein, A. (1915*a*), 'Explanation of the perihelion motion of Mercury from the General Theory of Relativity', *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)* **1915**, 831–839.

Einstein, A. (1915*b*), 'On the General Theory of Relativity', *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)* **1915**, 778–786. [Addendum: Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)1915,799(1915)].

Einstein, A. (1915*c*), 'The Field Equations of Gravitation', *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)* **1915**, 844–847.

Eisenstein, D. J. and Hu, W. (1998), 'Baryonic features in the matter transfer function', *The Astrophysical Journal* **496**(2), 605.

Eisenstein, D. J., Hu, W. and Tegmark, M. (1998), 'Cosmic complementarity: H_0 and $\Omega_{\rm M}$ from combining cosmic microwave background experiments and redshift surveys', *The Astrophysical Journal Letters* **504**(2), L57.

Eisenstein, D. J., Seo, H.-J. and White, M. (2007), 'On the robustness of the acoustic scale in the low-redshift clustering of matter', *The Astrophysical Journal* **664**(2), 660.

Eisenstein, D. J., Zehavi, I., Hogg, D. W., Scoccimarro, R., Blanton, M. R., Nichol, R. C., Scranton, R., Seo, H.-J., Tegmark, M., Zheng, Z. et al. (2005), 'Detection of the baryon acoustic peak in the large-scale correlation function of SDSS luminous red galaxies', *The Astrophysical Journal* **633**(2), 560.

Feroz, F. and Hobson, M. (2008), 'Multimodal nested sampling: an efficient and robust alternative to Markov Chain Monte Carlo methods for astronomical data analyses', *Monthly Notices of the Royal Astronomical Society* **384**(2), 449–463.

Feroz, F., Hobson, M., Cameron, E. and Pettitt, A. (2013), 'Importance nested sampling and the MultiNest algorithm', *arXiv preprint arXiv:1306.2144* .

Fisher, R. A. (1935), *The design of experiments.*, Oliver & Boyd.

Ford, K. W. and Wheeler, J. A. (1998), 'Geons, black holes, and quantum foam: A life in physics', *New York* **644**.

Foreman, S., Coulton, W., Villaescusa-Navarro, F. and Barreira, A. (2019), 'Baryonic effects on the matter bispectrum', *arXiv preprint arXiv:1910.03597* .

Friedmann, A. (1922), 'Über die krümmung des raumes', *Zeitschrift fur Physik* **10**, 377–386.

Fry, J. N. (1984), 'The galaxy correlation hierarchy in perturbation theory', *The Astrophysical Journal* **279**, 499–510.

Fu, L., Kilbinger, M., Erben, T., Heymans, C., Hildebrandt, H., Hoekstra, H., Kitching, T. D., Mellier, Y., Miller, L., Semboloni, E. et al. (2014), 'CFHTLenS: cosmological constraints from a combination of cosmic shear two-point and three-point correlations', *Monthly Notices of the Royal Astronomical Society* **441**(3), 2725–2743.

Gamow, G. (1948), 'The evolution of the universe', *Nature* **162**(4122), 680.

Gaztanaga, E., Cabre, A., Castander, F., Crocce, M. and Fosalba, P. (2009), 'Clustering of luminous red galaxies - III. Baryon acoustic peak in the three-point correlation', *Monthly Notices of the Royal Astronomical Society* **399**(2), 801–811.

Gelman, A., Rubin, D. B. et al. (1992), 'Inference from iterative simulation using multiple sequences', *Statistical science* **7**(4), 457–472.

Gerbino, M., Massimiliano, L., Migliaccio, M., Pagano, L., Salvati, L., Colombo, L., Gruppuso, A., Natoli, P. and Polenta, G. (2019), 'Likelihood methods for CMB experiments', *arXiv preprint arXiv:1909.09375* .

Gil-Marín, H., Wagner, C., Fragkoudi, F., Jimenez, R. and Verde, L. (2012), 'An improved fitting formula for the dark matter bispectrum', *Journal of Cosmology and Astroparticle Physics* **2012**(02), 047.

Goroff, M., Grinstein, B., Rey, S.-J. and Wise, M. B. (1986), 'Coupling of modes of cosmological mass density fluctuations', *The Astrophysical Journal* **311**, 6–14.

Gualdi, D., Gil-Marín, H., Schuhmann, R. L., Manera, M., Joachimi, B. and Lahav, O. (2019), 'Enhancing BOSS bispectrum cosmological constraints with maximal compression', *Monthly Notices of the Royal Astronomical Society* **484**(3), 3713–3730.

Gualdi, D., Manera, M., Joachimi, B. and Lahav, O. (2018), 'Maximal compression of the redshift-space galaxy power spectrum and bispectrum', *Monthly Notices of the Royal Astronomical Society* **476**(3), 4045–4070.

Hahn, C., Beutler, F., Sinha, M., Berlind, A., Ho, S. and Hogg, D. W. (2019), 'Likelihood non-Gaussianity in large-scale structure analyses', *Monthly Notices of the Royal Astronomical Society* **485**(2), 2956–2969.

Hamilton, A. J., Rimes, C. D. and Scoccimarro, R. (2006), 'On measuring the covariance matrix of the non-linear power spectrum from simulations', *Monthly Notices of the Royal Astronomical Society* **371**(3), 1188–1204.

Hamilton, A. J. S. (2000), 'Uncorrelated modes of the non-linear power spectrum', *Monthly Notices of the Royal Astronomical Society* **312**(2), 257–284.

Hamimeche, S. and Lewis, A. (2008), 'Likelihood analysis of CMB temperature and polarization power spectra', *Physical Review D* **77**(10), 103013.

Hanson, D., Challinor, A. and Lewis, A. (2010), 'Weak lensing of the CMB', *General Relativity and Gravitation* **42**(9), 2197–2218.

Harnois-Déraps, J., Amon, A., Choi, A., Demchenko, V., Heymans, C., Kannawadi, A., Nakajima, R., Sirks, E., van Waerbeke, L., Cai, Y.-C. et al. (2018), 'Cosmological simulations for combined-probe analyses: covariance and neighbour-exclusion bias', *Monthly Notices of the Royal Astronomical Society* **481**(1), 1337–1367.

Harnois-Deraps, J., Giblin, B. and Joachimi, B. (2019), 'Cosmic shear covariance matrix in wCDM: Cosmology matters', *Astronomy & Astrophysics* **631**, A160.

Harnois-Déraps, J. and van Waerbeke, L. (2015), 'Simulations of weak gravitational lensing–II. Including finite support effects in cosmic shear covariance matrices', *Monthly Notices of the Royal Astronomical Society* **450**(3), 2857–2873.

Hartlap, J., Schrabback, T., Simon, P. and Schneider, P. (2009), 'The non-Gaussianity of the cosmic shear likelihood or how odd is the Chandra Deep Field South?', *Astronomy & Astrophysics* **504**(3), 689–703.

Hearin, A. P., Zentner, A. R. and Ma, Z. (2012), 'General requirements on matter power spectrum predictions for cosmology with weak lensing tomography', *Journal of Cosmology and Astroparticle Physics* **2012**(04), 034.

Heavens, A. (2003), '3D weak lensing', *Monthly Notices of the Royal Astronomical Society* **343**(4), 1327–1334.

Heavens, A. (2009), 'Statistical techniques in cosmology', *arXiv preprint arXiv:0906.0664* .

Heitmann, K., Higdon, D., Nakhleh, C. and Habib, S. (2006), 'Cosmic calibration', *The Astrophysical Journal Letters* **646**(1), L1.

Heitmann, K., White, M., Wagner, C., Habib, S. and Higdon, D. (2010), 'The coyote universe. I. Precision determination of the nonlinear matter power spectrum', *The Astrophysical Journal* **715**(1), 104.

Heymans, C., Van Waerbeke, L., Bacon, D., Berge, J., Bernstein, G., Bertin, E., Bridle, S., Brown, M. L., Clowe, D., Dahle, H. et al. (2006), 'The Shear Testing Programme–I. Weak lensing analysis of simulated ground-based observations', *Monthly Notices of the Royal Astronomical Society* **368**(3), 1323–1339.

Heymans, C., Van Waerbeke, L., Miller, L., Erben, T., Hildebrandt, H., Hoekstra, H., Kitching, T. D., Mellier, Y., Simon, P., Bonnett, C. et al. (2012), 'Cfhtlens: the canada–france–hawaii telescope lensing survey', *Monthly Notices of the Royal Astronomical Society* **427**(1), 146–166.

Hikage, C., Oguri, M., Hamana, T., More, S., Mandelbaum, R., Takada, M., Köhlinger, F., Miyatake, H., Nishizawa, A. J., Aihara, H. et al. (2019), 'Cosmology from cosmic shear power spectra with Subaru Hyper Suprime-Cam first-year data', *Publications of the Astronomical Society of Japan* **71**(2), 43.

Hilbert, S., Xu, D., Schneider, P., Springel, V., Vogelsberger, M. and Hernquist, L. (2017), 'Intrinsic alignments of galaxies in the Illustris simulation', *Monthly Notices of the Royal Astronomical Society* **468**(1), 790–823.

Hildebrandt, H., Köhlinger, F., van den Busch, J., Joachimi, B., Heymans, C., Kannawadi, A., Wright, A., Asgari, M., Blake, C., Hoekstra, H. et al. (2020), 'KiDS+ VIKING-450: Cosmic shear tomography with optical and infrared data', *Astronomy & Astrophysics* **633**, A69.

Hildebrandt, H., Viola, M., Heymans, C., Joudaki, S., Kuijken, K., Blake, C., Erben, T., Joachimi, B., Klaes, D., Miller, L. et al. (2016), 'KiDS-450: Cosmological parameter constraints from tomographic weak gravitational lensing', *Monthly Notices of the Royal Astronomical Society* **465**(2), 1454–1498.

Hinshaw, G., Nolta, M., Bennett, C., Bean, R., Doré, O., Greason, M., Halpern, M., Hill, R., Jarosik, N., Kogut, A. et al. (2007), 'Three-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Temperature analysis', *The Astrophysical Journal Supplement Series* **170**(2), 288.

Hirata, C. M. and Seljak, U. (2004), 'Intrinsic alignment-lensing interference as a contaminant of cosmic shear', *Physical Review D* **70**(6), 063526.

Hu, W. (1999), 'Power spectrum tomography with weak lensing', *The Astrophysical Journal Letters* **522**(1), L21.

Hu, W. (2000), 'Weak lensing of the CMB: A harmonic approach', *Physical Review D* **62**(4).

Hu, W. and White, M. (1996), 'Acoustic signatures in the cosmic microwave background', *The Astrophysical Journal* **471**(1), 30.

Hubble, E. (1929), 'A relation between distance and radial velocity among extra-galactic nebulae', *Proceedings of the National Academy of Sciences* **15**(3), 168–173.

Hubble, E. and Humason, M. L. (1931), 'The velocity-distance relation among extra-galactic nebulae', *The Astrophysical Journal* **74**, 43.

Huterer, D. and Takada, M. (2005), 'Calibrating the nonlinear matter power spectrum: Requirements for future weak lensing surveys', *Astroparticle Physics* **23**(4), 369–376.

Huterer, D., Takada, M., Bernstein, G. and Jain, B. (2006), 'Systematic errors in future weak-lensing surveys: requirements and prospects for self-calibration', *Monthly Notices of the Royal Astronomical Society* **366**(1), 101–114.

Jain, B. and Bertschinger, E. (1993), 'Second order power spectrum and nonlinear evolution at high redshift', *The Astrophysical Journal* **431**, 495.

Jarvis, M., Bernstein, G. and Jain, B. (2004), 'The skewness of the aperture mass statistic', *Monthly Notices of the Royal Astronomical Society* **352**(1), 338–352.

Jee, M. J., Tyson, J. A., Schneider, M. D., Wittman, D., Schmidt, S. and Hilbert, S. (2013), 'Cosmic shear results from the Deep Lens Survey. I. Joint constraints on Ω_M and σ_8 with a two-dimensional analysis', *The Astrophysical Journal* **765**(1), 74.

Jing, Y. P. and Borner, G. (1997), 'Three-point correlation function in the quasilinear regime', *Astronomy and Astrophysics* **318**(3), 667–672.

Joachimi, B. and Bridle, S. (2010), 'Simultaneous measurement of cosmology and intrinsic alignments using joint cosmic shear and galaxy number density correlations', *Astronomy & Astrophysics* **523**, A1.

Joachimi, B., Cacciato, M., Kitching, T. D., Leonard, A., Mandelbaum, R., Schäfer, B. M., Sifón, C., Hoekstra, H., Kiessling, A., Kirk, D. et al. (2015), 'Galaxy alignments: An overview', *Space Science Reviews* **193**(1-4), 1–65.

Joachimi, B., Mandelbaum, R., Abdalla, F. and Bridle, S. (2011), 'Constraints on intrinsic alignment contamination of weak lensing surveys using the MegaZ-LRG sample', *Astronomy & Astrophysics* **527**, A26.

Joachimi, B. and Schneider, P. (2008), 'The removal of shear-ellipticity correlations from the cosmic shear signal via nulling techniques', *Astronomy & Astrophysics* **488**(3), 829–843.

Joachimi, B., Shi, X. and Schneider, P. (2009), 'Bispectrum covariance in the flat-sky limit', *Astronomy & Astrophysics* **508**(3), 1193–1204.

Joachimi, B., Taylor, A. and Kiessling, A. (2011), 'Cosmological information in Gaussianized weak lensing signals', *Monthly Notices of the Royal Astronomical Society* **418**(1), 145–169.

Joudaki, S., Blake, C., Heymans, C., Choi, A., Harnois-Deraps, J., Hildebrandt, H., Joachimi, B., Johnson, A., Mead, A., Parkinson, D. et al. (2016), 'CFHTLenS revisited: assessing concordance with Planck including astrophysical systematics', *Monthly Notices of the Royal Astronomical Society* p. stw2665.

Kacprzak, T., Kirk, D., Friedrich, O., Amara, A., Refregier, A., Marian, L., Dietrich, J., Suchyta, E., Aleksić, J., Bacon, D. et al. (2016), 'Cosmology constraints from shear peak statistics in Dark Energy Survey Science Verification data', *Monthly Notices of the Royal Astronomical Society* **463**(4), 3653–3673.

Kaiser, N. (1984), 'On the spatial correlations of Abell clusters', *The Astrophysical Journal* **284**, L9–L12.

Kaiser, N. (1992), 'Weak gravitational lensing of distant galaxies', *The Astrophysical Journal* **388**, 272–286.

Kalus, B., Percival, W. J. and Samushia, L. (2015), 'Cosmological parameter inference from galaxy clustering: The effect of the posterior distribution of the power spectrum', *Monthly Notices of the Royal Astronomical Society* **455**(3), 2573–2581.

Kayo, I. and Takada, M. (2013), 'Cosmological parameters from weak lensing power spectrum and bispectrum tomography: including the non-Gaussian errors', *arXiv preprint arXiv:1306.4684* .

Kayo, I., Takada, M. and Jain, B. (2012), 'Information content of weak lensing power spectrum and bispectrum: including the non-Gaussian error covariance matrix', *Monthly Notices of the Royal Astronomical Society* **429**(1), 344–371.

Kehagias, A., Perrier, H. and Riotto, A. (2014), 'Equal-time consistency relations in the large-scale structure of the universe', *Modern Physics Letters A* **29**(29), 1450152.

Kendall, M. G. and Stuart, A. (1966), *The Advanced Theory of Statistics Volume 2, Inference and Relationship*, Charles Griffin & Company Ltd., London.

Kenney, F. and Keeping, E. (1951), *Mathematics of Statistics*, D. Van Nostrand Company, Inc Princeton,; New Jersey; Toronto; New York; London.

Kilbinger, M. (2015), 'Cosmology with cosmic shear observations: a review', *Reports on Progress in Physics* **78**(8), 086901.

Kilbinger, M., Fu, L., Heymans, C., Simpson, F., Benjamin, J., Erben, T., Harnois-Déraps, J., Hoekstra, H., Hildebrandt, H., Kitching, T. D. et al. (2013), 'CFHTLenS: combined probe cosmological model comparison using 2D weak gravitational lensing', *Monthly Notices of the Royal Astronomical Society* **430**(3), 2200–2220.

Kilbinger, M., Heymans, C., Asgari, M., Joudaki, S., Schneider, P., Simon, P., Van Waerbeke, L., Harnois-Déraps, J., Hildebrandt, H., Köhlinger, F. et al. (2017), 'Precision calculations of the cosmic shear power spectrum projection', *Monthly Notices of the Royal Astronomical Society* **472**(2), 2126–2141.

Kirk, D., Rassat, A., Host, O. and Bridle, S. (2012), 'The cosmological impact of intrinsic alignment model choice for cosmic shear', *Monthly Notices of the Royal Astronomical Society* **424**(3), 1647–1657.

Kitching, T. D., Alsing, J., Heavens, A. F., Jimenez, R., McEwen, J. D. and Verde, L. (2017), 'The limits of cosmic shear', *Monthly Notices of the Royal Astronomical Society* **469**(3), 2737–2749.

Knabenhans, M., Stadel, J., Marelli, S., Potter, D., Teyssier, R., Legrand, L., Schneider, A., Sudret, B., Blot, L. et al. (2019), 'Euclid preparation: II. The EuclidEmulator–A tool to compute the cosmology dependence of the nonlinear matter power spectrum', *Monthly Notices of the Royal Astronomical Society* **484**(4), 5509–5529.

Kodwani, D., Alsono, D. and Ferreira, P. (2019), 'The effect on cosmological parameter estimation of a parameter dependent covariance matrix', *The Open Journal of Astrophysics* .

Köhlinger, F., Viola, M., Joachimi, B., Hoekstra, H., Van Uitert, E., Hildebrandt, H., Choi, A., Erben, T., Heymans, C., Joudaki, S. et al. (2017), 'Kids-450: The tomographic weak lensing power spectrum and constraints on cosmological parameters', *Monthly Notices of the Royal Astronomical Society* **471**(4), 4412–4435.

Köhlinger, F., Viola, M., Valkenburg, W., Joachimi, B., Hoekstra, H. and Kuijken, K. (2015), 'A direct measurement of tomographic lensing power spectra from CFHTLenS', *Monthly Notices of the Royal Astronomical Society* **456**(2), 1508–1527.

Kratochvil, J. M., Haiman, Z. and May, M. (2010), 'Probing cosmology with weak lensing peak counts', *Physical Review D* **81**(4), 043519.

Kratochvil, J. M., Lim, E. A., Wang, S., Haiman, Z., May, M. and Huffenberger, K. (2012), 'Probing cosmology with weak lensing Minkowski functionals', *Physical Review D* **85**(10), 103513.

Krause, E., Eifler, T., Zuntz, J., Friedrich, O., Troxel, M., Dodelson, S., Blazek, J., Secco, L., MacCrann, N., Baxter, E. et al. (2017), 'Dark Energy Survey Year 1 Results: Multi-Probe Methodology and Simulated Likelihood Analyses', *arXiv preprint arXiv:1706.09359* .

Kuijken, K., Heymans, C., Hildebrandt, H., Nakajima, R., Erben, T., de Jong, J. T., Viola, M., Choi, A., Hoekstra, H., Miller, L. et al. (2015), 'Gravitational lensing analysis of the Kilo-Degree Survey', *Monthly Notices of the Royal Astronomical Society* **454**(4), 3500–3532.

Kwan, J., Heitmann, K., Habib, S., Padmanabhan, N., Lawrence, E., Finkel, H., Frontiere, N. and Pope, A. (2015), 'Cosmic emulation: fast predictions for the galaxy power spectrum', *The Astrophysical Journal* **810**(1), 35.

Labatie, A., Starck, J. and Lachieze-Rey, M. (2012), 'Effect of model-dependent covariance matrix for studying Baryon Acoustic Oscillations', *The Astrophysical Journal* **760**(2), 97.

Laplace, P. S. (1812), 'Théorie analytique des probabilités, 2 vols', *Paris: Courcier Imprimeur* .

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

Lawrence, E., Heitmann, K., White, M., Higdon, D., Wagner, C., Habib, S. and Williams, B. (2010), 'The coyote universe. III. simulation suite and precision emulator for the nonlinear matter power spectrum', *The Astrophysical Journal* **713**(2), 1322.

Lazanu, A., Giannantonio, T., Schmittfull, M. and Shellard, E. P. S. (2016), 'Matter bispectrum of large-scale structure: Three-dimensional comparison between theoretical models and numerical simulations', *Physical Review D* **93**(8), 49.

Leclercq, F. (2018), 'Bayesian optimization for likelihood-free cosmological inference', *Physical Review D* **98**(6), 063511.

Lemaître, G. (1927), Un univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques, *in* 'Annales de la Société scientifique de Bruxelles', Vol. 47, pp. 49–59.

Lewis, A. and Challinor, A. (2006), 'Weak gravitational lensing of the CMB', *Physics Reports-Review Section of Physics Letters* **429**(1), 1–65.

Li, Y., Hu, W. and Takada, M. (2014*a*), 'Super-sample covariance in simulations', *Physical Review D* **89**(8), 083519.

Li, Y., Hu, W. and Takada, M. (2014*b*), 'Super-sample signal', *Physical Review D* **90**(10), 103530.

Limber, D. N. (1953), 'The analysis of counts of the extragalactic nebulae in terms of a fluctuating density field.', *The Astrophysical Journal* **117**, 134.

Lin, C.-H., Harnois-Déraps, J., Eifler, T., Pospisil, T., Mandelbaum, R., Lee, A. B. and Singh, S. (2019), 'Non-Gaussianity in the Weak Lensing Correlation Function Likelihood – Implications for Cosmological Parameter Biases', *arXiv preprint arXiv:1905.03779* .

Liu, J., Petri, A., Haiman, Z., Hui, L., Kratochvil, J. M. and May, M. (2015), 'Cosmology constraints from the weak lensing peak counts and the power spectrum in CFHTLenS data', *Physical Review D* **91**(6), 063507.

Loureiro, A., Moraes, B., Abdalla, F. B., Cuceu, A., McLeod, M., Whiteway, L., Balan, S. T., Benoit-Lévy, A., Lahav, O., Manera, M. et al. (2019), 'Cosmological measurements from angular power spectra analysis of BOSS DR12 tomography', *Monthly Notices of the Royal Astronomical Society* **485**(1), 326–355.

LoVerde, M. and Afshordi, N. (2008), 'Extended Limber approximation', *Physical Review D* **78**(12), 123506.

Ma, Z., Hu, W. and Huterer, D. (2006), 'Effects of photometric redshift uncertainties on weak-lensing tomography', *The Astrophysical Journal* **636**(1), 21.

Mandelbaum, R. (2018), 'Weak lensing for precision cosmology', *Annual Review of Astronomy and Astrophysics* **56**, 393–433.

Mangilli, A., Plaszczynski, S. and Tristram, M. (2015), 'Large-scale cosmic microwave background temperature and polarization cross-spectra likelihoods', *Monthly Notices of the Royal Astronomical Society* **453**(3), 3174–3189.

Martinet, N., Schneider, P., Hildebrandt, H., Shan, H., Asgari, M., Dietrich, J. P., Harnois-Déraps, J., Erben, T., Grado, A., Heymans, C. et al. (2017), 'KiDS-450: cosmological constraints from weak-lensing peak statistics–II: Inference from shear peaks using N-body simulations', *Monthly Notices of the Royal Astronomical Society* **474**(1), 712–730.

Massey, R., Hoekstra, H., Kitching, T., Rhodes, J., Cropper, M., Amiaux, J., Harvey, D., Mellier, Y., Meneghetti, M., Miller, L. et al. (2012), 'Origins of weak lensing systematics, and requirements on future instrumentation (or knowledge of instrumentation)', *Monthly Notices of the Royal Astronomical Society* **429**(1), 661–678.

McClintock, T. and Rozo, E. (2019), 'Reconstructing probability distributions with Gaussian processes', *Monthly Notices of the Royal Astronomical Society* **489**(3), 4155–4160.

McKellar, A. (1941), 'Molecular lines from the lowest states of diatomic molecules composed of atoms probably present in interstellar space', *Publications of the Dominion Astrophysical Observatory Victoria* **7**.

Mead, A., Peacock, J., Heymans, C., Joudaki, S. and Heavens, A. (2015), 'An accurate halo model for fitting the non-linear cosmological power spectra and baryonic feedback models', *Monthly Notices of the Royal Astronomical Society* **454**(2), 1958–1975.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953), 'Equation of state calculations by fast computing machines', *The Journal of Chemical Physics* **21**(6), 1087–1092.

Mo, H. and White, S. D. (1996), 'An analytic model for the spatial clustering of dark matter haloes', *Monthly Notices of the Royal Astronomical Society* **282**(2), 347–361.

Morrison, C. B. and Schneider, M. D. (2013), 'On estimating cosmology-dependent covariance matrices', *Journal of Cosmology and Astroparticle Physics* **2013**(11), 009.

Mukhanov, V. F., Feldman, H. A. and Brandenberger, R. H. (1992), 'Theory of cosmological perturbations', *Physics Reports* **215**(5-6), 203–333.

Mukherjee, P., Parkinson, D. and Liddle, A. R. (2006), 'A nested sampling algorithm for cosmological model selection', *The Astrophysical Journal Letters* **638**(2), L51.

Munshi, D. and Coles, P. (2017), 'The integrated bispectrum and beyond', *Journal of Cosmology and Astroparticle Physics* **2017**(02), 010.

Namikawa, T. (2016), 'Cmb lensing bispectrum from nonlinear growth of the large scale structure', *Physical Review D* **93**(12), 121301.

Navarro, J. F., Frenk, C. S. and White, S. D. (1997), 'A universal density profile from hierarchical clustering', *The Astrophysical Journal* **490**(2), 493.

Navarro, J. F., Frenk, C. S. and White, S. D. M. (1996), 'The Structure of Cold Dark Matter Halos', *The Astrophysical Journal* **462**, 563.

Okoli, C. (2017), 'Dark matter halo concentrations: a short review', *arXiv preprint arXiv:1711.05277* .

O'Raifeartaigh, C. (2013), 'The contribution of VM Slipher to the discovery of the expanding universe'.

O'Raifeartaigh, C., O'Keeffe, M., Nahm, W. and Mitton, S. (2017), 'Einstein's 1917 static model of the universe: a centennial review', *The European Physical Journal H* **42**(3), 431–474.

O'Raifeartaigh, Cormac and Mitton, Simon (2018), 'Interrogating the Legend of Einstein's "Biggest Blunder"', *Physics in Perspective* **20**(4), 318–341.

Peacock, J. and Dodds, S. (1996), 'Non-linear evolution of cosmological power spectra', *Monthly Notices of the Royal Astronomical Society* **280**(3), L19–L26.

Peebles, P. J. and Yu, J. (1970), 'Primeval adiabatic perturbation in an expanding universe', *The Astrophysical Journal* **162**, 815.

Pellejero-Ibañez, M., Angulo, R. E., Aricó, G., Zennaro, M., Contreras, S. and Stücker, J. (2019), 'Cosmological parameter estimation via iterative emulation of likelihoods', *arXiv preprint arXiv:1912.08806* .

Pen, U.-L., Zhang, T., Van Waerbeke, L., Mellier, Y., Zhang, P. and Dubinski, J. (2003), 'Detection of dark matter skewness in the VIRMOS-DESCART survey: implications for ω_0', *The Astrophysical Journal* **592**(2), 664.

Penzias, A. A. and Wilson, R. W. (1965), 'A measurement of excess antenna temperature at 4080 mc/s.', *The Astrophysical Journal* **142**, 419–421.

Percival, W. J. and Brown, M. L. (2006), 'Likelihood techniques for the combined analysis of CMB temperature and polarization power spectra', *Monthly Notices of the Royal Astronomical Society* **372**(3), 1104–1116.

Perlmutter, S., Aldering, G., Goldhaber, G., Knop, R., Nugent, P., Castro, P., Deustua, S., Fabbro, S., Goobar, A., Groom, D. et al. (1999), 'Measurements of ω and λ from 42 high-redshift supernovae', *The Astrophysical Journal* **517**(2), 565.

Petri, A., Haiman, Z., Hui, L., May, M. and Kratochvil, J. M. (2013), 'Cosmology with Minkowski functionals and moments of the weak lensing convergence field', *Physical Review D* **88**(12), 123002.

Pielorz, J., Rödiger, J., Tereno, I. and Schneider, P. (2010), 'A fitting formula for the non-Gaussian contribution to the lensing power spectrum covariance', *Astronomy & Astrophysics* **514**, A79.

Press, W. H. and Schechter, P. (1974), 'Formation of galaxies and clusters of galaxies by self-similar gravitational condensation', *The Astrophysical Journal* **187**, 425–438.

Rees, M. J. and Sciama, D. W. (1968), 'Large-scale density inhomogeneities in the universe', *Nature* **217**(5128), 511.

Reischke, R., Kiessling, A. and Schäfer, B. M. (2016), 'Variations of cosmic large-scale structure covariance matrices across parameter space', *Monthly Notices of the Royal Astronomical Society* **465**(4), 4016–4025.

Riess, A. G., Filippenko, A. V., Challis, P., Clocchiatti, A., Diercks, A., Garnavich, P. M., Gilliland, R. L., Hogan, C. J., Jha, S., Kirshner, R. P. et al. (1998), 'Observational evidence from supernovae for an accelerating universe and a cosmological constant', *The Astronomical Journal* **116**(3), 1009.

Rimes, C. D. and Hamilton, A. J. (2006), 'Information content of the non-linear power spectrum: the effect of beat-coupling to large scales', *Monthly Notices of the Royal Astronomical Society* **371**(3), 1205–1215.

Rizzato, M., Benabed, K., Bernardeau, F. and Lacasa, F. (2019), 'Tomographic weak lensing bispectrum: a thorough analysis towards the next generation of galaxy surveys', *Monthly Notices of the Royal Astronomical Society* **490**(4), 4688–4714.

Robertson, H. P. (1935), 'Kinematics and world-structure', *The Astrophysical Journal* **82**, 284.

Rogers, K. K., Peiris, H. V., Pontzen, A., Bird, S., Verde, L. and Font-Ribera, A. (2019), 'Bayesian emulator optimisation for cosmology: application to the Lyman-alpha forest', *Journal of Cosmology and Astroparticle Physics* **2019**(02), 031.

Ross, A. J., Beutler, F., Chuang, C.-H., Pellejero-Ibanez, M., Seo, H.-J., Vargas-Magaña, M., Cuesta, A. J., Percival, W. J., Burden, A., Sánchez, A. G. et al. (2016), 'The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: Observational systematics and baryon acoustic oscillations in the correlation function', *Monthly Notices of the Royal Astronomical Society* **464**(1), 1168–1191.

Sachs, R. and Wolfe, A. (1967), 'Perturbations of a cosmological model and angular variations of the microwave background', *The Astrophysical Journal* **147**, 73.

Samtleben, D., Staggs, S. and Winstein, B. (2007), 'The cosmic microwave background for pedestrians: a review for particle and nuclear physicists', *Annual Review of Nuclear and Particle Science* **57**, 245–283.

Sato, M. and Nishimichi, T. (2013), 'Impact of the non-Gaussian covariance of the weak lensing power spectrum and bispectrum on cosmological parameter estimation', *Physical Review D* **87**(12), 123538.

Schaan, E., Takada, M. and Spergel, D. N. (2014), 'Joint likelihood function of cluster counts and n-point correlation functions: Improving their power through including halo sample variance', *Physical Review D* **90**(12), 123523.

Schmidt, F., Jeong, D. and Desjacques, V. (2013), 'Peak-background split, renormalization, and galaxy clustering', *Physical Review D* **88**(2), 023515.

Schmidt, F., Vallinotto, A., Sefusatti, E. and Dodelson, S. (2008), 'Weak lensing effects on the galaxy three-point correlation function', *Physical Review D* **78**(4), 21.

Schneider, M. D. and Bridle, S. (2010), 'A halo model for intrinsic alignments of galaxy ellipticities', *Monthly Notices of the Royal Astronomical Society* **402**(4), 2127–2139.

Schneider, M. D., Holm, O. and Knox, L. (2011), 'Intelligent design: on the emulation of cosmological simulations', *The Astrophysical Journal* **728**(2), 137.

Schneider, P. (1996), 'Detection of (dark) matter concentrations via weak gravitational lensing', *Monthly Notices of the Royal Astronomical Society* **283**(3), 837–853.

Schneider, P. (2006), Weak gravitational lensing, *in* 'Gravitational Lensing: Strong, Weak and Micro', Springer, pp. 269–451.

Schneider, P., Eifler, T. and Krause, E. (2010), 'COSEBIs: Extracting the full E-/B-mode information from cosmic shear correlation functions', *Astronomy & Astrophysics* **520**, A116.

Schulz, E., Speekenbrink, M. and Krause, A. (2018), 'A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions', *Journal of Mathematical Psychology* **85**, 1–16.

Schwarzschild, K. (1916), Über das gravitationsfeld einer kugel aus inkompressibler flüssigkeit nach der einsteinschen theorie, *in* 'Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften zu Berlin, Phys.-Math. Klasse, 424-434 (1916)'.

Scoccimarro, R., Colombi, S., Fry, J. N., Frieman, J. A., Hivon, E. and Melott, A. (1998), 'Nonlinear evolution of the bispectrum of cosmological perturbations', *The Astrophysical Journal* **496**(2), 586.

Scoccimarro, R., Sefusatti, E. and Zaldarriaga, M. (2004), 'Probing primordial non-gaussianity with large-scale structure', *Physical Review D* **69**(10), 103513.

Scoccimarro, R., Zaldarriaga, M. and Hui, L. (1999), 'Power spectrum correlations induced by nonlinear clustering', *The Astrophysical Journal* **527**(1), 1.

Sefusatti, E., Crocce, M., Pueblas, S. and Scoccimarro, R. (2006), 'Cosmology and the bispectrum', *Physical Review D* **74**(2), 023522.

Sefusatti, E. and Scoccimarro, R. N. (2005), 'Galaxy bias and halo-occupation numbers from large-scale clustering', *Physical Review D* **71**(6), 13.

Seljak, U. (1996), 'Gravitational lensing effect on cosmic microwave background anisotropies: A power spectrum approach', *Astrophysical Journal* **463**, 1.

Sellentin, E. and Heavens, A. F. (2017), 'On the insufficiency of arbitrarily precise covariance matrices: non-Gaussian weak-lensing likelihoods', *Monthly Notices of the Royal Astronomical Society* **473**(2), 2355–2363.

Sellentin, E., Heymans, C. and Harnois-Déraps, J. (2018), 'The skewed weak lensing likelihood: why biases arise, despite data and theory being sound', *Monthly Notices of the Royal Astronomical Society* **477**(4), 4879–4895.

Semboloni, E., Heymans, C., Van Waerbeke, L. and Schneider, P. (2008), 'Sources of contamination to weak lensing three-point statistics: constraints from n-body simulations', *Monthly Notices of the Royal Astronomical Society* **388**(3), 991–1000.

Semboloni, E., Schrabback, T., van Waerbeke, L., Vafaei, S., Hartlap, J. and Hilbert, S. (2010), 'Weak lensing from space: first cosmological constraints from three-point shear statistics', *Monthly Notices of the Royal Astronomical Society* **410**(1), 143–160.

Seo, H. and Eisenstein, D. J. (2003), 'Probing dark energy with baryonic acoustic oscillations from future large galaxy redshift surveys', *The Astrophysical Journal* **598**(2), 720–740.

Sheth, R. K. and Tormen, G. (1999), 'Large-scale bias and the peak background split', *Monthly Notices of the Royal Astronomical Society* **308**(1), 119–126.

Shi, X., Joachimi, B. and Schneider, P. (2010), 'Controlling intrinsic-shear alignment in three-point weak lensing statistics', *Astronomy & Astrophysics* **523**, A60.

Shi, X., Joachimi, B. and Schneider, P. (2014), 'How well do third-order aperture mass statistics separate E-and B-modes?', *Astronomy & Astrophysics* **561**, A68.

Sievers, J., Bond, J., Cartwright, J., Contaldi, C., Mason, B., Myers, S., Padin, S., Pearson, T. J., Pen, U.-L., Pogosyan, D. et al. (2003), 'Cosmological parameters from Cosmic Background Imager observations and comparisons with BOOMERANG, DASI, and MAXIMA', *The Astrophysical Journal* **591**(2), 599.

Silk, J. (1968), 'Cosmic black-body radiation and galaxy formation', *The Astrophysical Journal* **151**, 459.

Simon, P. (2005), *Weak gravitational lensing and galaxy bias*, Rheinische Friedrich-Wilhelms Universitaet Bonn, Germany.

Skilling, J. et al. (2006), 'Nested sampling for general bayesian computation', *Bayesian analysis* **1**(4), 833–859.

Slepian, Z. and Eisenstein, D. J. (2015), 'Computing the three-point correlation function of galaxies in $\mathcal{O}(n^2)$ time', *Monthly Notices of the Royal Astronomical Society* **454**(4), 4142.

Slepian, Z. et al. (2017*a*), 'Detection of baryon acoustic oscillation features in the large-scale three-point correlation function of SDSS BOSS DR12 CMASS galaxies', *Monthly Notices of the Royal Astronomical Society* **469**(2), 1738–1751.

Slepian, Z. et al. (2017*b*), 'The large-scale three-point correlation function of the SDSS BOSS DR12 CMASS galaxies', *Monthly Notices of the Royal Astronomical Society* **468**(1), 1070–1083.

Slipher, V. M. (1915), 'Spectrographic observations of nebulae', *Popular astronomy* **23**, 21–24.

Smith, R. E., Peacock, J. A., Jenkins, A., White, S., Frenk, C., Pearce, F., Thomas, P. A., Efstathiou, G. and Couchman, H. (2003), 'Stable clustering, the halo model and non-linear cosmological power spectra', *Monthly Notices of the Royal Astronomical Society* **341**(4), 1311–1332.

Smith, S., Challinor, A. and Rocha, G. (2006), 'What can be learned from the lensed cosmic microwave background B-mode polarization power spectrum?', *Physical Review D* **73**(2), 023517.

Staggs, S., Dunkley, J. and Page, L. (2018), 'Recent discoveries from the cosmic microwave background: a review of recent progress', *Reports on Progress in Physics* **81**(4), 044901.

Sun, L., Wang, Q. and Zhan, H. (2013), 'Likelihood of the power spectrum in cosmological parameter estimation', *The Astrophysical Journal* **777**(1), 75.

Sunyaev, R. and Zel'dovich, Y. B. (1972), 'The observations of relic radiation as a test of the nature of X-ray radiation from the clusters of galaxies', *Comments on Astrophysics and Space Physics* **4**, 173.

Takada, M. and Bridle, S. (2007), 'Probing dark energy with cluster counts and cosmic shear power spectra: including the full covariance', *New Journal of Physics* **9**(12), 446.

Takada, M. and Hu, W. (2013), 'Power spectrum super-sample covariance', *Physical Review D* **87**(12), 123504.

Takada, M. and Jain, B. (2004), 'Cosmological parameters from lensing power spectrum and bispectrum tomography', *Monthly Notices of the Royal Astronomical Society* **348**(3), 897–915.

Takada, M. and Jain, B. (2009), 'The impact of non-Gaussian errors on weak lensing surveys', *Monthly Notices of the Royal Astronomical Society* **395**(4), 2065–2086.

Takahashi, R., Nishimichi, T., Namikawa, T., Taruya, A., Kayo, I., Osato, K., Kobayashi, Y. and Shirasaki, M. (2019), 'Bihalofit: A new fitting formula of non-linear matter bispectrum', *arXiv preprint arXiv:1911.07886* .

Takahashi, R., Sato, M., Nishimichi, T., Taruya, A. and Oguri, M. (2012), 'Revising the halofit model for the nonlinear matter power spectrum', *The Astrophysical Journal* **761**(2), 152.

Tansella, V., Bonvin, C., Durrer, R., Ghosh, B. and Sellentin, E. (2018), 'The full-sky relativistic correlation function and power spectrum of galaxy number counts. Part I: theoretical aspects', *Journal of Cosmology and Astroparticle Physics* **2018**(03), 019.

Tasinato, G., Tellarini, M., Ross, A. J. and Wands, D. (2014), 'Primordial non-Gaussianity in the bispectra of large-scale structure', *Journal of Cosmology and Astroparticle Physics* (3), 37.

Taylor, A., Joachimi, B. and Kitching, T. (2013), 'Putting the precision in precision cosmology: How accurate should your data covariance matrix be?', *Monthly Notices of the Royal Astronomical Society* **432**(3), 1928–1946.

Taylor, P. L., Kitching, T. D., Alsing, J., Wandelt, B. D., Feeney, S. M. and McEwen, J. D. (2019), 'Cosmic shear: Inference from forward models', *Physical Review D* **100**(2), 023519.

Tegmark, M., Taylor, A. N. and Heavens, A. F. (1997), 'Karhunen-Loeve eigenvalue problems in cosmology: how should we tackle large data sets?', *The Astrophysical Journal* **480**(1), 22.

Tenneti, A., Mandelbaum, R. and Di Matteo, T. (2016), 'Intrinsic alignments of disc and elliptical galaxies in the MassiveBlack-II and Illustris simulations', *Monthly Notices of the Royal Astronomical Society* **462**(3), 2668–2680.

Tinker, J., Kravtsov, A. V., Klypin, A., Abazajian, K., Warren, M., Yepes, G., Gottlöber, S. and Holz, D. E. (2008), 'Toward a halo mass function for precision cosmology: the limits of universality', *The Astrophysical Journal* **688**(2), 709.

Trotta, R. (2017), 'Bayesian methods in cosmology', *arXiv preprint arXiv:1701.01467* .

Troxel, M. A. and Ishak, M. (2011), 'Self-calibration technique for three-point intrinsic alignment correlations in weak lensing surveys', *Monthly Notices of the Royal Astronomical Society* **419**(2), 1804–1823.

Troxel, M. A. and Ishak, M. (2012), 'Self-calibration for three-point intrinsic alignment autocorrelations in weak lensing surveys', *Monthly Notices of the Royal Astronomical Society* **423**(2), 1663–1673.

Troxel, M. A., MacCrann, N., Zuntz, J., Eifler, T., Krause, E., Dodelson, S., Gruen, D., Blazek, J., Friedrich, O., Samuroff, S. et al. (2018), 'Dark Energy Survey Year 1 results: Cosmological constraints from cosmic shear', *Physical Review D* **98**(4), 043528.

Troxel, M. and Ishak, M. (2015), 'The intrinsic alignment of galaxies and its impact on weak gravitational lensing in an era of precision cosmology', *Physics Reports* **558**, 1–59.

Valageas, P. (2014), 'Angular-averaged consistency relations of large-scale structures', *Physical Review D* **89**(12), 123522.

Vallinotto, A., Dodelson, S., Schimd, C. and Uzan, J. P. (2007), 'Weak lensing of baryon acoustic oscillations', *Physical Review D* **75**(10), 10.

van Uitert, E., Joachimi, B., Joudaki, S., Amon, A., Heymans, C., Köhlinger, F., Asgari, M., Blake, C., Choi, A., Erben, T. et al. (2018), 'KiDS+ GAMA: cosmology constraints from a joint analysis of cosmic shear, galaxy–galaxy lensing, and angular clustering', *Monthly Notices of the Royal Astronomical Society* **476**(4), 4662–4689.

Verde, L., Peiris, H., Spergel, D. N., Nolta, M., Bennett, C., Halpern, M., Hinshaw, G., Jarosik, N., Kogut, A., Limon, M. et al. (2003), 'First-Year Wilkinson Microwave Anisotropy Probe (WMAP)* Observations: Parameter Estimation Methodology', *The Astrophysical Journal Supplement Series* **148**(1), 195.

Verde, L., Treu, T. and Riess, A. G. (2019), 'Tensions between the early and late Universe', *arXiv preprint arXiv:1907.10625* .

Viola, M., Cacciato, M., Brouwer, M., Kuijken, K., Hoekstra, H., Norberg, P., Robotham, A. S. G., Van Uitert, E., Alpaslan, M., Baldry, I. et al. (2015), 'Dark matter halo properties of GAMA galaxy groups from 100 square degrees of KiDS weak lensing data', *Monthly Notices of the Royal Astronomical Society* **452**(4), 3529–3550.

Vlah, Z., Chisari, N. E. and Schmidt, F. (2020), 'An EFT description of galaxy intrinsic alignments', *Journal of Cosmology and Astroparticle Physics* **2020**(01), 025.

von Soldner, J. (1804), 'On the deflection of a light ray from its rectilinear motion', *Berliner Astronomisches Jahrbuch* pp. 161–172.

Wagner, C., Schmidt, F., Chiang, C.-T. and Komatsu, E. (2015), 'The angle-averaged squeezed limit of nonlinear matter n-point functions', *Journal of Cosmology and Astroparticle Physics* **2015**(08), 042.

Walker, A. G. (1937), 'On Milne's Theory of World-Structure', *Proceedings of the London Mathematical Society* **2**(1), 90–127.

Walsh, D., Carswell, R. F. and Weymann, R. J. (1979), '0957+ 561 A, B: twin quasistellar objects or gravitational lens?', *Nature* **279**(5712), 381.

White, M. and Padmanabhan, N. (2015), 'Including parameter dependence in the data and covariance for cosmological inference', *Journal of Cosmology and Astroparticle Physics* **2015**(12), 058.

Wick, G.-C. (1950), 'The evaluation of the collision matrix', *Physical review* **80**(2), 268.

Yoo, J. and Zaldarriaga, M. (2014), 'Beyond the linear-order relativistic effect in galaxy clustering: Second-order gauge-invariant formalism', *Physical Review D* **90**(2).

Zhang, P. (2010), 'A proposal on the galaxy intrinsic alignment self-calibration in weak lensing surveys', *Monthly Notices of the Royal Astronomical Society: Letters* **406**(1), L95–L99.

Zuntz, J., Paterno, M., Jennings, E., Rudd, D., Manzotti, A., Dodelson, S., Bridle, S., Sehrish, S. and Kowalkowski, J. (2015), 'Cosmosis: Modular cosmological parameter estimation', *Astronomy and Computing* **12**, 45–59.