COSMOLOGY WITH DARK MATTER MAPS

Niall Jeffrey

Submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of

University College London.

Department of Physics & Astronomy

University College London

September 2019

I, Niall Jeffrey, 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 work.

Abstract

Physics is experiencing an exciting period of exploration into the nature of dark energy, dark matter, and gravitation. With 95% of the mass-energy of the Universe still unexplained, the answers to many further fundamental questions of astro-, theoretical- and particle-physics are being hampered. In the coming years, DES, HSC, KiDS, Euclid and LSST will image billions of galaxies, aiming to use observational data from the late Universe to infer cosmological parameters and compare cosmological models.

One of the most promising observables is the weak gravitational lensing effect. Using the statistical power from many small distortions, called *shear*, DES has provided excellent constraints. However, the standard 2-point statistics do not capture the full information in the data. In the late Universe, gravitational collapse has led to a highly non-Gaussian density field, for which 2-point correlations are not a unique statistical description, and even all N-point functions cannot completely characterize. The research presented in this thesis focuses on methods to reconstruct mass maps from DES weak lensing data and using map-based statistics to infer cosmological parameters and assess theoretical models in a principled Bayesian framework.

In Chapter 2, I compare three mass mapping methods with closed-form priors using DES SV data and simulations. In Chapter 3, I demonstrate how the Wiener filter (one of the above methods) computation can be sped up by an order of magnitude using Dataflow Engines (reconfigurable hardware). In Chapter 4, I present a Bayesian hierarchical model which takes into account added uncertainty introduced when noisy simulations are used to generate theoretical predictions. In Chapter 5, with my publicly available DeepMass code, I demonstrate how mass maps reconstructions can be improved ($> 10\%$ mean-square-error compared with previously presented methods) using deep learning techniques trained on simulations. In Chapter 6, I discuss future work and the applicability of likelihood-free inference methods for map-based statistics.

Impact Statement

The work in this thesis provides novel techniques to tackle fundamental problems in our understanding of cosmology. In the next decades, Euclid and LSST will observe billions of galaxies across more than 20,000 sq deg of the sky, which will provide unprecedented data challenges and opportunities. The research in this thesis combines theory and data analysis methods to reconstruct and use dark matter maps from weak gravitational lensing to better constrain cosmology.

This work, particularly Chapter 2, has directly contributed to the science exploitation of Dark Energy Survey (DES) data. DES is an international collaboration, which is in co-funded by the UK Science and Technology Facilities Council (STFC), that aims to constrain the cosmological model and provide understanding of dark energy using observational data. This galaxy survey covers the largest sky footprint of any currently active weak lensing survey. The work here will also be of use within the Large Synoptic Survey Telescope (LSST) and future European Space Agency (ESA) Euclid mission (at whose collaboration meetings this work has been discussed).

In Chapter 3, the work was carried out with industry partners, Maxeler Technologies, who provided innovative non-standard hardware and expertise. The resulting code has been made free and public, and the interaction with these partners has led to better insight for hardware development with different scientific use-cases.

The statistical results of Chapter 4, which demonstrate how to take into account theoretical predictions from noisy simulations, have since been cited and implemented by independent groups in the cosmology community working with different datasets.

In Chapter 5, I present the first reconstruction of a dark matter map from weak lensing observational data using deep learning. The code, DeepMass, has been made freely available. This mass mapping technique has been shown to outperform existing methods, and is therefore a leading candidate for methods to be used by the next generation of experiments.

The work of Chapters 2 and 4 were part of the European Union (EU) project DEDALE - Data Learning on Manifolds and Future Challenges. This project brought together industrial partners, astrophysicists, computer scientists and mathematicians to develop the next generation of data analysis methods for the new era of big data in astrophysics and compressed sensing. The work carried out in Chapter 2, was described by the European Commission's report:

"The EU-funded DEDALE project's notable contribution to the Dark Energy Survey, an international collaboration to investigate the nature of dark matter and dark energy, is one of several achievements demonstrating far-reaching applications for novel techniques to efficiently analyse vast volumes of data."

Acknowledgements

To begin, I would like to thank my supervisors, Ofer Lahav and Filipe Abdalla, who have guided me along the winding and branching path of this PhD. I imagine it is rare for a cosmology student, whose supervisor is not Ofer, to benefit from surprise meetings and discussions with passing philosophers.

I will also thank Jean-Luc Starck and Alan Heavens, with whom I have worked separately on research included in this thesis, and who have both provided wise counsel when it was sought. Scientifically, but closer to home, I would also like to thank Will Hartley and Lorne Whiteway, whose time was often generously given.

I would like to thank the many students and postdocs in the astrophysics group who have made UCL such an inviting and friendly place. In particular, my friends who began their PhDs simultaneously to mine have provided much in the way of entertainment and (necessary) biscuits throughout my time here. Similarly, my housemates of these past few years have helped make my PhD life a happy one, for which I am thankful.

I should acknowledge my family, who have been a steadfast source of encouragement and who, in times past, endured my many pyrophile kitchen-based “experiments”. Finally, I thank Clara, who has, in so many ways, made this immeasurably easier.

Contents

List of Figures

List of Tables

Chapter 1

Introduction and background

...that celebrated saying of Archytas of Tarentum, I think it was – a saying which I have heard repeated by our old men who in their turn heard it from their elders. It is to this effect:

"If a man should ascend alone into heaven and behold clearly the structure of the Universe and the beauty of the stars, there would be no pleasure for him in the awe-inspiring sight, which would have filled him with delight if he had had someone to whom he could describe what he had seen."

Cicero, Laelius de Amicitia (translation 1923)

The Heavens Reflect Our Labours

Motto of Scunthorpe

1.1 Cosmology and gravity

Gravity is the force of nature most clearly at work in the physical systems studied in this thesis. This is the same in much of modern cosmology. On cosmological scales, the interplay between gravitation[1] and the matter distribution (of which dark matter is likely the largest fraction) controls the expansion of the Universe, the formation of large-scale structure, and the apparently curved trajectories of particles and light.

As far as has been currently measured, dark matter interacts only through the gravitational force. Though the word "dark" has become the chosen adjective, its lack of interaction with electromagnetism would lead to "invisible" being a more ordinary description. Direct detection experiments in high energy physics collaborations have yet to show signs of dark matter interacting with known particles through the electroweak or strong interactions as understood in the standard model of particle physics.

Due to its importance to the content of this thesis, I shall begin by introducing the key concepts of gravitation as described by General Relativity. Given the work in this thesis, it is a coincidence that the year in which it was begun, 2015, is precisely 100 years since the publication of both the finalised Einstein field equations (Einstein, 1915a) and the derivation of the bending of light (lensing) in General Relativity (Einstein, 1915b). The year of submission, 2019, also marks 100 years since the Eddington Experiment measurement of this lensing effect (Dyson et al., 1920).

This four year gap between Einstein's conjecture and Eddington's observation seems remarkably short by modern standards. Another weak-field gravitational effect, the conjectured gravitational waves (Einstein, 1916) had to wait much longer for their direct detection from Earth (Abbott et al., 2016), which resulted in the 2017 Nobel Prize.

1.1.1 Homogeneity and isotropy

At the start of this introduction I described gravitation as a force. Famously in General Relativity (GR) we need not, or should not, think of gravity as a force, but rather as variations in the metric of a spacetime manifold. Newton's pull of gravity becomes Einstein's free fall in a curved spacetime.

The Newtonian concept that the inertial mass of $\mathbf{F} = m\mathbf{a}$ is the same as the gravitational mass suggests that there are trajectories of free fall, known as *inertial*, which are followed

[1] including "dark energy" or cosmological constant

by all particles not accelerated by an external[2] force. This motivates the *Weak Equivalence Principle*, which Carroll (2003) formulates as: "*the motion of freely-falling particles are the same in a gravitational field and a uniformly accelerated frame*". This leads us towards what is sometimes known as the *Einstein Equivalence Principle*, which states that locally (in small enough spacetime regions) physics reduces to that of special relativity.

It is impossible to use global inertial frames in this framework. On Earth, an inertial frame for a test particle A falling at the north pole will not be inertial for a test particle B falling at the south pole, in which the effect of gravity would appear like an external force acting on B. However, a test particle A' that is close enough in time and space to A, and is therefore sharing the same trajectory, can be thought of as sharing a *locally inertial frame*.

This all necessitates a geometry that has curvature but is approximately flat locally, for which mathematics provides the concept of a *differential manifold* as the relevant object.

The *line element*, defining infinitesimal distances, of a trajectory in a curved space can be written as

$$\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu \tag{1.1}$$

with a Lorentzian metric tensor $g_{\mu\nu}$. The line element is often a useful way of expressing a metric in a given coordinate system, and so is itself sometimes called the metric.

To arrive at a general metric that describes the global evolution of the Universe, we can invoke the *cosmological principle*. As an overall principle it uses the *a priori* expectation that we are not especially unique in the Universe. Firstly, as observers, we did not pop into existence in a spatial position that is noteworthy. There would be no difference to our average inferences from observations if we were instead translated to an arbitrary different position an arbitrarily large distance away. This translation invariance in space is known as *homogeneity*. The other part of the cosmological principle is *isotropy*, which corresponds to rotational symmetry of our average inferences. If we require that everywhere in space at a given time is isotropic we also get homogeneity for free.

The *Friedmann-Lemaître-Robertson-Walker* (FLRW) metric is a general 4-dimensional Lorentzian metric that is invariant under spatial translation and rotation (homogeneity and isotropy respectively). It is the maximally symmetric solution with these constraints. Using a $(+,-,-,-)$ signature, it is given by

[2] non-gravitational

$$\mathrm{d}s^2 = \mathrm{d}t^2 - a^2(t)\mathrm{d}\sigma^2 \,, \tag{1.2}$$

where $a(t)$ is the *scale factor* and the spatial term $\mathrm{d}\sigma^2$ is a metric on a maximally symmetric three-manifold with the speed of light c set to one. It is worth noting that this has already been slightly motivated by observation, rather than pure reason as it might appear to have been presented. We have assumed it is space, not spacetime, that follows the cosmological principle. As such, time and space are already separated in this metric. The scale factor $a(t)$ is a parameter that is free to vary as a function of time. Physically, this corresponds to how distances between points change as the Universe contracts or, as has been observed (Hubble, 1929), expands.

Using the most general $\mathrm{d}\sigma^2$ in polar coordinates gives us the full FLRW metric

$$\mathrm{d}\sigma^2 = \mathrm{d}t^2 - a^2(t)\left[\frac{\mathrm{d}r^2}{1-Kr^2} + r^2\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right] \,, \tag{1.3}$$

where K, which is the only degree of freedom in the maximally symmetric $\mathrm{d}s^2$, is a constant of global curvature. There is one convention where, with no loss of generality, we could restrict K to values of 0 or ± 1 (with appropriate adjustments to the units of a and r).

Whatever the convention, positive curvature $K > 0$ is known as a *closed* Universe, negative curvature $K < 0$ is known as an *open* Universe, and $K = 0$ is a *flat* Universe. For the flat case, the metric simplifies greatly in Cartesian coordinates

$$\mathrm{d}s^2 = \mathrm{d}t^2 - a(t)^2\delta_{ij}\mathrm{d}x^i\mathrm{d}x^j \,, \tag{1.4}$$

which, but for the scale factor $a(t)$, is the same as the metric for flat Minkowski space of special relativity: $\eta^{\mu\nu} = \mathrm{diag}(1,-1,-1,-1)$.

With homogeneity and isotropy still in mind, we can consider a *perfect fluid*; that is, one which is completely described by its rest-frame energy density ρ and momentum p. Consider the *energy-momentum* (or *stress-energy*) tensor $T^{\mu\nu}$, which is a symmetric (2,0) tensor that represents the flux of four-momentum

$$p^\mu = mU^\mu = m\frac{\mathrm{d}x^\mu}{\mathrm{d}\tau_\mathrm{p}} \,, \tag{1.5}$$

with mass m, and four-velocity U^μ with respect to proper time τ_p.

Taking the cosmological perfect fluid for our flat Universe in Cartesian coordinates, homogeneity makes it possible to define a *comoving* observer and isotropy ensures that momentum density T^{i0} and stress/shear T^{ij} (where $i \neq j$) are zero, as there can be no non-zero 3-vectors. This gives

$$T^{\mu\nu} = \begin{pmatrix} \rho & 0 & 0 & 0 \\ 0 & Pa^{-2} & 0 & 0 \\ 0 & 0 & Pa^{-2} & 0 \\ 0 & 0 & 0 & Pa^{-2} \end{pmatrix} , \tag{1.6}$$

where ρ is the energy density and P is the pressure (Dodelson, 2003) with the speed of light c set to one. The factors of a^{-2} take into account the expansion of the Universe. A comoving observer follows the expansion of the universe, hence the lack of (peculiar) velocity components.

For a non-comoving observer in a general coordinate system (Hobson et al., 2006), the energy-momentum tensor is given as

$$T^{\mu\nu} = (\rho + P)U^{\mu}U^{\nu} - Pg^{\mu\nu} \ . \tag{1.7}$$

For a comoving observer, such that $U^{\mu} = (1,0,0,0)$, we can lower the index of the energy-momentum tensor (equation 1.6) to give it a simpler (metric independent) form

$$T^{\mu}_{\nu} = T^{\mu\gamma}g_{\gamma\nu} = \begin{pmatrix} \rho & 0 & 0 & 0 \\ 0 & -P & 0 & 0 \\ 0 & 0 & -P & 0 \\ 0 & 0 & 0 & -P \end{pmatrix} . \tag{1.8}$$

1.1.2 Evolution of a smooth Universe

The previous section described how we can use homogeneity and isotropy to arrive at the FLRW metric and the energy-momentum tensor for a perfect fluid. The resulting metric has a single degree of freedom, namely the scale factor as a function of time $a(t)$, with the unspecified energy-momentum content encapsulated in the density and pressure terms, ρ and P respectively.

The evolution of the metric follows the Einstein field equations,

$$\begin{aligned} G_{\mu\nu} + \Lambda g_{\mu\nu} &= \frac{8\pi G_N}{c^4} T_{\mu\nu} \\ \text{where } G_{\mu\nu} &= R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} \end{aligned} \tag{1.9}$$

where $R_{\mu\nu}$ and R are the Ricci curvature tensor and scalar respectively, and G_N is the Newtonian gravitational constant. For completeness c has been included, but in what follows we take $c = 1$.

It is possible to subsume the cosmological constant term, Λ, into the stress-energy tensor, $T_{\mu\nu}$, by interpreting the cosmological constant as an energy density. This seems to be preferred by certain communities which wish to go on to interpret the cosmological constant as a quantum vacuum energy effect (e.g. Weinberg 1989) or by those who wish to explain the observed acceleration of the Universe by extensions to known physics. It can additionally be done to make the mathematics appear simpler (as will be done shortly).

I think it is of some importance, however, that when deriving the Einstein field equations from a principle of least action, according to various constraints and caveats, the cosmological constant term appears naturally.

Lovelock's theorem states that deriving a four-dimensional gravitational theory from an action involving the metric tensor and up to its second derivatives, results in the Einstein field equations (equation 1.9) with a cosmological constant term (Lovelock 1971, Lovelock 1972, Clifton et al. 2012). The following action fulfils these requirements,

$$S = \int \left[-\frac{1}{8\pi G_N} (\frac{1}{2} R - \Lambda) + \mathscr{L}_{\mathrm{M}} \right] \sqrt{-g}\, \mathrm{d}^4 x \ , \tag{1.10}$$

where g is the metric determinant, and this yields the Einstein field equation (after some calculus of variations). Written this way, the first term (with the Ricci scalar) is the Einstein-Hilbert action, and the second term is a constant (the cosmological constant) which is free to vary. The final term $\mathscr{L}_{\mathrm{M}}$ is the Lagrangian density for matter, which contributes to the energy-momentum tensor

$$T_{\mu\nu} = \frac{2}{\sqrt{-g}} \frac{\delta(\sqrt{-g}\mathscr{L}_M)}{\delta g^{\mu\nu}} \ . \tag{1.11}$$

Despite the "naturalness" of Λ as an independent term in the action, in what follows it will be treated as a form of matter. This is simpler because a cosmological constant term

can be treated as a perfect fluid (though a strange fluid with negative pressure).

Absorbing Λ into the energy-momentum tensor, we can rewrite the Einstein field equations as

$$R_{\mu\nu} = 8\pi G_N\left(T_{\mu\nu} - \frac{1}{2}g_{\mu\nu}T\right) \tag{1.12}$$

where T is the trace of $T_{\mu\nu}$. Evaluating this leads to only two distinct equations as the off-diagonal terms vanish (due to isotropy), giving a time 00 degree of freedom and only a single spatial ij degree of freedom (due to homogeneity). These two equations can be represented as the *Friedmann equations*,

$$\left(\frac{\dot{a}}{a}\right) = \frac{8\pi G_N}{3}\rho - \frac{K}{a^2} \ ; \tag{1.13}$$

and

$$\left(\frac{\ddot{a}}{a}\right) = -\frac{4\pi G_N}{3}(\rho + 3P) \ , \tag{1.14}$$

where K takes the convention of having units and is free to be an arbitrary real number (not necessarily $0,-1$, or $+1$). The Hubble parameter is defined as

$$H \equiv \frac{\partial_t a}{a} = \frac{\dot{a}}{a} \ . \tag{1.15}$$

Conservation of energy-momentum gives us the final equation constraining the evolution of the scale factor. In General Relativity, this conservation law is expressed as

$$\nabla_\mu T^\mu_\nu = 0 \ , \tag{1.16}$$

where the covariant derivative is defined as

$$\nabla_\mu A^\nu = \partial_\mu A^\nu + \Gamma^\nu_{\mu\sigma}A^\sigma \ , \tag{1.17}$$

$$\nabla_\mu A_\nu = \partial_\mu A_\nu - \Gamma^\sigma_{\mu\nu}A_\sigma \ , \tag{1.18}$$

and

$$\nabla_\mu F^\nu_\gamma = \partial_\mu F^\nu_\gamma + \Gamma^\nu_{\sigma\mu} F^\sigma_\gamma - \Gamma^\sigma_{\gamma\mu} F^\nu_\sigma \ , \tag{1.19}$$

for vectors, covectors and $(1,1)$ tensors respectively.

Combining this with equation 1.8, gives the conservation equation

$$\frac{\dot{\rho}}{\rho} = -3(1+w)\frac{\dot{a}}{a} \ , \tag{1.20}$$

where we have assumed that the fluid obeys the *equation of state*

$$P = w\rho \ . \tag{1.21}$$

We can consider three relevant fluids: *dust* with density ρ_M and negligible pressure $w = 0$ (e.g. stars and galaxies), radiation (or any relativistic particles) with density ρ_R and $w = \frac{1}{3}$, or cosmological constant with density ρ_Λ and $w = -1$. If w is constant (as it is for these examples), we can integrate equation 1.20 to get

$$\rho \propto a^{-3(1+w)} \ . \tag{1.22}$$

For radiation the value $w = \frac{1}{3}$ can be derived using the energy momentum tensor for electromagnetism: $T^{\mu\nu} = F^{\mu\lambda}F^\nu_\lambda - \frac{1}{4}g^{\mu\nu}F^{\lambda\sigma}F_{\lambda\sigma}$ (Carroll, 2003). For a massless photon, evaluating the energy density for $w = \frac{1}{3}$ gives

$$\rho_R \propto a^{-4} \ . \tag{1.23}$$

To explain this result, the energy density of radiation ρ_R changes due to the number density falling as a^{-3} as space expands and a further factor of a^{-1} due to the photons losing energy in an expanding Universe – this is known as *redshift* (c.f. equation 1.32 discussion).

If we define a *density parameter*

$$\Omega_i = \frac{\rho_i}{\rho_c} \tag{1.24}$$

with *critical density*

$$\rho_c = \frac{3H^2}{8\pi G_N} \tag{1.25}$$

for the different cosmological fluids labelled i, then we can rewrite the Hubble parameter using equation 1.13 as

$$H^2 = H_0^2\left(\Omega_{0,R}a^{-4} + \Omega_{0,M}a^{-3} + \Omega_{0,k}a^{-2} + \Omega_{0,\Lambda}\right) \ , \tag{1.26}$$

where the zero subscript refers to the value today. The curvature is subsumed into the curvature density as

$$\Omega_k = 1 - \Omega_R - \Omega_m - \Omega_\Lambda \ , \tag{1.27}$$

where $\Omega_K = 0$ is a flat Universe[3]. At late times (large a), equation 1.26 shows that the radiation density ρ_R becomes negligible compared to the other components.

Equation 1.26 is succinct in showing some of the unknown (or to-be-inferred) quantities of this model. The Hubble parameter on the left hand side is observable[4], and is related to the unknown cosmological parameters on the right hand side.

These equations are the beginnings of the Λ *Cold Dark Matter* (ΛCDM) model, the standard model of cosmology. There are, of course, further elements and possible extensions. One clear extension from these equations is known as wCDM, which allows the *dark energy* fluid which causes acceleration to be something other than Λ (e.g. $w = f(a)$).

1.1.3 Distances

What follows is a brief overview of some of the distance measures in the FLRW Universe.

Freely falling test particles in a curved spacetime follow geodesics, with a path $x^\mu(\lambda)$ obeying the geodesic equation

$$\frac{\mathrm{d}^2x^\mu}{\mathrm{d}\lambda^2} + \Gamma^\mu_{\nu\gamma}\frac{\mathrm{d}x^\nu}{\mathrm{d}\lambda}\frac{\mathrm{d}x^\gamma}{\mathrm{d}\lambda} = 0 \ . \tag{1.28}$$

The affine parameter λ monotonically increases between points, and therefore monotonically increases with time. If it were necessary, we could choose to have the affine parameter be the proper time of the test particle.

A photon is massless, so its four momentum $p_\mu = (E, p^i)$ has zero magnitude

$$p_\mu p^\mu = 0 \ . \tag{1.29}$$

[3] As predicted by inflation (section 1.2) and measured with the Cosmic Microwave Background (CMB, section 1.4.1).

[4] Either directly or indirectly.

Combining this with the Friedmann equations, the energy of the photon E can be related to the scale factor

$$E \propto a^{-1} \quad . \tag{1.30}$$

This matches the redshift result from equation 1.23. From this we can see that the observed wavelength of a photon L_o is redshifted from its emitted wavelength L_e according to the value of the scale factors at those respective points

$$\frac{L_e}{L_o} = \frac{a_e}{a_o} \quad . \tag{1.31}$$

The redshift value

$$z = \frac{L_o - L_e}{L_e} \tag{1.32}$$

is one of the most important quantities for observational cosmology. It can be measured if the emitted frequency of a photon is known, for example by its relation to known features in an object's *spectra* (distribution of emitted photos) for example. Its measured value is then related to the value of the scale factor at the point the photon was emitted

$$a_e = \frac{1}{1+z} \tag{1.33}$$

where the scale factor now a_o is normalised to one by convention. In this regard z, in the absence of peculiar[5] velocities, acts as both a measure of distance and time.

A comoving distance is a rescaling by the time-evolving scale factor of the physical distance, such that two particles initially at rest in the evolving FLRW metric would stay equidistant in comoving coordinates (whatever the behaviour of $a(t)$). The comoving distance (or time with $c = 1$) to an object observed today, which emitted light when the scale factor was a_e, is

$$D_c(a_e) = \int_{t(a_e)}^{t(a_0)} \frac{\mathrm{d}t'}{a(t')} \quad . \tag{1.34}$$

The true amount of time that has elapsed, however, is called the *lookback* time

[5](non-comoving)

$$t_{\text{lookback}}(a_e) = \int_{t(a_e)}^{t(a_0)} \mathrm{d}t' \ . \tag{1.35}$$

The final two distances used often in cosmology are the *luminosity distance* and *angular diameter distance* (Dodelson, 2003).

The luminosity distance is the extension of the inverse square law for observed flux, extended to an FLRW spacetime. If we have observed flux

$$F = \frac{L_*}{4\pi D_L^2} \ , \tag{1.36}$$

from an object with luminosity L_*, then D_L is the luminosity distance. The luminosity distance in terms of comoving distance D_c and redshift is

$$D_L = (1+z) f_K(D_c) \tag{1.37}$$

with function f_K dependent on the curvature K

$$f_K(D_c) = \begin{cases} \frac{1}{H_0\sqrt{|\Omega_K|}} \sin(\sqrt{-\Omega_K} H_0 D_c) & \Omega_K < 0 \\ D_c & \Omega_K = 0 \\ \frac{1}{H_0\sqrt{|\Omega_K|}} \sinh(\sqrt{\Omega_K} H_0 D_c) & \Omega_K > 0 \ . \end{cases} \tag{1.38}$$

Angular diameter distance is an extension to FLRW of a geometric reckoning of distance using an object's physical size X and its angular size θ (assumed to be small), such that

$$D_A = \frac{X}{\theta} \ . \tag{1.39}$$

Evaluating this (Dodelson, 2003) in terms of comoving distance D_c gives

$$D_A = \frac{f_K(D_c)}{1+z} \ . \tag{1.40}$$

1.2 Cosmic inflation

1.2.1 Motivation

Cosmic inflation is a mechanism proposed to solve a series of perceived problems with the initial conditions in the Big Bang cosmological model. Some of these include the problem of the initial singularity ($a \to 0$), the *flatness problem*, and the *horizon problem*. The initial steps towards the inflationary model were made by Guth and Tye (1980) to address the apparent lack of magnetic monopoles predicted by Grand Unified Theories (GUT) .

The view held by the Astronomer Royal, Martin Rees, regarding this monopole argument,

> "*Sceptics about exotic physics might not be hugely impressed by a theoretical argument to explain the absence of particles that are themselves only hypothetical*" (Rees, 1997) ,

might be extended to some of the justifications listed so far. Some of these perceived problems may only be problems from a certain prior world-view.

As will be explained in more detail, the flatness and horizon problems (Guth, 1981) are related to the observed lack of global curvature[6] and to the principle of homogeneity. Inflation could explain why the Universe seems to have these properties, but it could be that the Universe is just the way it is.

It is a tricky business trying to assign probabilities to different hypothetical physics. Without a prior probability for a given configuration of the Universe, it becomes difficult to then be surprised that the Universe is a certain way (e.g. flat or homogeneous). This is not to say that resolving these problems is not good a justification for inflation; it might be that explaining flatness is better than just accepting it.

It seems, however, that the value of the inflationary theory is really in its prediction of overall homogeneity with small fluctuations. Primordial fluctuations from the inflationary period (Mukhanov and Chibisov 1981, Mukhanov and Chibisov 1982, Hawking 1982, Starobinsky 1982, Guth and Pi 1982, Bardeen et al. 1983) are an ideal way to seed the growth of structure in the Universe. The prediction of an almost (but not quite) flat spectrum of primordial fluctuations at different length scales has been since verified by Cosmic Microwave Background experiments (Planck Collaboration et al. 2016, Planck Collaboration et al. 2018a).

[6] Ω_K observed as zero or close to zero

A further prediction of many inflationary models is near-Gaussian initial conditions, which will be of importance throughout this thesis.

1.2.2 Solution to flatness and the horizon problems

We can consider the *comoving Hubble radius*

$$r_{\text{Hubble}} = \frac{1}{aH} \tag{1.41}$$

and compare it with Friedmann equation 1.13 rewritten as

$$\Omega_K = \frac{-K}{(aH)^2} \ , \tag{1.42}$$

which shows that deviations from flatness increase with increasing comoving Hubble radius. In particular, as long as $w > -\frac{1}{3}$, any small deviation from $\Omega_K = 0$ will grow over time. If the Universe appears close to flat now, it must have been much closer to flat in its primordial state.

For the horizon problem (see Baumann 2009 for a review), we take comoving distance (equation 1.34) as a "time since the initial conditions" coordinate. This is known as conformal time

$$\eta(t) = \int_0^t \frac{\mathrm{d}t'}{a(t')} \ . \tag{1.43}$$

This can be interpreted as a *causal*, *particle* or *comoving horizon*. To explain this, we can consider the FLRW metric (equation 1.3) of a flat Universe in one spatial dimension,

$$\mathrm{d}s^2 = \mathrm{d}t^2 - a^2(t)\mathrm{d}\chi^2 \tag{1.44}$$

for a particle with radial trajectory (with coordinate χ). This can be rewritten in term of conformal time as

$$\mathrm{d}s^2 = a^2(\eta)(\mathrm{d}\eta^2 - \mathrm{d}\chi^2) \ . \tag{1.45}$$

In this new coordinate system, a photon following a null geodesic $\mathrm{d}s^2 = 0$ will have a trajectory

$$\chi = \text{constant} \pm \eta \ , \tag{1.46}$$

which corresponds to two straight lines in the χ-η plane crossing at a single point representing an event. This is known as the *light cone*.

In the $(+,-,-,-)$ metric, two points in spacetime separated by $\mathrm{d}s^2 > 0$ are *timelike* and two separated by $\mathrm{d}s^2 < 0$ are *spacelike*. A pair of points with a timelike or null (also known as *lightlike*) interval can be in causal contact, meaning they can share information by exchange of particles from one to the other.

The comoving horizon (in terms of the scale factor or the comoving Hubble radius) is

$$\eta(a) = \int_0^a \frac{\mathrm{d}a'}{Ha'^2} = \int_{-\infty}^{\ln a} r_{\mathrm{Hubble}}(a')\mathrm{d}[\ln a'] \ . \tag{1.47}$$

Comparing this with the Friedmann equations (1.13 and 1.14), we can see that the comoving horizon grows monotonically with time for any combination of matter and radiation (as $w_i > 0$). This means that in a Universe with matter and radiation, if a length scale drops below the comoving horizon as the Universe expands, then it had previously always been larger than the comoving horizon and it spans points that have never been in causal contact.

In the classical example of the horizon problem, we consider temperature measurements of the CMB from photons emitted at last-scattering ($z \sim 1100$) from different patches on the sky. Patches that should not have been in causal contact appear to have the same average temperature and have correlated perturbations over causally disconnected regions.

As a possible solution to this, we can look again at equation 1.47. The comoving Hubble radius r_{Hubble} can be thought of as a distance over which particles can travel by a given time, or as the largest scales observable at a given time. The comoving horizon η can be thought of as a distance over which information could travel in the whole history of the Universe. To make the patches of the CMB in causal contact, we need η to be much bigger now than r_{Hubble} (Dodelson, 2003).

Inflation was proposed to be a period in the early Universe during which the comoving Hubble radius r_{Hubble} shrank. During inflation, length scales that were previously well within the comoving Hubble radius (and which were therefore in causal contact) became larger than r_{Hubble}. In the late Universe, as r_{Hubble} increases, length scales that previously had been lost re-enter the growing *Hubble sphere*. To an observer who wasn't aware of inflation, regions that seem to be newly coming into contact with each other appear to have already been interacting (see figure 1.1).

The Friedmann equations (1.13 & 1.14) show that for the Hubble radius r_{Hubble} to

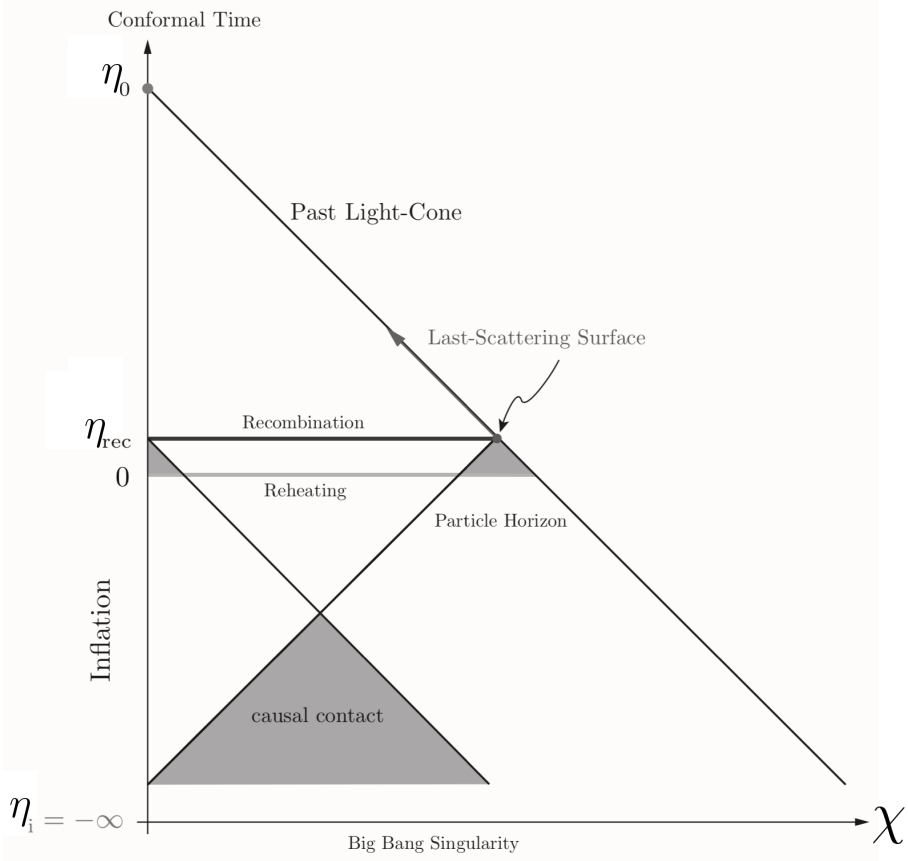

Figure 1.1: This conformal diagram shows how patches of the CMB (with photons emitted at the last scattering surface) could have been in causal contact in a previous inflationary period. Adapted from Baumann (2009), where Baumann defines conformal time at the end of inflation (reheating) with $\eta = 0$ as an *"apparent" Big Bang*; this would be the initial singularity if inflation were not considered and would lead to the existence of the horizon problem.

shrink, we require $w < -\frac{1}{3}$, which corresponds to an accelerating Universe ($\ddot{a} > 0$). For this reason, inflation is often referred to as an early period of accelerated expansion.

Accelerated expansion during an inflationary period not only solves the horizon problem, but also the flatness problem. As can be seen from equation 1.14, acceleration corresponds to $w < -\frac{1}{3}$, which for flatness will bring Ω_K extremely close to zero (equation 1.42). Physically, one can think of space stretching out any initial curvature.

1.2.3 Slow-roll inflation

A standard approach begins by assuming that inflation is driven by a scalar field ϕ called the *inflaton*[7]. Many of the details of inflation, including its duration, are somewhat model dependent (see Carroll 2003 and Baumann 2009 for overviews).

[7]Multiple field inflation is an important area of research, but its results are not directly relevant to the work in this thesis.

We can take a typical Lagrangian density for a single scalar field

$$\mathscr{L}_\phi = \frac{1}{2} g^{\mu\nu}\, \nabla_\mu \phi\, \nabla_\nu \phi - V(\phi) \quad . \tag{1.48}$$

The energy-momentum tensor associated with this Lagrangian is

$$T_{\mu\nu} = \nabla_\mu \phi\, \nabla_\nu \phi - g_{\mu\nu}\Big(\frac{1}{2} g^{\gamma\sigma}\, \nabla_\gamma \phi\, \nabla_\sigma \phi - V(\phi)\Big) \quad , \tag{1.49}$$

which can be compared to the energy-momentum tensor of the perfect fluid for a comoving observer (see equations 1.6 & 1.8), to get the equation of state

$$w = \frac{P_\phi}{\rho_\phi} = \frac{\frac{1}{2}\dot{\phi}^2 - V}{\frac{1}{2}\dot{\phi}^2 + V} \quad . \tag{1.50}$$

From this equation of state, we can see that if the potential term dominates over the kinetic, such that $V(\phi) \gg \dot{\phi}^2$, then we get an accelerating Universe with $w \approx -1$.

This fulfils our requirements for an inflationary period ($w < -\frac{1}{3}$), but only if $V(\phi) \gg \dot{\phi}^2$ is maintained for a sufficient duration. This is where *slow roll* (Linde 1982, Albrecht and Steinhardt 1982) comes in. We can imagine a potential with a few conditions that leads to an inflationary period of expansion that then stops after a certain amount of time.

The Langrangian for the inflaton ϕ (equation 1.48) yields the equation of motion

$$0 = g^{\mu\nu} \nabla_\mu \nabla_\nu \phi + \frac{\mathrm{d}V}{\mathrm{d}\phi} \quad . \tag{1.51}$$

For a homogeneous inflaton field ($\partial_i \phi = 0$) the equation of motion becomes

$$0 = \ddot{\phi} + 3H\dot{\phi} + \frac{\mathrm{d}V}{\mathrm{d}\phi} \quad . \tag{1.52}$$

The behaviour of ϕ can be related to the expansion of the Universe using the Friedmann equation 1.13 combined with the equation of state for inflation (equation 1.50), giving

$$\left(\frac{\dot{a}}{a}\right)^2 = H^2 = \frac{8\pi G_N}{3}\left(\frac{1}{2}\dot{\phi}^2 + V(\phi)\right) \quad . \tag{1.53}$$

We can require that slow roll keeps both the kinetic $\dot{\phi}^2$ and acceleration terms $\ddot{\phi}$ of equations 1.52 & 1.53 small

$$\begin{aligned} V &\gg \dot{\phi}^2 \\ |3H\dot{\phi}| &\gg |\ddot{\phi}| \\ \left|\frac{\mathrm{d}V}{\mathrm{d}\phi}\right| &\gg |\ddot{\phi}| \ , \end{aligned} \tag{1.54}$$

and define the slow roll parameters as

$$\begin{aligned} \varepsilon_V &= \frac{1}{2}\left(\frac{V_{,\phi}}{V}\right)^2 \\ \eta_V &= \frac{1}{2}\frac{V_{,\phi\phi}}{V} \ , \end{aligned} \tag{1.55}$$

with notation $V_{,\phi\phi} = \partial_\phi \partial_\phi V$, where sometimes these are defined with a Planck mass factor $M_{\rm pl}^2$ to make them dimensionless.

While these conditions are being fulfilled, the inflaton field is "slowly rolling" down the potential V and accelerated expansion is continuing. At the end of inflation, ϕ falls into a minimum $V = 0$, at which point the field is conjectured to couple with the standard model (and hypothesised dark matter) fields and oscillates away its energy. This period is known as *reheating*.

1.2.4 Initial conditions

During inflation small-scale quantum fluctuations could be rapidly expanded to extremely large scales. These fluctuations could then form the initial perturbations that grew into the large-scale structure in the Universe today (see Lahav and Liddle 2014 and Tanabashi et al. 2018 for modern implications).

In harmonic (Fourier) space, any oscillating mode with length $\frac{2\pi}{k}$ is *frozen out* if, during inflation, the Hubble radius drops such that $r_{\rm Hubble} \ll \frac{2\pi}{k}$. At this point, scales which are larger than the Hubble radius can no longer have any effect as they span regions that are not in causal contact. The quantum fluctuations imprint random initial conditions which only then begin evolving once a length scale comes back into play as the post-inflationary Universe expands.

I will not include any depth of detail regarding this process. Though a rich subject, its relevance to the work in this thesis is through its generation of near-Gaussian initial conditions with a *primordial power spectrum*.

As an overall picture, we can imagine an inflating Universe flattening out any existing perturbations and rapidly diluting any existing particles, leaving a vacuum in an accelerating Universe. The Gibbons-Hawking temperature

$$T_{\mathrm{GH}} = \frac{H}{2\pi} \propto \sqrt{V} \ , \tag{1.56}$$

is the temperature of a quantum vacuum state in expanding de Sitter ($w = -1$) space (Gibbons and Hawking, 1977). This is analogous to the Unruh effect and the corresponding temperature apparent to an accelerating observer in a vacuum (Unruh, 1976).

During inflation, fluctuations at different scales are frozen out at $k \approx aH$ with amplitudes related to $|\Delta\phi| \approx T_{\mathrm{GH}}$. The variance of the fluctuations at different $\frac{2\pi}{k}$ scales can be related to the value of the potential V and its derivative $V_{,\phi}$, which are combined into the slow roll parameters (equation 1.55).

The variance of harmonic modes at different length scales can be represented by the *power spectrum*. For an isotropic and homogeneous field $\delta(\vec{x})$ with three-dimensional coordinate $\vec{x}$, it is defined by

$$\langle \hat{\delta}(\vec{k}) \hat{\delta}(\vec{k}')^* \rangle = (2\pi)^3 \delta_{\mathrm{D}}(\vec{k} - \vec{k}') P(k) \ , \tag{1.57}$$

where $\hat{\delta}(\vec{k})$ is the Fourier transform of $\delta(\vec{x})$, given by

$$\hat{\delta}(\vec{k}) = \int_{\mathbb{R}^3} \mathrm{d}^3 x \, \mathrm{e}^{i\vec{x}\cdot\vec{k}} \, \delta(\vec{x}) \ , \tag{1.58}$$

and k is the magnitude of $\vec{k}$ (Peebles, 1980). The statistics of a homogeneous field are invariant under spatial transformation so the Fourier modes are uncorrelated, hence the Dirac delta factor $\delta_{\mathrm{D}}(\vec{k} - \vec{k}')$. Isotropy means the power spectrum is a function of only the magnitude, k, not of $\vec{k}$.

In most inflationary models, correlations of order higher than two are suppressed (Weinberg, 2008), so the two-point correlation terms dominate. The initial energy-density perturbations $\delta\rho$ imprinted by inflation are therefore near-Gaussian (see section 1.2.5) for many, often simpler (e.g. Starobinsky 1982), models. More complex models of inflation may lead to deviations from Gaussianity (e.g. Allen et al. 1987, Kofman and Pogosyan 1988, Salopek et al. 1989), a detection of which may give evidence for a given inflationary model (Verde and Heavens, 2001).

The primordial power spectrum is close to scale invariant. If we define a *dimensionless power spectrum*

$$\Delta^2(k) = \frac{k^3}{2\pi^2} P(k) \ , \tag{1.59}$$

then

$$\frac{\mathrm{d}[\ln \Delta^2(k)]}{\mathrm{d}[\ln k]} = 0 \tag{1.60}$$

would correspond to a scale-free spectrum. This had already been proposed as a primordial power spectrum before any prediction from inflation (Harrison 1970, Zeldovich 1972, Peebles and Yu 1970).

Fortunately for our ability to test inflation, the primordial power spectrum from inflation is only nearly scale-free. For slow-roll models that are entirely characterised by V and $V_{,\phi}$, the spectrum of the scalar metric perturbations can be expressed as

$$\frac{\mathrm{d}[\ln \Delta_s^2(k)]}{\mathrm{d}[\ln k]} = n_s - 1 \ , \tag{1.61}$$

where

$$n_s = 2\eta_V - 6\varepsilon_V + 1 \ . \tag{1.62}$$

As I will discuss in section 1.3.1, perturbations to the metric (caused by quantum fluctuations in this case) can be decomposed into scalar, vector and tensor components. Vector components are irrelevant here, but the tensor component corresponds to gravitational waves sourced by the same fluctuations.

In addition to a scalar spectrum, there is a tensor spectrum

$$\frac{\mathrm{d}[\ln \Delta_t^2(k)]}{\mathrm{d}[\ln k]} = n_t = -2\varepsilon_V \ . \tag{1.63}$$

Having a nearly scale-free spectrum corresponds to $n_s - 1 \ll 1$ and $n_t \ll 1$, which is expected from the slow-roll conditions. Measurement, or inference, of n_s and n_t is another task for observational cosmology. As their values are model dependent, they provide a way towards understanding the form of the potential $V(\phi)$.

In section 1.4.1 I will discuss the current constraints on the inflationary model from

the CMB.

1.2.5 Gaussian random fields

Gaussianity is a statistical property which means a field (e.g. perturbations imprinted by inflation) are randomly drawn from a Gaussian distribution. A zero-mean Gaussian field $f(\vec{x})$ has a probability distribution given by

$$P(f(\vec{x}))\mathrm{d}^D x = \frac{1}{\det|2\pi\Sigma|}\exp\Big(-\frac{1}{2}f(\vec{x})^T\Sigma^{-1}f(\vec{x})\Big)\mathrm{d}^D x \ , \tag{1.64}$$

where $\langle f(\vec{x})f(\vec{x})^T\rangle = \Sigma$ is the covariance matrix and the dimension D is the number of elements in the vector $\vec{x}$. It is simple to generalise to a field $g(\vec{x})$ with a non-zero mean μ by the transformation $g(\vec{x}) = f(\vec{x}) + \mu$. Equivalently this transformation can remove the mean. This has allowed us so far to think of perturbations $\delta(\vec{x})$ as being Gaussian without, as yet, thinking about the total density.

We can imagine that if the Universe were run many times over, the perturbations would be initialised differently, with each realisation of these perturbations being independent and identically distributed (according the same probability distribution). As we only observe one Universe, the property of *ergodicity* is invoked to say that averaging over large volumes would be the same as averaging over many realisations of the Universe (Peacock, 1999).

The covariance matrix is the generalisation of variance to include correlations between different elements of $f(\vec{x})$. As previously discussed, homogeneity means that Fourier modes are uncorrelated, which would mean that the covariance matrix of $\hat{\delta}(\vec{k})$ would be diagonal.

This view of things makes it possible to think again about homogeneity and isotropy as presented initially. The Universe is clearly not completely homogeneous; there are observable perturbations on all scales (e.g. galaxies, the cosmic-web, people). The Universe is *statistically homogeneous*, meaning that these perturbations average away over many realisations.

For correlations of Gaussian random fields with greater than order two, we can invoke Wick's (or more fairly Isserlis') theorem (Wick 1950, Isserlis 1918). With our notation, it implies that

$$\begin{aligned}\langle f(\vec{x}_1)...f(\vec{x}_{2n+1})\rangle &= 0 \\ \langle f(\vec{x}_1)...f(\vec{x}_{2n})\rangle &= \sum_{\substack{\text{distinct} \\ \text{partitions} \\ \text{into pairs}}} \prod_{\substack{n \text{ pairs} \\ i,j}} \langle f(\vec{x}_i) f(\vec{x}_j)\rangle \end{aligned} \tag{1.65}$$

for integer n. Taking the examples of $n = 3$ & 4 with a shorthand notation for our Gaussian field $f(\vec{x}_i) = \vec{f}_i$, Wick's theorem gives

$$\begin{aligned}\langle \vec{f}_1 \vec{f}_2 \vec{f}_3\rangle &= 0 \\ \langle \vec{f}_1 \vec{f}_2 \vec{f}_3 \vec{f}_4\rangle &= \langle \vec{f}_1 \vec{f}_2\rangle\langle \vec{f}_3 \vec{f}_4\rangle + \langle \vec{f}_1 \vec{f}_3\rangle\langle \vec{f}_2 \vec{f}_4\rangle + \langle \vec{f}_1 \vec{f}_4\rangle\langle \vec{f}_2 \vec{f}_3\rangle \ .\end{aligned} \tag{1.66}$$

For a Gaussian random field, the expected odd-number correlations are zero and the even-number correlations are combinations of the two-point correlations.

This confirms something that was clear from the form of equation 1.64, that the Gaussian random field is entirely characterised by its mean and its two-point correlations (i.e. the covariance matrix). It additionally provides ways to measure non-Gaussianity, for example by measuring a non-zero three-point correlation.

If we do not want to work in Fourier space with the power spectrum, but rather in *configuration space* where $\vec{x}$ lives, then we can use the two-dimensional correlation function

$$\xi(\vec{r}) = \langle \delta(\vec{x})\delta(\vec{x}+\vec{r})\rangle \ , \tag{1.67}$$

which can be related to the power spectrum using equation 1.57 (Peacock, 1999) to give

$$\xi(\vec{r}) = \frac{1}{(2\pi)^3}\int \mathrm{d}^3 k P(k) \mathrm{e}^{-i\vec{k}\cdot\vec{r}} \ . \tag{1.68}$$

Keeping in harmonic space, but remembering that observations are often on the celestial sphere, we turn to spherical harmonic analysis. For a field on the sphere $f(\vec{\phi})$ defined at angular positions[8] $\vec{\phi}$ (where the two elements of $\vec{\phi}$ could be *right ascension* and *declination*), the spherical harmonic transform is given by

[8] note the arrow indicating a vector, which is used to distinguish this from the inflaton

$$f(\vec{\phi}) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{+\ell} f_{\ell m} Y_{\ell m}(\vec{\phi}) \ . \tag{1.69}$$

The spherical harmonic coefficients $f_{\ell m}$ are the spherical analogue of $\hat{\delta}(\vec{k})$. Similarly, the $Y_{\ell m}$ form a complete set of eigenfunctions for an expansion on the sphere in the way that as $e^{i\vec{k}\cdot\vec{x}}$ do for Euclidean space in Fourier analysis (Dodelson, 2003). The *angular power spectrum* is given by

$$\langle f_{\ell m} f^*_{\ell' m'} \rangle = \delta^K_{\ell\ell'} \delta^K_{mm'} C_\ell \tag{1.70}$$

where δ^K is the Kronecker delta function.

1.3 Growth of structure

1.3.1 Scalar perturbations

Solving the Einstein equations in the isotropic and homogeneous FLRW Universe does recover the global time evolution in the form of the Friedmann equations (equations 1.13 & 1.14), which are functions of $\rho(t)$ and $P(t)$ for different spatially-constant fluids. To go further and predict the formation of structure and the trajectories of light, we need to consider perturbations to the FLRW metric that represent *inhomogeneities*.

We can split the metric into an FLRW ($\bar{g}$) and perturbative (h) part

$$g_{\mu\nu} = \bar{g}_{\mu\nu} + h_{\mu\nu}. \tag{1.71}$$

We assume that $h_{\mu\nu}$ is small and does not effect the global evolution of the scale factor a of the background $\bar{g}_{\mu\nu}$. There is a proposed, and controversial, process by which structure formation does lead to changes in the evolution of the scale factor; this *backreaction* debate appears to not yet have disappeared (Buchert and Räsänen 2012, Kaiser 2017).

The metric tensor must be symmetric, so there are 10 degrees of freedom in $h_{\mu\nu}$. For the flat, $K = 0$, case, the perturbed metric can be written as

$$\mathrm{d}s^2 = a(\eta)^2\Big((1+2A)\mathrm{d}\eta^2 - 2B_i\mathrm{d}x^i\mathrm{d}\eta - (\delta^K_{ij} + h_{ij})\mathrm{d}x^i\mathrm{d}x^j\Big) \ . \tag{1.72}$$

Scalar-vector-decomposition can then split B_i into the sum of its scalar and (divergence-free) vector parts, and by extension h_{ij} into 2 scalar, 1 vector and 1 tensor part. At linear order these decomposed parts are decoupled and evolve independently (Dodelson, 2003). A certain linear combination of these, for which the metric is invariant under any general coordinate transformation, are the Bardeen variables which are *gauge independent* (Bardeen, 1980).

I will not discuss the issue of gauge with great depth, as the final standard, quasi-Newtonian result is the goal here. The choice of gauge is related to how $h_{\mu\nu}$ is separated from the background, as the background itself can be changed by a coordinate transformation (Ellis and van Elst, 1999). The fact that changing gauge can give the same observable prediction is *gauge freedom*.

It is possible to fix the gauge, by choosing a coordinate transformation that simplifies an expression or calculation. The metric in the Newtonian gauge (equivalent to fixing $B =$

$E_{\text{Bardeen}} = 0$), with only the scalar perturbations, is given by

$$\mathrm{d}s^2 = a(\eta)^2\Big((1+2\Phi)\mathrm{d}\eta^2 - (1-2\Psi)\delta_{ij}\mathrm{d}x^i\mathrm{d}x^j\Big) \tag{1.73}$$

where $\Phi = A$ and -2Ψ is the scalar part of h_{ij}. The gauge looks similar to the Newtonian limit in Minkowski space, as the time and space parts are orthogonal, and if there is no anisotropic stress ($\Psi = \Phi$) these scalar perturbations behave like the Newtonian gravitational potential.

We are focussing on the scalar part, which represents compressional modes, as these are the (primary) cause of structure formation in the Universe (Weinberg, 2008). Tensor modes have already been discussed as primordial gravitational waves in the context of inflation. The vector components, which represent vortical modes, can be safely ignored here as they rapidly decay with the expansion of the Universe.

Solving the Einstein equations (1.9) to predict the evolution of the scalar perturbation in the Newtonian gauge Φ at linear order, a long exercise not repeated here (Dodelson 2003, Weinberg 2008, Baumann 2014), gives

$$-k^2\Phi - \mathscr{H}\big(\Phi' + \mathscr{H}\Phi\big) = 4\pi G_N a^2\ \delta\rho \ , \tag{1.74}$$

$$ik_i(\Phi' + \mathscr{H}\Phi) = -4\pi G_N a^2\ \big(\bar{\rho} + \bar{P}\big)\delta U_i \ , \tag{1.75}$$

$$\Phi'' + 3\mathscr{H}\Phi' + \big(2\mathscr{H}' + \mathscr{H}^2\big)\Phi = 4\pi G_N \delta P \ , \tag{1.76}$$

where the derivative with respect to conformal time is

$$A' = \frac{\mathrm{d}A}{\mathrm{d}\eta} \ , \tag{1.77}$$

and the conformal Hubble parameter, $\mathscr{H}(\eta)$, is defined as

$$\mathscr{H} = \frac{d\ln a}{d\eta} \ . \tag{1.78}$$

We can define *overdensity* δ with respect to the mean background density $\bar{\rho}$ as

$$\delta(t,\mathbf{x}) = \frac{\delta\rho(\mathbf{x},t)}{\bar{\rho}(t)} = \frac{\rho(\mathbf{x},t)-\bar{\rho}(t)}{\bar{\rho}(t)} \ . \tag{1.79}$$

For scales which are far below the Hubble radius $k^{-1} << r_{\rm Hubble}$, which are often called *sub-horizon*, equation 1.74 reduces to the Newtonian form of the Poisson equation

$$\nabla^2\Phi = 4\pi G_N a^2\ \bar{\rho}\ \delta \ , \tag{1.80}$$

with a scale factor contribution accounting for expansion[9].

In addition to solutions using the Einstein equations, we can also make use of the conservation equation (1.20). The time component of this gives the relativistic *continuity equation*

$$\delta\rho' + 3\mathscr{H}\left(\delta\rho+\delta P\right) + \left(\bar{\rho}+\bar{P}\right)\left(ik_i\delta U^i - 3\Phi'\right) = 0 \ , \tag{1.81}$$

and the spatial component gives the relativistic *Euler equation*

$$\delta U_i{}' + \mathscr{H}\ \delta U_i + ik_i\Phi + \frac{\bar{P}'\ \delta U_i + ik_i\delta P}{\bar{\rho}+\bar{P}} = 0 \ . \tag{1.82}$$

This acts per fluid component, unlike the Einstein equations which act on the sum. For a single evolving fluid, conservation is already intrinsic to General Relativity. The set for all fluids form a set of Boltzmann equations (Dodelson, 2003)[10] which combined with the Einstein equation results can be used to model the evolution of perturbations.

1.3.2 Growing modes

As the work in this thesis is focussed on weak lensing and maps of dark matter from the late Universe, here I will present a few results relevant to structure formation in the late Universe.

In this case "late" means that the Friedmann equations are dominated by the matter (or late-time dark energy) terms, so $\bar{\rho}_R$ is negligible. It also means the time after baryons and radiation have decoupled, so we can assume there is a single matter component with $P \ll \rho$. I will also ignore the (comparatively small) effect of neutrinos, which free-stream in the late Universe and suppress the formation of structure (Park et al., 2012).

[9] This can also be achieved by eliminating the second term of equation 1.74 using equation 1.74 and redefining only the right hand side of equation 1.80 to be in a different gauge.

[10] We have already used a few assumptions (e.g. anisotropic stress) which need not be made in the full analysis (e.g. including neutrinos).

In this regime (Dodelson 2003, Baumann 2014) the continuity and Euler equations reduce to

$$\delta' + \nabla \cdot \mathbf{u} - 3\Phi' = 0 \ , \tag{1.83}$$

$$\frac{\mathrm{d}\mathbf{u}}{\mathrm{d}\eta} + \mathscr{H}\mathbf{u} + \nabla\Phi = 0 \ . \tag{1.84}$$

where, having moved out of Fourier space, the simpler notation for the peculiar velocity $\mathbf{u}$ is used.

Combining equations 1.74 & 1.76 with $w = 0$ gives $\Phi = constant$, so we can drop the Φ' term in equation 1.83. We can say that Φ is frozen during matter domination. Then, rewriting the Poisson equation in the sub-horizon limit (equation 1.80) as

$$\nabla^2\Phi(\mathbf{x}) = \frac{3}{2}\Omega_m(\eta)\mathscr{H}^2(\eta)\delta(\mathbf{x},\eta) \ , \tag{1.85}$$

allows us to eliminate the spatial derivatives from equations 1.83 & 1.84, which gives

$$\delta''(\mathbf{k},\eta) + \mathscr{H}\delta'(\mathbf{k},\eta) = \frac{3}{2}\Omega_m(\eta)\mathscr{H}^2(\eta)\delta(\mathbf{k},\eta) \ , \tag{1.86}$$

a second order differential equation (Scoccimarro 1998, Bernardeau et al. 2002). I have arbitrarily (as there are no spatial derivatives) written these perturbations as functions of $\mathbf{k}$ here, as it is often more useful to imagine evolving harmonic modes.

We now go to linear order in perturbation theory

$$\delta(\mathbf{k},\eta) = D_1(\eta)\delta(\mathbf{k},0) + \ldots \tag{1.87}$$

for which we ignore the decaying solution of equation 1.86 (which is second-order and therefore permits two independent solutions) and call D_1 the *linear growing mode*.

A simple solution for D_1 is with $\Omega_m = 1$ and $\Omega_K = 0$, the so-called Einstein-de Sitter Universe (Einstein and de Sitter, 1932). In this case the linear growing mode would have a scaling given by

$$D_1(t) \propto t^{\frac{2}{3}} \ , \tag{1.88}$$

showing that structure grows with time in a flat, matter-dominated Universe.

The evolution of D_1 is often represented by

$$f = \frac{\mathrm{d}\ln D_1}{\mathrm{d}\ln a} \tag{1.89}$$

which for a flat Universe, $1 = \Omega_m + \Omega_\Lambda$, can be parameterised as approximately $f = \Omega^\gamma$ with $\gamma \approx \frac{5}{9}$ (Peebles 1980, Lightman and Schechter 1990, Lahav et al. 1991, Bouchet et al. 1995). Modifications to ΛCDM or General Relativity would give a different value of γ.

1.3.3 Non-linear structure

The research in this thesis focusses on dark matter maps from weak lensing data. The large-scale structures featured in these maps are from the late Universe, where non-linear collapse has led to a highly non-Gaussian density field. To understand and model the density distribution in this case, we cannot rely on the linear techniques introduced in the previous section.

There are many techniques to model non-linear structures in the Universe. In the initial small δ regime, perturbation theory (PT) has some success (as will be discussed shortly), but will quickly become unhelpful for truly non-linear dynamics. Of particular historical significance was the theory of Press and Schechter (1974), which introduced a simple model of spherically symmetric collapse resulting in a mass distribution of virialised *halos*. Though this theory and its extension are a good approximation (Sheth and Tormen 1999, Jenkins et al. 2001), the theory does not generate predictions which match results from N-body simulations exactly. N-body simulations, and software[11] combining N-body and PT, will be used throughout this thesis to model the non-linear density field in the late Universe.

PT is a powerful predictive tool to model the formation of structure evolving from the Gaussian initial conditions, even into the mildly non-linear regime. The technique of *Langrangian perturbation theory* (LPT), which can can be contrasted with the more natural *Eulerian perturbation theory* (Scoccimarro 1998, Bernardeau et al. 2002) uses a new, Lagragian, coordinate system. We replace the standard Eulerian coordinates $\mathbf{x}$ with a Lagrangian coordinate $\mathbf{q}$,

$$\mathbf{x}(\eta) = \mathbf{q} + \psi(\mathbf{q}, \eta) \ . \tag{1.90}$$

[11] L-PICOLA

The two coordinate systems are related through a *displacement field* ψ, which becomes the time-evolving object. As $\mathbf{q}$ does not evolve, it corresponds to initial particle positions.

To transform between densities in the different coordinate systems the Jacobian is used

$$J(\mathbf{q},\eta) = \det\left|\frac{\partial q_i}{\partial x_j}\right| = \det\left|\delta^K_{ij} + \frac{\partial \psi_i}{\partial x_j}\right| \ . \tag{1.91}$$

We can also enforce a conservation of mass constraint in the new coordinate system

$$\rho(\mathbf{x},\eta)\mathrm{d}^3x = \rho(\mathbf{q})\mathrm{d}^3q \ , \tag{1.92}$$

which, after dividing by the mean density $\bar{\rho}$, gives the Jacobian as

$$J(\mathbf{q},\eta) = \frac{1}{1+\delta(\mathbf{x},\eta)} \ . \tag{1.93}$$

With the new coordinates, equation 1.84 is simply now in terms of the displacement field

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}\eta^2} + \mathscr{H}\frac{\mathrm{d}\psi}{\mathrm{d}\eta} + \nabla_{\mathbf{x}}\Phi = 0 \ , \tag{1.94}$$

where the gradient is still in terms of the Eulerian coordinates. Multiplying by the Jacobian (equation 1.93) and taking the gradient gives

$$J\nabla_{\mathbf{x}}\left(\psi'' + \mathscr{H}\psi'\right) = \frac{3}{2}\mathscr{H}^2(J-1) \ , \tag{1.95}$$

where we have used the Poisson equation 1.85 on the right hand side. To get the gradient in Lagrangian coordinates, the Jacobian can be used for coordinate transformation.

The *Zel'dovich Approximation* (ZA) is the linear order solution to equation 1.95 (Zel'dovich, 1970). We expand the displacement to linear order

$$\psi = \psi^{(1)} + \dots \ , \tag{1.96}$$

where

$$\psi^{(1)} = b(\eta)\mathbf{p}(\mathbf{x}) \ . \tag{1.97}$$

The solution to first order is

$$\nabla_{\mathbf{q}}\psi^{(1)}(\mathbf{q},\eta) = -D_1(\eta)\delta(\mathbf{q}) \ , \tag{1.98}$$

where D_1 is the linear growth from equation 1.87 and $\delta(\mathbf{q})$ are the initial (Gaussian) matter fluctuation.

The equations of motion can then evolve the linear growth factor, $D_1(\eta)$, independently. The dynamics here lead to growth of the density perturbations into flattened *Zel'dovich pancakes*, due to structures collapsing in the direction with negative eigenvalues λ_i (with our sign definition) of the matrix $p_{i,j}$. This behaviour can be shown by expanding the Jacobian to linear order, which gives

$$\begin{aligned} 1+\delta(\mathbf{x},\eta) &= \det\left|\delta_{ij} + b(\tau)\frac{\partial p_i}{\partial q_j}\right|^{-1} \\ &= 1 - b(\tau)\sum_i \lambda_i + \mathscr{O}(\psi^2) \ , \end{aligned} \tag{1.99}$$

where λ_i are the eigenvalues of the matrix $p_{i,j} = \frac{\partial p_i}{\partial q_j}$. Continuing at linear order gives

$$\begin{aligned} \delta(\mathbf{x},\eta) &= -b(\eta)\sum_i \lambda_i \\ &= -b(\eta)\nabla_{\mathbf{q}}\mathbf{p}(\mathbf{q}) \\ &= -\nabla_{\mathbf{q}}\psi^{(1)} \ . \end{aligned} \tag{1.100}$$

The appearance of ZA structure is in agreement with the filamentary structure of full N-body dynamics, though the lack of higher order "damping" terms lead to catastrophic caustic behaviour (Shandarin and Zeldovich, 1984). Even a simple collapsing cloud of particle leads to a singularity (Peacock, 1999).

The point of failure is sometimes known as *shell crossing*, and can be interpreted as the coordinate transformation becoming ill-defined due to multiple streams being displaced from initial $\mathbf{q}$ points to the same $\mathbf{x}$ (Bernardeau et al., 2002). This failure is not only restricted to ZA; problems with shell crossing are inherent to all LPT due to the coordinate transform (Carlson et al., 2009).

Second order Lagrangian perturbation theory (2LPT)

$$\psi = \psi^{(1)} + \psi^{(2)} + \dots \ , \tag{1.101}$$

gives the next order displacement (Bouchet et al., 1995), which takes into account tidal gravitational effects,

$$\nabla_{\mathbf{q}} \psi^{(2)} = \frac{1}{2} D_2(\eta) \sum_{i \neq j} \psi^{(1)}_{i,i} \psi^{(1)}_{i,j} - \psi^{(1)}_{i,j} \psi^{(1)}_{j,i} \ , \tag{1.102}$$

where D_2 is the second order growth factor (Bernardeau et al., 2002).

The structures represented in weak lensing maps from the late Universe, which are of interest given the work in this thesis, are extremely non-linear. This PT approach is restrictive as we can no longer assume we are in a perturbative regime when $\delta \ll 1$. Though we will see that the results from LPT can still play a part, we rely on simulations to model truly non-linear growth.

N-body simulations evolve a large number of particles according to the equations of motion, primarily the Euler and Poisson equations. *Particle-mesh* (PM) is a popular and simple approach where the forces (or potentials) at each time-step are calculated from a grid. In the simplest case, the grid points are assigned the total mass from all particles within the point's own unique volume (Efstathiou et al., 1985). As these particles are meant to represent a fluid, but actually have large individual masses and are not continuous in space, the forces can be set to a constant below a certain length scale. This *softening* procedure avoids large scattering-like events between particles in the simulation (Peacock, 1999).

One of the most popular software packages, GADGET (Springel, 2005), is primarily an N-body method, if we ignore its *smoothed-particle hydrodynamics*. GADGET models the non-interacting (cold) dark matter density field as a large number of collisionless particles which only interact through gravitation. This assumption neglects the effect of baryonic matter or neutrinos, which both can affect the matter evolution.

The grid-interpolation of pure PM inevitably introduces inaccuracies in the force calculation at length scales corresponding to the grid spacing. GADGET's TreePM algorithm uses both a PM, which predominantly solves for the large distance scales, and a tree method (Hernquist et al., 1991), which computes the short range forces. The PM grids the particles onto a mesh on which the Poisson equation (equation 1.85) calculates the forces at each grid position. The forces are calculated in Fourier space, achieved using a

fast-Fourier-transform (FFT).

For large volume cosmological simulations, standard methods such as GADGET can be computationally slow and memory-intensive. Even without the more complex baryon interactions, full N-body methods are often too computationally intensive to run enough simulations for certain problems. For dark matter map statistics, the simulations must simultaneously have a volume as large as a modern galaxy survey and high enough resolution to accurately resolve the peak structure. L-PICOLA (Howlett et al., 2015) is a good candidate to generate a large enough sample of simulations (rather than using the full N-body methods).

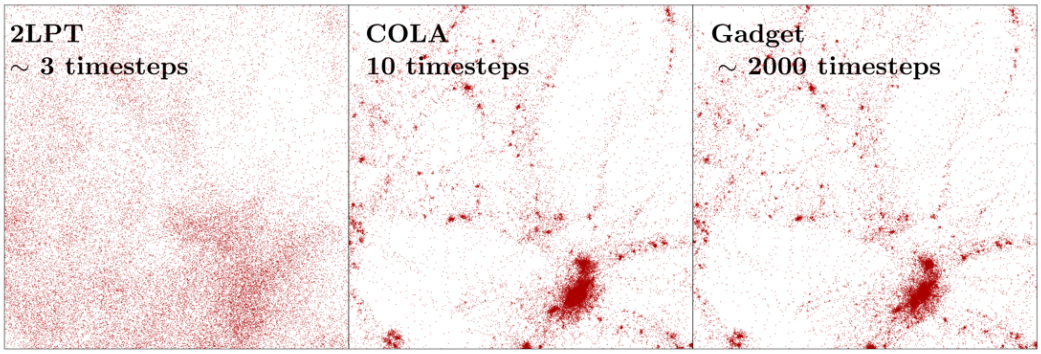

Figure 1.2: Image taken from Tassev et al. (2013): *"We show slices through three N-body simulations evolving the same initial conditions up to $z = 0$. The particles (each of mass 4.6×10^9 $M_\odot/h$) are shown as red points. Each slice is* 20 Mpc/h *on the side (the full simulation box is* 100 Mpc/h *on the side), and about* 3 Mpc/h *thick. The left panel shows the 2LPT approximation used for building mock catalogs using the PTHalos approach. Calculating the 2LPT particle positions requires an equivalent of roughly* 3 *timesteps performed by an N-body code. The middle panel shows the result obtained with our modified N-body code with as few as* 10 *timesteps. The rightmost panel shows the "true" result obtained from* GADGET-2 *after approximately* 2000 *timesteps starting with 2LPT initial conditions at* $z = 49$*."*

L-PICOLA is a light-cone implementation of the COLA algorithm, which is an N-body extension of 2LPT. The COLA method (Tassev et al., 2013) allows the large scale quasi-linear dynamics to be solved by the perturbative approach but also uses N-body methods to help in recovering the small scales. This requires far fewer timesteps to recover the non-linear and large scale dynamics than an N-body code such as GADGET, as can be seen in figure 1.2.

The L-PICOLA implementation uses an initial theoretical power spectrum, with parameters Ω_m, Ω_b (baryon density), h, σ_8, and n_s, and calculates the initial density up to a user-defined initial redshift using 2LPT.

The σ_8 parameter is the standard deviation of linear matter fluctuations in 8 Mpc h^{-1} spheres, which defines the amplitude of the power spectrum. Here h is the rescaled Hubble parameter

$$h = H_0/100 \ (\mathrm{km/s})/\mathrm{Mpc} \ . \tag{1.103}$$

At every user-defined iteration step from the initial redshift to the final, the particles are evolved according to both 2LPT and the PM method. The PM implemented in L-PICOLA accounts for the effect of 2LPT, and applies the PM as a final correction. Such evolution can be seen in figure 1.3.

The dark matter halos recovered from L-PICOLA are not as collapsed as with a full N-body. When compared with GADGET-2 simulations, Howlett et al. (2015) found that the lack of the smaller scale forces lead to "puffed out" halos. This propagates to the correlation functions, where the recovered power spectrum is very accurate up until a cut-off at a certain small length scale that depends on the specific simulation parameters. However, above this scale the power spectrum and 3-point statistics agree very well with theoretical and simulated power spectra and bispectra.

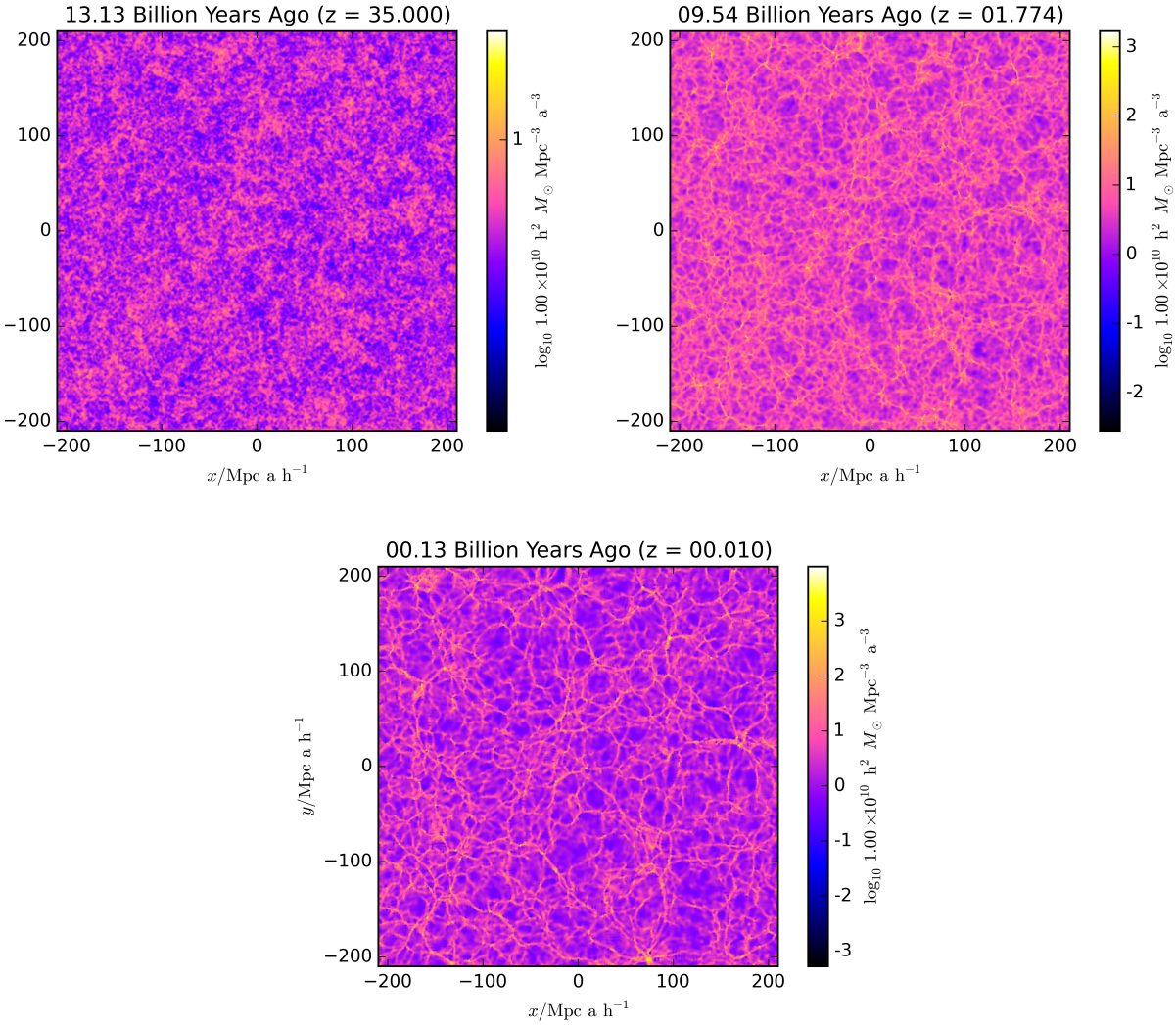

Figure 1.3: Three projected density snapshots from different timesteps of a L-PICOLA simulation using a comoving (600 Mpc) simulation box with periodic boundary conditions. A near-Gaussian density field at early times evolves to highly non-Gaussian cosmic-web distribution at late times. Density estimates were calculated from particle positions using PYNBODY (Pontzen et al., 2013).

1.4 Overview of cosmological observables

In the previous sections the main aspects of the standard model of cosmology were presented, with a main focus on results which will become relevant in the following chapters. Similarly, in this section I will show a few main ways this model has been constrained from observational data, results which will used throughout this thesis.

A brief presentation of the mechanism of the Cosmic Microwave background is outlined in section 1.4.1. I will present select results from the Planck analysis that have affected some of the analysis choices made in the work of this thesis.

In section 1.4.2, I will present how distance ladders and Type Ia supernovae can be used to constrain cosmological parameters and models, and particularly how H_0 seems to disagree with the result from the CMB. In the work described in Chapter 4, supernovae distance ladder (JLA) data will be used to demonstrate the effects of noisy simulations on parameter inference.

I will discuss how cosmological models and parameters can be constrained from N-point functions of the galaxy distribution in section 1.4.3. This has much in common with, and can be contrasted with, the main topic of this thesis, weak gravitational lensing, which is presented in section 1.4.4.

1.4.1 Cosmic Microwave Background

This extremely brief description of some of the elements of the CMB will be to allow the presentation of a few relevant results. Those which will be of particular relevance to the work in this thesis are: flatness, a dark matter component, and primordial spectrum that matches inflation.

As discussed in section 1.2, in the post-inflationary Universe the Hubble sphere expands, reintroducing modes with increasingly large length scales. Once a mode with a given length scale becomes sub-horizon, it is able to evolve once more.

In the early Universe, the baryon fluid and photon fluid were tightly coupled due to efficient Thompson scattering between the photons and electrons[12]. As such, they can be treated as one single fluid which has pressure, unlike the dark matter component which could only interact gravitationally with the baryon plasma. Solving the Boltzmann equations shows that modes in this fluid start oscillating once they are sub-horizon.

As the Universe expanded, the temperature of the baryon-photon plasma decreased.

[12] Baryons in this context include electrons, much to the confusion of particle physicists.

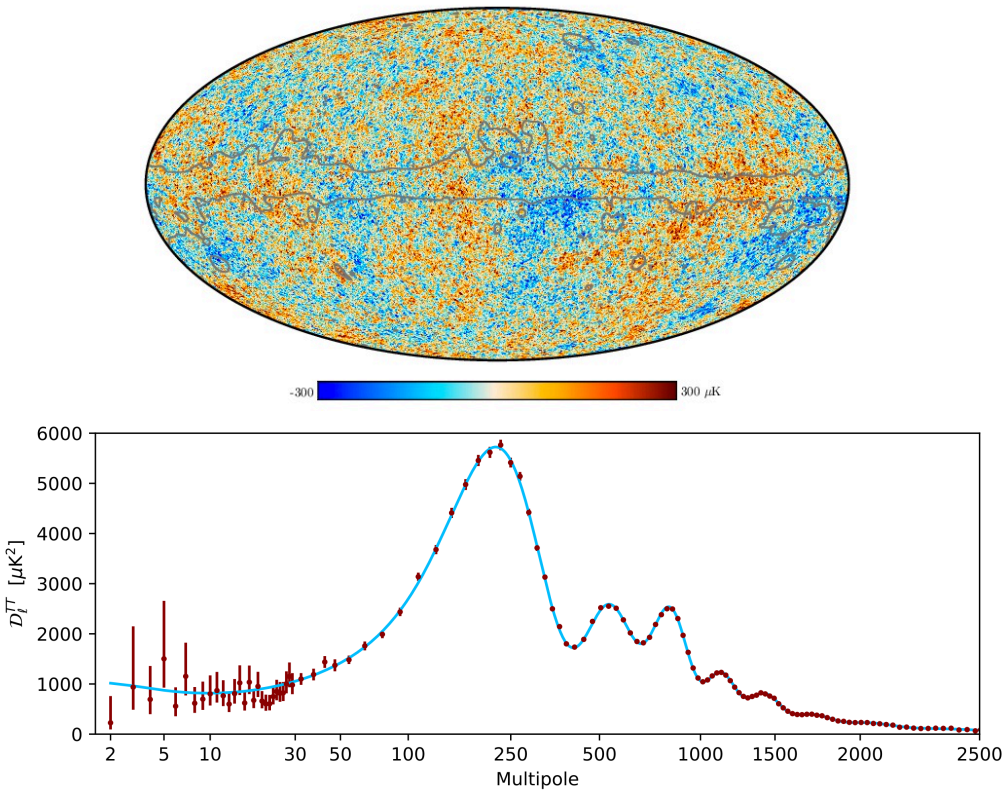

Figure 1.4: Adapted from Planck Collaboration et al. (2018b). The *upper panel* is the foreground removed (SMICA) map. The grey outline shows the mask, which covers the central Milky Way galaxy region, which has been inpainted. The *lower panel* is the foreground-subtracted and frequency-averaged angular power spectra for temperature, where $D_\ell = C_\ell \ell(\ell+1)/(2\pi)$.

Once the temperature dropped below a critical level, neutral hydrogen could form out of the plasma; this time period is known as *recombination*. At this time ($\sim 4 \times 10^4$ years after inflation), photons decoupled from electrons and free-streamed. Observations of these photons at the present day give a snapshot of this *surface of last scattering*.

These photons, now redshifted by a factor of ~ 1100, form the Cosmic Microwave Background (CMB). The discovery of the CMB by Penzias and Wilson (1965) gave spectacular evidence for the hot Big Bang model (Dicke et al., 1965). This was followed by a precise measurement of the photons' black body spectrum with a temperature of 2.725 ± 0.002 K (Fixsen et al. 1996, Mather et al. 1999).

In recent years, the angular anisotropies in the CMB temperature and polarisation have provided unique constraints on cosmological parameters and models. Figure 1.4 shows the angular power spectrum (defined in equation 1.70) for the CMB temperature fluctuations along with a reconstructed map of the underlying temperature fluctuations.

The acoustic peaks at different ℓ multipoles in the angular power spectrum can be understood in terms of the different length scales coming into the growing Hubble sphere at different times. Smaller scales in the baryon fluid, corresponding to larger ℓ, began oscillating at earlier times and so had more oscillatory periods before recombination; hence the oscillations at different ℓ scales are out of phase. At recombination the baryon fluid stopped oscillating as the pressure from radiation was lost; some modes that were at the extrema of their oscillation show enhanced power. The exact form of the temperature angular power spectrum can be calculated using linear Boltzmann codes, such as CMBFAST (Seljak and Zaldarriaga, 1996) or CAMB (Lewis et al., 2000).

As the CMB is in the relatively early Universe, non-linear structure formation has not taken place. The primordial temperature anisotropies are therefore extremely Gaussian (assuming Gaussian initial conditions from inflation). The angular power spectrum therefore contains all of the information about the temperature field (excluding the mean).

Despite this, there is additional information encoded in the linear polarisation of the CMB photons. Polarisation occurs due to scattering from quadrupole distributions of incoming photons. A polarisation map across the celestial sphere can be composed of independent linear Stokes Q and U parameters. These can be combined into rotationally invariant E and B fields, which have the benefit that tensor (gravitational wave) modes from inflation induce only B-modes and not E-modes in the primary CMB (Kamionkowski et al. 1997, Seljak and Zaldarriaga 1997, Zaldarriaga 2001, Kovac et al. 2002).

The photon distribution is not exactly a snapshot of the last-scattering surface. Secondary anisotropies, which I won't explore in detail, include inverse Compton scattering, the Sunyaev-Zel'dovich effect (Sunyaev and Zeldovich, 1980), and the late-time integrated Sachs-Wolfe effect (Sachs and Wolfe, 1967). Gravitational lensing, changing the apparent trajectories of photons, causes distortions which are a probe of the matter distribution between the time of last scattering and now (see Hu and Dodelson 2002 for review).

These following results, which are selected for their relevance to the work in this thesis, are all from the Planck Collaboration's final 2018 analysis release (Planck Collaboration et al., 2018a); Planck was an ESA space mission which took CMB temperature and polarisation measurements for 30 months between 2009 and 2013.

Flatness from the CMB has been measured to a great enough precision that global spatial curvature Ω_K will be assumed negligible in all of the work in this thesis. The combi-

nation of temperature, polarisation and lensing gives $\Omega_K = 0.0106 \pm 0.0065$, where 0.0106 is the mean of the marginal posterior probability distribution[13] and 0.0065 is the standard deviation (i.e. 1σ).

An inferred cold dark matter density of $\Omega_c h^2 = 0.1200 \pm 0.0012$ can be compared with the total matter density $\Omega_m h^2 = 0.1430 \pm 0.0011$. The factors of h^2 reflect the dependence on the Hubble parameter. This is clear evidence (using temperature, polarisation, and lensing) that the matter distribution in the Universe is primarily non-baryonic dark matter.

Though the tensor contribution to the CMB has disappointingly remained elusive, the index of the scalar primordial power spectrum n_s is well constrained. The most recent measurement from Planck gives $n_s = 0.9649 \pm 0.0042$. This statistical inference combines CMB temperature, polarisation and lensing data using a ΛCDM model to give a parameter constraint which is "8σ" away from $n_s = 1$ (Planck Collaboration et al. 2016, Planck Collaboration et al. 2018a). This is compelling evidence of a prediction of inflation that is explicitly different from the pre-inflationary theory (i.e. the Harrison-Zel'dovich-Peebles spectrum with $n_s = 1$).

1.4.2 Distance ladder

Objects with a known luminosity can be thought of as being a *standard candle*, which can be used to estimate distances (Leavitt, 1908). Type Ia supernovae (SNe Ia), which are expected to occur due to mass transfer onto a white dwarf star from a second star (Whelan and Iben, 1973), are often termed *standardisable candles*. Their observed peak luminosity can be related to the luminosity distance with corrections that include the observed colour, light-curve shape, and mass of the host galaxy (e.g. Betoule et al. 2014).

By combining the luminosity distance and the observed redshift of the host galaxy, SNe Ia were famously used to great effect by Perlmutter et al. (1999) and Riess et al. (1998) to provide strong evidence that $\Omega_\Lambda > 0$, work which led to the 2011 Nobel Prize in Physics.

The *Hubble tension* is a name that has been given to the apparently discrepant results for the value of H_0, Hubble parameter at $z = 0$, from the CMB (e.g. Planck) and so-called "direct measures" using SNe Ia in a *distance ladder*. The final Planck analysis (Planck Collaboration et al., 2018a) gives a "low" result for the reduced Hubble constant as $h = 0.674 \pm 0.005$. Based on data which included the Hubble Space Telescope (HST) SH0ES[14] programme, Riess et al. (2019) present a "high" result of $h = 0.74 \pm 0.014$.

[13] See section 1.5 for the definition of marginal posterior.

[14] Supernovae, H0, for the Equation of State of dark energy

In order to make such a local measurement of H_0 from supernovae alone, the intrinsic magnitude of the SNe Ia would need to be known. The technique used, which is sometimes referred to as *calibrating* the distance ladder, joins up local measurements of absolute distances to the low redshift tip of SNe Ia observed magnitude-redshift curve. In broad terms[15] the ladder connects: (1) local geometric distance measurements (including parallax), (2) pulsating cepheids variable stars using an assumed period-luminosity relation (Leavitt and Pickering 1912, Eddington 1917), and (3) observed SNe Ia.

The benefit of the local distance ladder, rather than the full ΛCDM Planck analysis, is that a cosmological model does not need to be assumed, and $H(z)$ can be modelled as a flexible function in terms of z (e.g. Taylor expansion, Visser 2004) where $H_0 = H(z=0)$.

It is possible to have an *inverse distance ladder*, where the absolute distance is calibrated at high redshift using only the sound horizon from the CMB and having a flexible form for $H(z)$ (e.g. Lemos et al. 2019). Combining 207 SNe Ia from DES, 122 low redshift SNe Ia from external data sources, BAO measurements from external spectrocopic data, and the sound horizon result from Planck, Macaulay et al. (2019) infer a value of $h = 0.678 \pm 0.013$. As this result used the sound horizon from the CMB at high redshift, it is perhaps not too surprising that this result agrees with Planck, rather than the local distance ladder.

1.4.3 Galaxy clustering

The distribution of galaxies in the Universe is expected to trace the underlying matter distribution. Using this assumption, galaxy surveys are able constrain cosmological parameters and models using observed N-point correlation functions or the N-point spectra (in harmonic space).

The link between the bright observable galaxies and the dark underlying density field is encapsulated in *galaxy biasing*. We think of galaxies as biased tracers, as they do not exactly trace the underlying matter distribution. This is clear enough when one considers that different types of galaxies (e.g. spiral or elliptical galaxies) are observed to cluster differently (Dekel and Lahav, 1999).

Initial work (Kaiser 1984, Bardeen et al. 1986) assumed that galaxies would form at growing peaks in the primordial Gaussian density field, where the potential well is deepest. In the linear regime, this leads to a galaxy overdensity given by

[15] See Riess et al. 2019 for details, additional routes and "half-steps" up the ladder.

$$\delta_g(\mathbf{x}) = b\ \delta(\mathbf{x}) \quad , \tag{1.104}$$

where, in the most general case, b is the bias parameter for a certain type of galaxy at a given time or redshift.

This then gives a simple relation for the power spectrum of the galaxy distribution

$$P_g(k) = b^2\ P(k) \quad , \tag{1.105}$$

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

One clearly might desire to expand on the simple assumptions made up to this point. For example, adding a redshift dependence to the bias parameter $b(z)$ has been proposed (Nusser and Davis 1994, Fry 1996, Tegmark and Peebles 1998). Indeed, choosing to ignore redshift evolution has been shown to affect the inferences made about the dark energy equation of state w (Clerkin et al., 2015).

The form of equation 1.105 might suggest a perturbation theory point of view. One could consider adding higher-order (non-linear) biasing terms. It was suggested quite early on that linear bias was incompatible with data (Gaztanaga, 1992). With this view, one could consider expanding to increasingly higher order with operators combining perturbed density, the potential, and scale k (see Desjacques et al. 2018 for a recent review). Bias with scale k variation, $b(k)$, would be a particularly difficult problem, as it can introduce effects that are indistinguishable from the power spectrum $P(k)$.

It is typical to choose some flexible model and treat any free bias parameters as *nuisance parameters*. These are free to be constrained by the data and are included in the statistical inference, but are of little eventual interest. Whatever the bias model, it provides a link between the N-point functions of galaxy overdensity and the N-point functions of the matter overdensity.

The most accurate strategy to map the distributions of the galaxies is to measure their *spectral energy distribution* (SED), the frequency distribution of observed photons per galaxy. The various absorption and emission features, which are known in the rest-frame, are redshifted. This allows a spectroscopic experiment like the Sloan Digital Sky Survey (SDSS, Abolfathi et al. 2018) to map galaxies with three coordinates: z, right ascension (RA), declination (DEC).

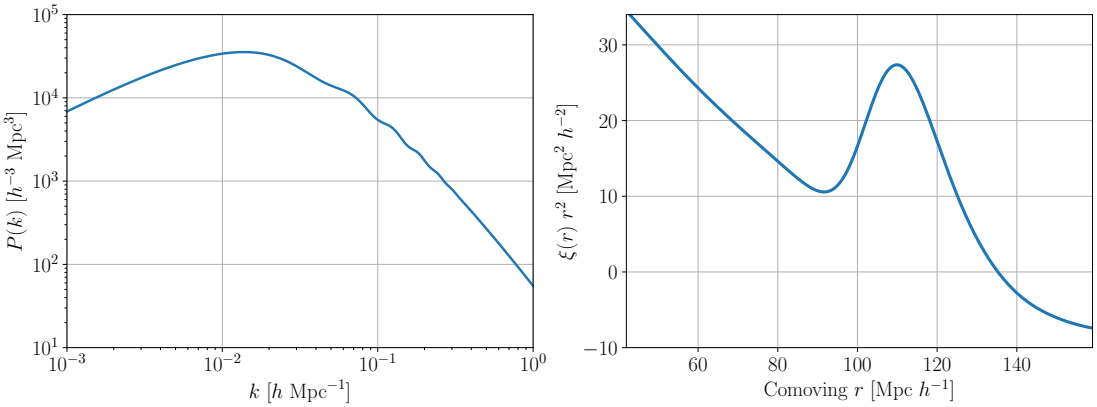

Figure 1.5: *Left panel*: the matter power spectrum at $z = 0$ using linear theory and excluding the effects of redshift space. *Right panel*: the corresponding two-point correlation function. Both show the baryonic acoustic oscillation (BAO) feature. The chosen cosmological parameters match the MICE simulations (e.g. Fosalba et al. 2015) and calculations were performed with CLASS and NBODYKIT (Lesgourgues 2011, Audren and Lesgourgues 2011, Hand et al. 2018).

One feature of observing the apparent redshift of a galaxy with high-quality spectrometry is *redshift space distortion*. The z coordinate only corresponds to a unique distance in the absence of peculiar velocities. The quasi-random paths of galaxies within a halo leads to a Doppler broadening, elongating the redshift distribution, which is grandly known as the *Fingers-of-God* effect (Jackson, 1972). A second *Kaiser effect* (1987) flattens the structures in the z direction due to infalling galaxies.

Figure 1.5 shows the linear theory matter power spectrum at $z = 0$ and the corresponding two-point correlation function $\xi(r)$ evaluated using CLASS and NBODYKIT (Lesgourgues 2011, Audren and Lesgourgues 2011, Hand et al. 2018). The baryon acoustic oscillation (BAO) feature can be clearly seen. The comoving length scale of the oscillations corresponds to the comoving sound horizon at recombination (c.f. section 1.4.1), which has a Planck value of $r_d = 147.18 \pm 0.29$ Mpc (Planck Collaboration et al., 2018a). This BAO oscillation corresponds to a sharp feature at this comoving length scale in the configuration space (i.e. not harmonic space) correlation function. This can be used as a *standard ruler* across different redshifts, providing particularly good constraints on the cosmological parameters when combined with CMB results (e.g. Baryon Oscillation Spectroscopic Survey, Alam et al. 2017).

The next generation spectroscopic survey DESI (DESI Collaboration et al., 2016), based at the Mayall 4-meter telescope at Kitt Peak National Observatory, will take spec-

tra and determine the redshift of 35 million galaxies. In comparison, both the upcoming space-based ESA Euclid mission (Amendola et al., 2016) and the ground-based Large Synoptic Survey Telescope (LSST, LSST Science Collaboration et al. 2009) will observe approximately 1 billion galaxies. This gain in the number of galaxies is at the expense of the quality of the redshift measurement. Both Euclid and LSST will primarily[16] be *photometric* surveys.

A photometric redshift measurement relies on measuring a galaxy's flux in a relatively low number of colour filters, measurements which are then used to estimate the redshift of a galaxy (or the redshift distribution of galaxies). The ongoing Dark Energy Survey (DES) experiment, which will feature often in the work of this thesis, is expected to have observed 300 million galaxies in such a way by the completion of its analysis. DES uses a wide-field camera (DECam) on the 4m Blanco Telescope in Chile to image 5000 sq deg of the sky in five filters (Dark Energy Survey Collaboration et al., 2016).

1.4.4 Weak gravitational lensing

Light propagating through non-uniformities in the metric between a distant source galaxy and an observer leaves a distortion imprinted on the image of the galaxy. This distortion, though it may be small, is controlled by the distribution of matter along the path of the photon. Given enough galaxies across the sky and with a suitable distribution in radial distance, it is possible to infer properties of the matter distribution.

The power of gravitational lensing is that all matter contributes to the effect. Unlike with galaxy counts (section 1.4.3) where we were reduced to proposing galaxy biasing schemes to infer the dark matter distribution, weak lensing can "see" the dark matter directly.

In this section I will present some standard results from weak gravitational lensing, primarily based on Bartelmann and Schneider (2001), Castro et al. (2005), and Kilbinger (2015). Throughout it will be assumed that we are in the weak field limit, such that the potential is small $\Phi \ll c^2$, as are the angles through which the light's path is deviated. To make the weak field explicit, the factors of the speed of light c will be reintroduced at first. The assumption from section 1.3.1, that $\Phi = \Psi$, will carry through here.

To begin lensing, we can again consider scalar perturbations to the FLRW metric (equation 1.73), but dropping the scale factor a as we only care about local lensing[17]. This

[16] Euclid will actually have an additional grism-based spectrograph.

[17] We can also think that $\delta\eta = a\delta t$, so the argument from Fermat's principle still applies.

gives

$$\mathrm{d}s^2 = \left(1+\frac{2\Phi}{c^2}\right)c^2\mathrm{d}t^2 - \left(1-\frac{2\Phi}{c^2}\right)\mathrm{d}r^2 \ , \tag{1.106}$$

where we are using the proper (non-conformal) time coordinate t and are only considering the radial spatial coordinate r. To first order

$$c\frac{\mathrm{d}t}{\mathrm{d}r} = 1-\frac{2\Phi}{c^2} \ , \tag{1.107}$$

where the right hand side is acting like a refractive index term in classical lensing.

Similarly to classical lensing, we can use Fermat's principle for $\delta t = 0$. Solving the calculus of variations (Blandford and Narayan 1986, Schneider 1985) gives a *deflection angle* $\hat{\vec{\alpha}}$,

$$\hat{\vec{\alpha}} = -\frac{2}{c^2}\int \vec{\nabla}_\perp \Phi \, \mathbf{d}r \ , \tag{1.108}$$

which can be thought of as the scattering angle of the incoming photon. The gradient is with respect to the physical coordinates of the perpendicular plane at the lens.

We consider the unperturbed perpendicular distance between two incoming light rays $\vec{x}^0$ at some comoving distance ω, which can be given in terms of the observed angular separation $\vec{\theta}$

$$\vec{x}^0 = f_K(\omega)\vec{\theta} \ . \tag{1.109}$$

If a light ray is perturbed by $\mathrm{d}\hat{\vec{\alpha}}$, the perpendicular distance will change by

$$\mathrm{d}\vec{x} = f_K(\omega)\mathrm{d}\hat{\vec{\alpha}} \ . \tag{1.110}$$

The physical separation between two close light rays will therefore be given by

$$\vec{x}(\omega) = f_K(\omega)\vec{\theta} - \frac{2}{c^2}\int_0^\omega f_K(\omega-\omega')\left(\vec{\nabla}_\perp\Phi^{(1)}(\vec{x}(\omega'),\omega') - \vec{\nabla}_\perp\Phi^{(2)}(\omega')\right)\mathrm{d}\omega' \ . \tag{1.111}$$

The angle in the absence of any lensing $\vec{\beta}$, which can be thought of as a true angle in the classical lensing analogy, can be written as

$$\vec{\beta} = \vec{\theta} - \vec{\alpha} \ , \tag{1.112}$$

where $\vec{\alpha}$ is called the *reduced deflection angle* and corresponds to

$$\vec{\alpha} = \frac{2}{c^2}\int_0^{\omega} \frac{f_K(\omega-\omega')}{f_K(\omega)}\Big(\vec{\nabla}_{\perp}\Phi^{(1)}(\vec{x}(\omega'),\omega') - \vec{\nabla}_{\perp}\Phi^{(2)}(\omega')\Big)\mathrm{d}\omega' \ . \tag{1.113}$$

The Born approximation assumes that the observed angle to a point, $\vec{\theta}$, deviates only a small amount from the true angle $\vec{\beta}$, so the change in distance of the photon's path is negligible. To first order, this corresponds to substituting $\Phi^{(1)}(\vec{x}(\omega'),\omega') = \Phi^{(1)}(f_K(\omega)\vec{\theta},\omega')$ in the integral.

We consider the lensing matrix, a Jacobian which transforms between observed angular coordinates and (hypothetical) true angular coordinates

$$\mathscr{A}_{ij} = \frac{\partial\beta_i}{\partial\theta_j} = \delta_{ij} - \frac{\partial\alpha_i}{\partial\theta_j} \ , \tag{1.114}$$

If we define a *lensing potential* as

$$\psi_l(\theta,\omega) = \frac{2}{c^2}\int_0^{\omega}\mathrm{d}\omega'\Big[\frac{f_K(\omega)-f_K(\omega)'}{f_K(\omega)f_K(\omega)'}\Big]\Phi(f_K(\omega)'\theta,\omega') \ , \tag{1.115}$$

the lensing matrix can be rewritten as

$$\mathscr{A}_{ij} = \delta_{ij} - \partial_i\partial_j\psi_l \ . \tag{1.116}$$

This now gives us the relationship between scalar perturbations to the metric, which in our weak-field limit and chosen gauge corresponds to the Newtonian potential, and the angular distortion of light bundles. That is, we can see how images of distant galaxies are distorted by fluctuations in the foreground matter distribution.

The $\mathscr{A}_{ij}$ matrix can be decomposed into the functions $\kappa(\vec{\theta})$ and $\gamma(\vec{\theta}) = \gamma_1 + i\gamma_2$, by

$$\begin{aligned} \mathscr{A}_{ij} &= \delta_{ij} - \frac{\partial^2\psi_l(\vec{\theta})}{\partial\theta_i\partial\theta_j} \\ \boldsymbol{\mathcal{A}} &= \begin{pmatrix} 1-\kappa-\gamma_1 & -\gamma_2 \\ -\gamma_2 & 1-\kappa+\gamma_1 \end{pmatrix} \ . \end{aligned} \tag{1.117}$$

Through the definition of the lensing potential and the Poisson equation (1.85), the convergence can be expressed as an integral of the overdensity along the line of sight,

$$\kappa(\vec{\theta},\omega) = \frac{3H_0^2\Omega_M}{2}\int_0^{\omega} d\omega' \frac{\omega'(\omega-\omega')}{\omega}\frac{\delta(\vec{\theta},\omega)}{a(\omega')} \quad . \tag{1.118}$$

From this we would be able to calculate quantities such as the angular power spectrum C_ℓ^κ from the matter power spectrum evolving with comoving distance $P(k,\omega)$.

By inspection of equation 1.117, we can see that the shear and convergence are related to derivatives of the lensing potential

$$\begin{aligned} \kappa &= \frac{1}{2}\left(\partial_1\partial_1 + \partial_2\partial_2\right)\psi_l = \frac{1}{2}\nabla_\theta^2\psi_l \\ \gamma_1 &= \frac{1}{2}\left(\partial_1\partial_1 - \partial_2\partial_2\right)\psi_l \\ \gamma_2 &= \partial_1\partial_2\psi_l \quad , \end{aligned} \tag{1.119}$$

and therefore we can transform between γ and κ (up to an integration constant). As shear is observable under certain definitions and convergence is the projected matter density (equation 1.118), it is possible to directly map dark matter (Kaiser and Squires, 1993). This will be discussed thoroughly in Chapters 2 and 5.

For the observable ellipticity quantities in lensing, we can define the *reduced shear* as

$$g_i = \frac{\gamma_i}{1+\kappa} \tag{1.120}$$

and rewrite equation 1.117 as

$$\mathcal{A} = (1-\kappa)\begin{pmatrix} 1-g_1 & -g_2 \\ -g_2 & 1+g_1 \end{pmatrix} \quad . \tag{1.121}$$

If we consider an image with elliptical isophotes (contours of equal brightness), at angle ϕ and major-to-minor axis ratio a/b, then the ellipticity (Seitz and Schneider, 1997) is given by

$$\varepsilon = \frac{(a-b)}{(a+b)}\exp(2i\phi) = \frac{g+\varepsilon_S}{1+g*\varepsilon_S} \tag{1.122}$$

for $|g| \leq 1$, and where the unlensed *source ellipticity* ε_S has mean zero. In the weak lensing

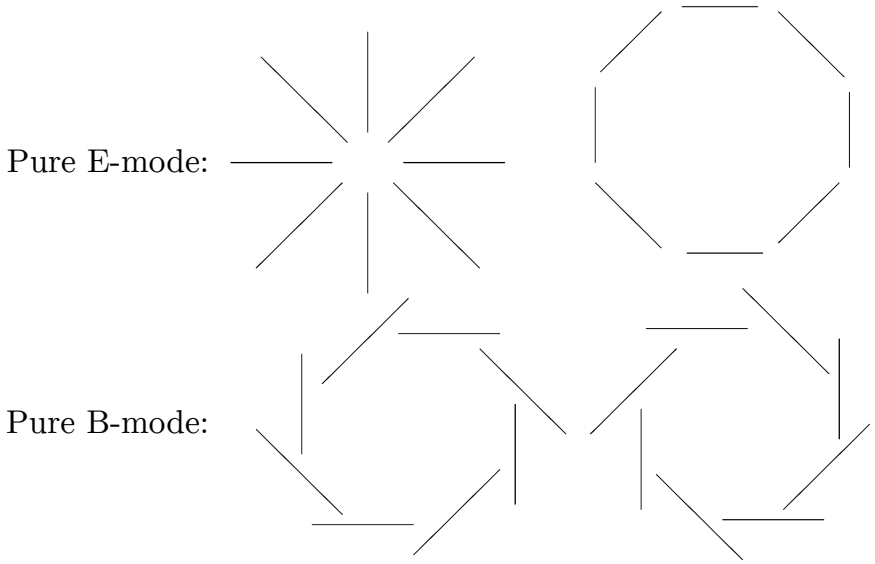

Figure 1.6: Headless vectors represent the shear field, where the direction corresponds to the major axis of an elliptical isophote (adapted from Cabella and Kamionkowski 2004). Gravitational lensing in the weak field limit generates no B-mode patterns.

limit ($\gamma \ll 1$) the observed ellipticity for a galaxy becomes analogous to a noisy measurement of the shear: $\varepsilon = \gamma + \varepsilon_S$.

Unfortunately for weak lensing, $\sigma_{\varepsilon_S} \sim \mathscr{O}(10^{-1})$, whereas $\sigma_\gamma \sim \mathscr{O}(10^{-2})$, as we will see in the following chapters. Weak lensing therefore benefits from a large angular density of observed galaxies, which will essentially average away the uncorrelated source ellipticity terms.

Similarly to polarisation in the CMB (section 1.4.1), the shear field can be decomposed into a curl-free E-mode and a divergence-free B-mode. As the convergence is a real scalar field (equation 1.118) in the weak lensing limit (based on equations 1.119) the lensing shear field should be free of B-modes. Higher order terms (e.g. Krause and Hirata 2010), biases and noise can all in fact contribute B-mode components.

Figure 1.6 shows the E- and B-mode patterns. The headless vectors represent the orientation and magnitude of the lensing induced shear. The π (180 deg) rotational symmetry of the complex shear, indicated by the factor of $\exp(2i\phi)$ in the ellipticity, demonstrates that the shear is a *spin 2 field*. If we wish to go beyond the small-angle, flat-sky approximation we made in the derivation above, on the sphere there is a natural formalism of spin 2-weighted spherical harmonics (Newman and Penrose, 1966).

In this (often formidable) framework one can derive the E- and B-mode angular power spectra with less recourse to simplifying assumptions. However, as the work in this thesis uses data which covers a small fraction of the celestial sphere (high multipole ℓ), we will only ever work with the results with small-angle approximations.

1.5 Inference

1.5.1 Bayesian background

The posthumous presentation of Thomas Bayes' *An Essay towards solving a Problem in the Doctrine of Chances* at the Royal Society in 1763, sent with the "abstract" opening

> "*Dear Sir, I Now send you an essay which I have found among the papers of our deceased friend Mr. Bayes, and which, in my opinion, has great merit, and well deserves to be preserved.*" Richard Price 1763,

presents a surprising introduction to a great moment of scientific and epistemological history. Despite brilliant work[18] by Laplace in the following decades (Laplace, 1840), the "Doctrine of Chances" seems to have been barely considered until the 20th Century.

The Bayesian approach, as it has come to be known, provides clear and precise relationships between unknown quantities and observations. For many people who were presented to "Statistics" as a box of byzantine tools to justify this or that after an experiment, a unified framework to quantify belief and uncertainty comes as a revelation. In many ways, it is similar to discovering how one can replace the bookkeeping of Newtonian vectors and conserved quantities with the elegance of Lagrangian mechanics and Noether's theorem (Noether, 1918).

Here, there is maybe a warning from history. The reemergence of Bayesian methods in the late 20th Century has probably more to do with our improved ability to calculate and compute, than a spontaneous shift in philosophy. With the growth of machine learning, and particularly deep learning, in this so-called *Artificial Intelligence (AI) revolution*, the flood of new techniques might feel like a return to messy "Statistics".

For some problems where the naive Bayesian approach might be impractical and suffer from the *curse of dimensionality*, deep learning solutions have appeared which seem miraculous, but perhaps also mysterious and lacking in principled interpretation. This may be very well in some fields, but not for cosmologists. We want to know exactly how much we know about the Universe. We have been gifted a challenge, to connect AI methods with our Bayesian principles.

[18] who was perhaps even unaware of Bayes

1.5.2 Bayes' theorem

To introduce Bayes' theorem, following Sivia and Skilling (2006), we can begin by considering a system with a single proposition or outcome A (e.g. rolling a five with dice). The probability of A and the probability of not A (denoted $\bar{A}$) are defined such that

$$P(A|\mathscr{M}) + P(\bar{A}|\mathscr{M}) = 1 \ . \tag{1.123}$$

This is known as the *sum rule*. Here $\mathscr{M}$ is the *model*, in which one might include any information about the system. Certainty about a given outcome would be represented as $P(A|\mathscr{M}) = 1$.

If we consider a second outcome B, we can say that the probability of both A and B occurring is the same as the probability of A given the knowledge B has occurred multiplied by the probability B occurs. This is known as the *product rule*

$$P(A,B|\mathscr{M}) = P(A|B,\mathscr{M})P(B|\mathscr{M}) \tag{1.124}$$

and is, of course, symmetric

$$P(A,B|\mathscr{M}) = P(B|A,\mathscr{M}) \times P(A|\mathscr{M}) \ . \tag{1.125}$$

These rules of probability were presented by Cox (1946) as "*credited by common sense with a wider validity than can be established by deducing them from the frequency definition of probability*". This contrasts with the *frequentist* definition, where probability is defined as a fraction (or frequency) of outcomes. The few requirements used, broadly that (1) knowing something about A tells us something about the converse $\bar{A}$ and (2) knowing about A and then knowing about B given that knowledge is sufficient to know about $A\&B$, lead to the product and sum rule[19].

As a direct result of these rules we arrive at Bayes' theorem

$$P(A|B) = \frac{P(B|A)\ P(A)}{P(B)} \ . \tag{1.126}$$

As a further result we arrive at the *marginal probability*

[19] after suitable definitions, including *transitivity* of belief

$$P(B|\mathscr{M}) = \sum_{A_i} P(A_i, B|\mathscr{M}) \tag{1.127}$$

for a complete set of A_i outcomes.

For typical problems in cosmology we can consider a set of parameters θ, where the elements may be continuous (rather than discrete A_i). This might be unknown parameters of our model $\mathscr{M}$, for example $\theta = \{\Omega_m, \sigma_8\}$. We may have observed some data encapsulated in a vector $\mathbf{d}$. Bayes' theorem gives us

$$P(\theta|\mathbf{d}, \mathscr{M}) = \frac{P(\mathbf{d}|\theta, \mathscr{M})\, P(\theta|\mathscr{M})}{P(\mathbf{d}|\mathscr{M})} \quad . \tag{1.128}$$

The left hand side of this equation is the *posterior probability distribution*; it quantifies the (*a posteriori*) belief in parameters θ having observed some data. In the switch to continuous parameters, this distribution became a *density*, meaning that a probability element would be given by $P(\theta|\ldots)\mathrm{d}^{N_\theta}\theta$, where N_θ is the dimension of the vector θ. This posterior is a *probability density function* (PDF) in N_θ dimensions.

The first factor in the numerator on the right hand side of equation 1.128 is the *likelihood distribution* $P(\mathbf{d}|\theta, \mathscr{M})$. This is a PDF in terms of $\mathbf{d}$ (meaning integrating over $\mathbf{d}$ would be unity), and gives the probability of observing some data, given some parameters and a model. This simultaneously encapsulates the relationship between the data and parameters in terms of causal predictions (e.g. the theoretical power spectrum for a given set of parameters) and the properties of random contribution (e.g. measurement noise).

The second factor in the numerator is the *prior probability distribution* $P(\theta|\mathscr{M})$. This must represent the state of belief in the unknown parameters prior to any influence by the new data $\mathbf{d}$. Other datasets may be included in the prior, as long as they are independent of the new data. If we combine datasets that are not independent, then a likelihood that includes both datasets must be used, e.g. $P(\mathbf{d}^{(1)}, \mathbf{d}^{(2)}|\theta, \mathscr{M})$.

The denominator in equation 1.128 is the *model evidence*, which will be a single number for a given set of data. The probability density of the data given the model may appear to be just a factor to ensure that the posterior normalises to unity. However, if the two models (α and β) are considered, its role as an *evidence* becomes clear

$$\frac{P(\mathbf{d}|\mathscr{M}^\alpha)}{P(\mathbf{d}|\mathscr{M}^\beta)} = \frac{P(\mathscr{M}^\alpha|\mathbf{d})}{P(\mathscr{M}^\beta|\mathbf{d})} \times \frac{P(\mathscr{M}^\beta)}{P(\mathscr{M}^\alpha)} \quad . \tag{1.129}$$

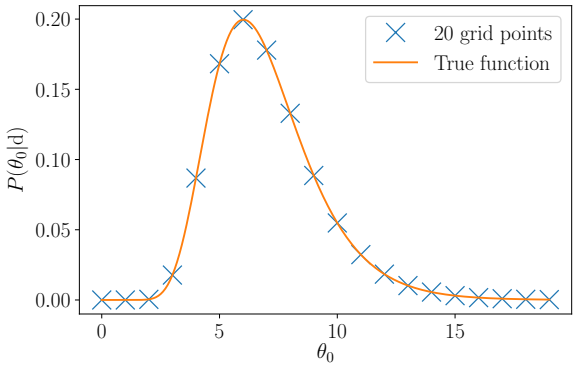

Figure 1.7: Evaluation at 20 grid positions of an underlying posterior probability distribution. For low dimensional parameter spaces, this can be the most efficient approach.

With no prior belief in either model α or β, the second factor is one, so the ratio of evidences becomes an ratio of probability for the models given the data. This is an odds ratio for different models.

The evidence is also called the *marginal likelihood*, as it can be calculated using the continuous form of the marginalisation equation 1.127 and the likelihood,

$$P(\mathbf{d}|\mathscr{M}) = \int P(\mathbf{d}|\boldsymbol{\theta},\mathscr{M})\, P(\boldsymbol{\theta}|\mathscr{M})\, \mathrm{d}^{N_\theta}\boldsymbol{\theta} \quad . \tag{1.130}$$

The prior in this integral, the second factor, can affect the evidence value as much as the likelihood here. Even if the prior is a uniform distribution, changing its range can give different values of the evidence, so care is often needed.

1.5.3 Calculation

For many problems in cosmology (see Trotta 2008 and Hobson et al. 2010 for review), the dimension of $\boldsymbol{\theta}$ will be greater than six. Already, this is quite a high dimensional problem for brute-force calculation. In figure 1.7, 20 evaluations of the posterior distribution (within a prior range) give a relatively good description of the distribution. Even if this distribution continued to be nicely behaved in higher dimension, with only five more parameters, the number of evaluations of the posterior distribution would be $20^6 = 6.4 \times 10^7$.

Markov Chain Monte Carlo (MCMC) methods provide a solution to this impractical scaling. In Chapter 3 the target posterior distribution will have dimension $N_\theta > 10^3$, where no amount of computer time or memory would justify the brute-force evaluation. MCMC methods randomly generate *samples* defined at points in parameter space, such that the

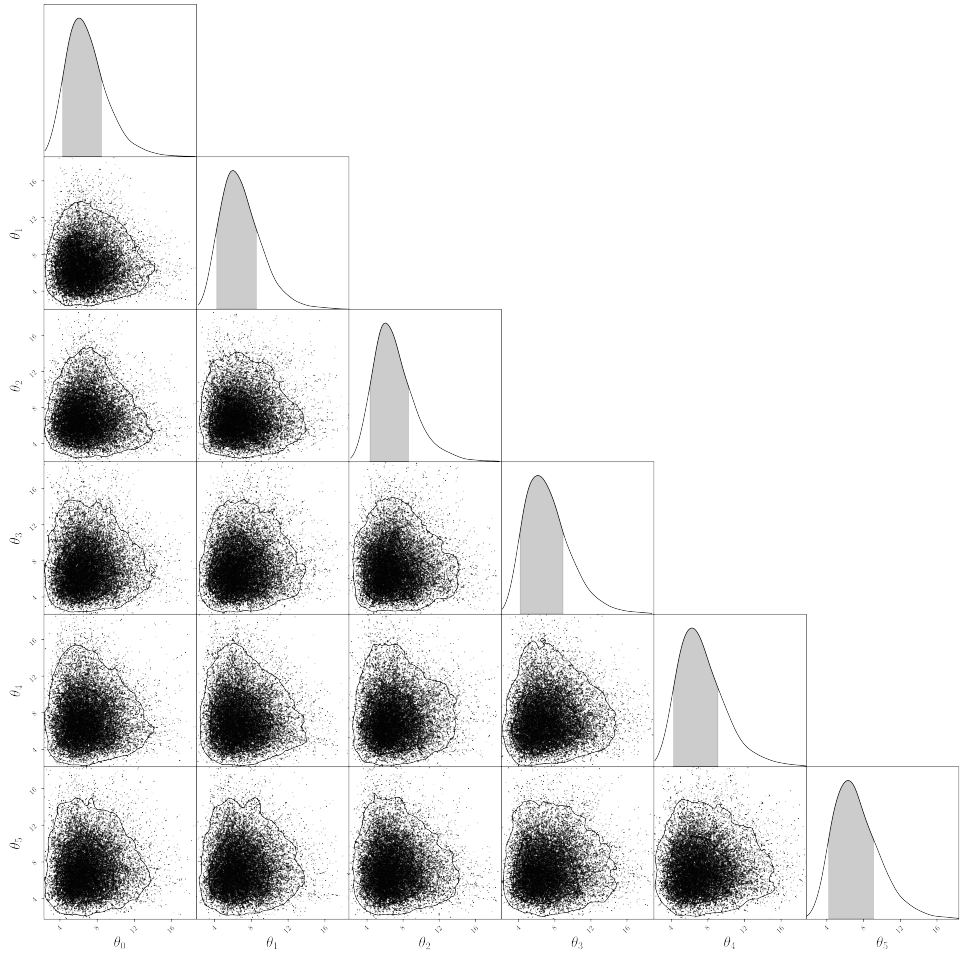

Figure 1.8: Evaluation of six-dimensional toy-model posterior using Markov Chain Monte Carlo (MCMC). The two-dimensional projections show contours of the estimated two-dimensional marginal distributions. The single one-dimensional marginal distributions are also shown.

density of the samples in parameter space match the target distribution. With sufficient samples, the underlying distribution, or properties of it, can be estimated.

The first such method by Metropolis et al. (1953) has many of the features shared by subsequent MCMC algorithms. As a Markov chain, the next sample added to the chain depends only on the current position. From a given position in parameter space at a given time i, a new point is chosen randomly at the next step $i+1$ according to some criteria (e.g. the $P(\theta_{i+1})/P(\theta_i)$ ratio) that ensures the final distribution matches the target.

Figure 1.8 shows the result of such a procedure. The marginal distributions (posterior distributions of a parameter subset following equation 1.127) are simple to calculate using the samples. The density of the samples in one parameter's dimension corresponds to the marginal distribution of that parameter.

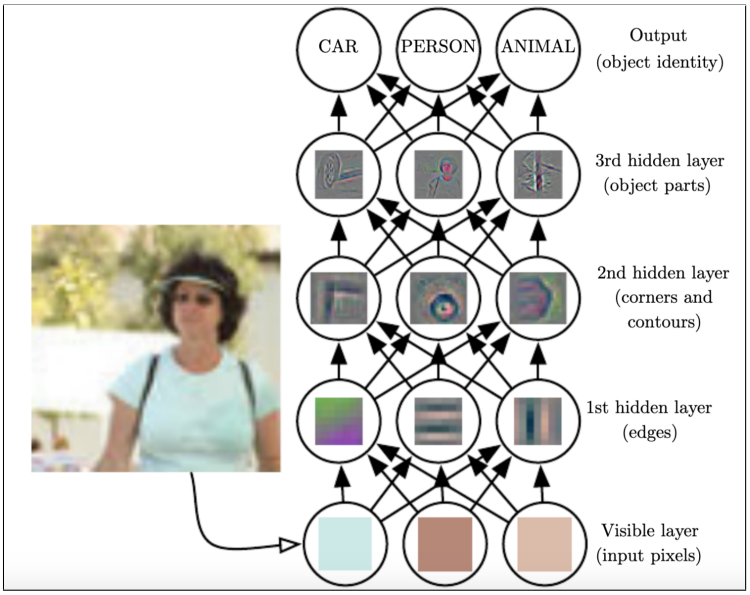

Figure 1.9: Image reproduced from Goodfellow et al. (2016), which was adapted from images from Zeiler and Fergus (2014). From Goodfellow et al. (2016): "*The input is presented at the visible layer, so named because it contains the variables that we are able to observe. Then a series of hidden layers extracts increasingly abstract features from the image.*"

As will be discussed in Chapter 3, knowing how many samples is enough and whether the distribution has properly been explored by the chain are important for interpreting the final result. Many large experiments currently releasing posterior distributions of parameters often omit to either report or perform these vital tests.

An extension of the simple scheme represented by equation 1.128 could be to use *multilevel* or *hierarchical models* (Gelman et al., 2006). If a statistical model has a number of variables, some of which depend on others but not all, their conditional dependencies can be represented as a *directed acyclic graph* (DAG). This is described as a graph in the mathematical sense of a network, and this field of *Bayesian networks* is included in the study of *probabilistic graphical models* (Koller and Friedman, 2009).

In Chapter 4, such a probabilistic model will be presented, where, because of the chosen priors, the marginal distribution of interest can be calculated analytically. However, this is not always so simple, and for certain models even recourse to sampling methods will not be computationally efficient. In this case, one may be able to approximate a target marginal distribution using variational inference methods (see Bishop 2011 for overview). These often benefit from optimization techniques shared with machine learning.

1.5.4 Learning

Deep learning is a branch of *machine learning*[20] whose methods generally solve statistical problems by finding complex[21] representations. A deep method is one with a large hierarchy of simple features that build to something with more complexity.

Supervised learning methods use training data to approximate some unknown function. It can be thought of as a high-dimensional interpolation or inference problem (Mallat, 2016). As will be reintroduced in Chapter 5 in the context of weak lensing, when using such methods, we hope that some unknown function $f(\mathbf{d})$ exists that takes observed data $\mathbf{d}$ as input and returns some unknown quantity Q. This $\mathbf{d}$ data may be pixels from an image and Q may be the name of a person in that image. The data could also be a noisy image and the unknown quantity might be the clean image.

We do not know this hypothetical function, but we seek an approximation $\mathscr{F}_\Theta$ to the function

$$\hat{Q} = \mathscr{F}_\Theta(\mathbf{d}) \ , \tag{1.131}$$

where the parameters of the function Θ are to be learned (Goodfellow et al., 2016). We learn these parameters by minimising a cost function $J(\Theta)$ evaluated with a set of training data which consists of pairs of $\mathbf{d}$ and known Q.

Convolution neural networks (CNN) are neural network architectures introduced by LeCun et al. (1990). These methods approximate the unknown function by applying repeated layers of convolution and non-linearity to input data (see Chapter 5 for more details). They have been shown to be particularly suited to image data, or any data with translation invariant features.

Figure 1.9 is a cartoon example showing how progressively more layers create complex features which correspond to structures in the image. This can be compared to transformations such as Fourier or wavelets (Starck et al., 2015), where the decomposition of an image gives coefficients corresponding to a fixed set of basis functions. It can also be compared with the approach where one chooses the important features in data which then become the only input to any machine learning. With the CNN, the series of convolutional filters and non-linear functions which act on the data are flexible and, after training, can correspond to

[20] which itself is a branch of AI

[21] "Complex" meaning in terms of complexity, not imaginary numbers.

features most relevant for a given problem.

Cosmology has now enjoyed more than a twenty-five year history of supervised learning, including examples of object classification (Connolly et al. 1995, Lahav et al. 1995, Weir et al. 1995), redshift estimation (Firth et al. 2003, Collister and Lahav 2004), and fast emulation of theoretical predictions (Habib et al., 2007).

In Chapter 5, I will discuss some interpretation of the outputs from deep learning in a Bayesian framework, specifically the relationship between the cost function J and Bayesian point estimates. In Chapter 6, I will discuss how combining outputs from machine learning methods with likelihood-free inference techniques can provide posterior distributions, rather than just point estimates, of unknown parameters. A promising future for statistics from dark matter maps might involve the combination of observed data with simulated training data, deep learning methods, and Bayesian uncertainties.

Chapter 2

Improving DES SV weak lensing maps with Gaussian and sparsity priors

This Chapter is based on Jeffrey et al. (DES Collab.) 2018: *Improving weak lensing mass map reconstructions using Gaussian and sparsity priors: application to DES SV.* Monthly Notices of the Royal Astronomical Society 479.3 (2018): 2871-2888.

Mapping the underlying density field, including non-visible dark matter, using weak gravitational lensing measurements is now a standard tool in cosmology. Due to its importance to the science results of current and upcoming surveys, the quality of the convergence reconstruction methods should be well understood. We compare three methods: Kaiser-Squires (KS), Wiener filter, and GLIMPSE. KS is a direct inversion, not accounting for survey masks or noise. The Wiener filter is well-motivated for Gaussian density fields in a Bayesian framework. GLIMPSE uses sparsity, aiming to reconstruct non-linearities in the density field. We compare these methods with several tests using public Dark Energy Survey (DES) Science Verification (SV) data and realistic DES simulations. The Wiener filter and GLIMPSE offer substantial improvements over smoothed KS with a range of metrics. Both the Wiener filter and GLIMPSE convergence reconstructions show a 12 per cent improvement in Pearson correlation with the underlying truth from simulations. To compare the mapping methods' abilities to find mass peaks, we measure the difference between peak counts from simulated ΛCDM shear catalogues and catalogues with no mass fluctuations (a standard data vector when inferring cosmology from peak statistics); the maximum signal-to-noise of these peak statistics is increased by a factor of 3.5 for the Wiener filter and 9 for GLIMPSE. With simulations we measure the reconstruction of the harmonic phases; the phase residuals' concentration is improved 17 per cent by GLIMPSE and 18 per cent by the Wiener filter. The correlation between reconstructions from data and foreground redMaPPer clusters is increased 18 per cent by the Wiener filter and 32 per cent by GLIMPSE.

2.1 Introduction

Mass map reconstruction from weak gravitational lensing recovers the underlying matter distribution in the universe from measurements of galaxy shapes. Images of distant galaxies are deformed by the inhomogeneous matter distribution along the line of sight. Any matter can contribute to the lensing effect, making it a direct probe of non-visible dark matter.

Weak lensing, which takes advantage of the statistical power from many small distortions (that is, gravitational lensing induced "shears"), is now a well established tool in constraining cosmology. The Dark Energy Survey (DES) has used the 2-point correlation function of shear to contribute to excellent constraints on cosmological parameters and models, including the nature of dark energy (DES Collaboration et al., 2017b). Shear 2-point correlation functions have been used to constrain cosmology from many other survey datasets (van Uitert et al. 2017, Kilbinger et al. 2013). These methods use the shear measurements directly, as the shear can be related to the underlying matter distribution without needing to explicitly reconstruct mass maps.

A zero-mean Gaussian random field can be characterised entirely by its 2-point correlations. The matter density field in the early universe is expected to be highly Gaussian, a property which persists into the late universe for the large scales that were less affected by gravitational collapse. For the smaller scales at late times, non-linear gravitational collapse has led to a highly non-Gaussian density field. Much valuable information can be extracted from this non-Gaussianity, although this requires additional methods beyond 2-point statistics.

Popular proposed methods to extract this information include N-point statistics and higher order moments (Cooray and Hu, 2001), peak statistics (Dietrich and Hartlap 2010, Kacprzak et al. 2016, Peel et al. 2017, Shan et al. 2017, Martinet et al. 2017), and Minkowksi functionals (Kerscher et al. 1996, Petri et al. 2013). It is often either essential or convenient to apply these methods to the density field directly (rather than in the space of the shear measurements), thereby necessitating a reliable mass map reconstruction.

Peak statistics are particularly promising, as peaks in the density field probe the non-Gaussian structure directly. Peaks can be identified from aperture mass maps, which are derived by convolving the shear data with a kernel, or from the reconstructed density field. The first approach has the advantage of having local noise, while the second is "closer" to the underlying density field and often has faster algorithms. Both methods often require

simulations to provide a link between the theory and data, with the exception of proposed semi-analytic models (Lin and Kilbinger 2015, Shan et al. 2017).

In addition to using mass maps for higher order statistics to constrain cosmological parameters and models, the mass maps can themselves be intrinsically useful. Clerkin et al. (2017), using the original DES Science Verification (SV) mass map, show evidence that the 1-point distribution of the density field is more consistent with Log-Normal than Gaussian. Combining mass maps with the spatial distributions of stellar mass, galaxies, or galaxy clusters allows the relationship between the visible baryonic matter and invisible dark matter to be studied. Using mass maps to constrain galaxy bias (Chang et al., 2016), the relation between the distribution of galaxies and matter, can in turn aid cosmological probes other than weak lensing. Maps also enable simple tests for systematic error in the galaxy shape catalogues.

Since the first application of mass mapping methods to wide-field surveys with the Canada-France Hawaii Telescope Lensing Survey (CFHTLenS) data (Van Waerbeke et al., 2013), mass maps have been a standard product of large weak lensing surveys. In addition to DES, current surveys reconstructing the density field from weak lensing data include the Kilo-Degree Survey (Giblin et al. in prep.) and the Hyper Supreme-Cam Subaru Strategic Program (HSC-SSP) (Oguri et al., 2017). Mapping dark matter is key to the science goals of the future Euclid Mission (Amendola et al., 2016) and the Large Synoptic Survey Telescope (LSST Science Collaboration et al., 2009).

DES is a ground based photometric galaxy survey, observing in the southern sky from the 4m Blanco telescope at the Cerro Tololo Inter-American Observatory (CTIO) in Chile with five photometric filters covering the optical and near-infrared spectrum using the Dark Energy Camera (Flaugher et al. 2015, Dark Energy Survey Collaboration et al. 2016). The SV data come from an initial run over a fraction of the final sky coverage, but to almost the full exposure time of the final survey. The sky coverage is still large, 139 $\deg^2$, and the nearly full exposure (Chang et al., 2015) gives a galaxy density almost equal to what is expected after the complete 5 years of DES observations.

This Chapter uses the public DES SV data to compare the quality of mass mapping reconstruction methods. The maps are of the two-dimensional convergence, κ, a weighted projection of the density field in the foreground of the observed background galaxies. Recovering the convergence from the shear data is an ill-posed inverse problem, troubled by

survey masks and galaxy "shape noise".

This work follows on from that of Chang et al. (2015) and Vikram et al. (2015) in which the original DES SV mass map was created using the Kaiser and Squires (1993) method. In this Chapter we compare three quite different methods: Kaiser-Squires (KS); Wiener filtering (Wiener, 1949); and GLIMPSE (Leonard et al. 2014, Lanusse et al. 2016), a sparsity-based reconstruction method. The Kaiser-Squires method is a direct inversion from shear to convergence, taking no account of missing data or the effect of noise. The Wiener filter and GLIMPSE assume different prior knowledge about the underlying convergence to account for the effects of noise and missing data.

In section 2.2 we describe the theoretical foundation for weak lensing mass mapping and the three different methods used for this work. In section 2.3 we describe the DES SV shear data, the accompanying simulations, and the redMaPPer galaxy cluster catalogue. Foreground galaxy clusters are expected to trace the true density field, and therefore should be correlated with the convergence reconstruction. The different methods are also applied to realistic data simulations where the true convergence is known. In section 2.4 we present our results on data and simulation, using various quality metrics for the reconstruction. On simulations these metrics are the Pearson correlation coefficient, the pixel root-mean-square error (RMSE), the variance of the 1-point distribution of pixel values, the phase residuals, and peak statistics. On data we compare the convergence reconstructions to the foreground galaxy clusters. We conclude in section 2.5.

2.2 Methodology

2.2.1 Weak gravitational lensing

We can use measurements of the distortion of background galaxy shapes by weak gravitational lensing to learn about the mass distribution in the foreground without making many physical assumptions or relying on phenomenological models. For convenience, here we summarise some of the existing literature relevant for mass mapping from weak lensing (Bartelmann and Schneider 2001, Kilbinger 2015).

The weak lensing formalism follows photon paths along geodesics in a perturbed Friedmann-Robertson-Walker (FRW) metric. The perturbations are sourced by the density field of large scale structure. Throughout we assume that the perturbations are small, and that the measurements are made over a small enough patch of the sky that the sky geometry is Euclidian. Consistent with the Planck CMB results (Planck Collaboration et al., 2016)

and motivated by inflationary theory, we assume that the global geometry of the universe is flat.

The density contrast, $\delta = (\rho - \bar{\rho})/\bar{\rho}$, of a pressureless fluid is related to the scalar gravitational potential perturbation, Φ, through the Poisson equation,

$$\nabla^2\Phi = \frac{3H_0\Omega_m}{2a}\delta \ , \tag{2.1}$$

where H_0 is the present value of the Hubble parameter, a is the cosmological scale factor, and ρ and $\bar{\rho}$ are the local and mean density respectively.

For a flat universe, the lensing potential is given by

$$\psi(\vec{\theta},\omega) = 2\int_0^{\omega} \mathrm{d}\omega' \Big[\frac{\omega-\omega'}{\omega\omega'}\Big]\Phi(\vec{\theta},\omega') \ , \tag{2.2}$$

where ω is the comoving distance.

The Born approximation assumes that the observed angle to a point, $\vec{\theta}$, deviates only a small amount from the true angle $\vec{\beta}$, so the change in distance of the photon's path is negligible. We can characterise the effect of lensing on the galaxies using the Jacobian of the transformation, $\mathscr{A}_{ij} = \partial\beta_i/\partial\theta_j$, which is decomposed into the functions $\kappa(\vec{\theta})$ and $\gamma(\vec{\theta}) = \gamma_1 + i\gamma_2$, and which is given by

$$\begin{aligned}\mathscr{A} &= \Big(\delta_{ij} - \frac{\partial^2\psi(\vec{\theta})}{\partial\theta_i\partial\theta_j}\Big) \\ &= \begin{pmatrix} 1-\kappa-\gamma_1 & -\gamma_2 \\ -\gamma_2 & 1-\kappa+\gamma_1 \end{pmatrix} .\end{aligned} \tag{2.3}$$

Using the definition of the lensing potential and the Poisson equation, the convergence can be expressed as an integral over the density along the line of sight,

$$\kappa(\vec{\theta},\omega) = \frac{3H_0^2\Omega_m}{2}\int_0^{\omega}\mathrm{d}\omega' \frac{\omega'(\omega-\omega')}{\omega}\frac{\delta(\vec{\theta},\omega')}{a(\omega')} \ . \tag{2.4}$$

For a distribution $n(\omega)$ of lensed galaxies, the lensing efficiency kernel is defined to be

$$p(\omega') = \int_{\omega'}^{\omega_\infty}\left(\frac{\omega-\omega'}{\omega}\right)n(\omega)\mathrm{d}\omega \ ; \tag{2.5}$$

this weights the contribution of the foreground density fluctuations to give the convergence

weighted over the redshift distribution of source galaxies,

$$\begin{aligned}\kappa(\vec{\theta}) &= \int_0^\infty n(\omega)\kappa(\vec{\theta},\omega)\mathrm{d}\omega \\ &= \frac{3H_0^2\Omega_m}{2}\int_0^\infty \mathrm{d}\omega' p(\omega')\omega'\frac{\delta(\vec{\theta},\omega')}{a(\omega')} \ .\end{aligned} \tag{2.6}$$

The shear, $\gamma(\vec{\theta})$, which is assumed to be an observable in the weak lensing limit, is given by

$$\begin{aligned}\gamma(\vec{\theta}) &= \frac{1}{\pi}\int_{\mathbb{R}^2}\mathrm{d}^2\theta' \mathscr{D}(\vec{\theta}-\vec{\theta}')\kappa(\vec{\theta}') \\ &\text{where } \mathscr{D}(\vec{\theta}) = -(\theta_1 - i\theta_2)^{-2} \ .\end{aligned} \tag{2.7}$$

For surveys where the integral is over large angles on the sky, this formulation breaks down, and requires a full treatment in spherical bases. Wallis et al. (2017) show that errors can be introduced at an $\mathscr{O}(1$ per cent$)$ level for correlations between points at DES SV angular separation depending on the projection. All of the methods used here use the small angle approximation, and should suffer equally.

The real and imaginary parts of the shear γ represent a chosen two dimensional coordinate system. In weak lensing, the observed ellipticity[1] of a galaxy ε_{obs} is related to the reduced shear g plus the intrinsic ellipticity of the source galaxy ε_s through

$$\begin{aligned}\varepsilon_{obs} &\approx g + \varepsilon_s \\ &\text{where } \ g = \frac{\gamma}{1-\kappa} \ .\end{aligned} \tag{2.8}$$

The reduced shear is approximately the true shear, $g \approx \gamma$, in the weak lensing limit. This allows a standard definition of observed shear, $\gamma_{obs} = \varepsilon_{obs}$, where the measurements are degraded by "shape noise", caused by the ε_s values of the observed galaxies:

$$\gamma_{obs} \approx \gamma + \varepsilon_s \ . \tag{2.9}$$

The shape noise for a given galaxy is modelled as a randomly-drawn Gaussian variate,

[1] Using the Bartelmann and Schneider (2001) equation 4.10 ellipticity definition for ε.

$\varepsilon_s \sim G(0, \sigma_\varepsilon)$, where σ_ε is estimated from data. The distribution of the ellipticity from the SV data in figure 2.2 is not an exact Gaussian, as the true distribution is the result of galaxy astrophysics, though a Gaussian still has properties that make it a good approximation. The Gaussian would be the maximum entropic, least informative, distribution for known mean and variance, and, by the central limit theorem, would be the correct distribution in the limit of large numbers of galaxies averaged in pixels.

It is possible to extend the simple Kaiser-Squires method (section 2.2.2) to use the reduced shear, g, for the mildly non-linear lensing regime when it is no longer appropriate to assume $g \approx \gamma$ (Schneider and Seitz 1995a, Seitz and Schneider 1995, Seitz and Schneider 2001). This is also done by GLIMPSE (section 2.2.4).

In matrix notation, the problem as given by equations 2.7 and 2.9 can be expressed as a linear model, with a data vector of observed shear measurements

$$\gamma = \mathbf{A}\kappa + \mathbf{n} \,, \tag{2.10}$$

where $\mathbf{A}$ is a discretised version of equation 2.7 and $\mathbf{n}$ is a noise vector due to shape noise (equation 2.9). The elements of the data vector can either correspond to the individual shear measurements or to measurements binned into angular pixels (in which case the noise vector would be the average noise in the pixel).

The convergence need not be reconstructed with the same pixelisation as the shear measurements, giving κ and γ vectors of different length. Missing data due to survey masks would correspond to a shorter γ vector; here one may wish to fill in the convergence in the masked region — this is known as *inpainting*. Different sized κ and γ vectors result in a non-square $\mathbf{A}$ matrix, potentially causing inversion problems.

2.2.2 Kaiser-Squires reconstruction

2.2.2.1 Theory

The convergence-to-shear relationship, equation 2.7, is a convolution in the two dimensional angular plane. The two-dimensional Fourier transforms of the shear and convergence, defined for κ as

$$\tilde{\kappa}(\vec{k}) = \int_{\mathbb{R}^2} \mathrm{d}^2\theta \, \kappa(\vec{\theta}) \exp(i\vec{\theta} \cdot \vec{k}) \,, \tag{2.11}$$

are related through an elementwise product via the convolution theorem

$$\tilde{\gamma}(\vec{k}) = \pi^{-1}\tilde{\mathscr{D}}(\vec{k})\tilde{\kappa}(\vec{k}) \,, \tag{2.12}$$

where the Fourier transform of the kernel is given by

$$\tilde{\mathscr{D}}(\vec{k}) = \pi\frac{(k_1^2 - k_2^2 + 2ik_1k_2)}{|\vec{k}|^2} \,; \tag{2.13}$$

here k_1 and k_2 are the components of $\vec{k}$. Using $\tilde{\mathscr{D}}\tilde{\mathscr{D}}^* = \pi^2$, equation 2.12 can be rewritten:

$$\tilde{\kappa}(\vec{k}) = \pi^{-1}\tilde{\gamma}(\vec{k})\tilde{\mathscr{D}}^*(\vec{k}) \quad \text{for } \vec{k} \neq \vec{0} \,. \tag{2.14}$$

The inverse Fourier transform then returns the convergence reconstruction in configuration space (Kaiser and Squires, 1993).

The real and imaginary parts of the reconstruction are the E- and B-modes respectively, where $\kappa_{\rm recon} = \kappa_E + i\kappa_B$. In standard cosmology (equation 2.7), the convergence sourced by a real density field should be a pure E-mode. Errors, noise or other systematic effects can lead to B-mode contributions to the reconstruction.

2.2.2.2 Implementation

In the matrix formulation of equation 2.10, this deconvolution corresponds to multiplying the Fourier space shear field with the inverse of $\mathbf{A}$ in Fourier space. For a case with no shape noise, that is

$$\tilde{\gamma} = \tilde{\mathbf{A}}\tilde{\kappa} \,, \tag{2.15}$$

the Kaiser-Squires method is identical to using the inverse matrix

$$\begin{aligned}\left[\tilde{\mathbf{A}}^{-1}\right]_{ij} &= \frac{k_{1,i}^2 - k_{2,i}^2 - 2ik_{1,i}k_{2,i}}{k_{1,i}^2 + k_{2,i}^2}\delta_{ij} \\ &= \left[\tilde{\mathbf{A}}^{\dagger}\right]_{ij} \,,\end{aligned} \tag{2.16}$$

where the Kronecker delta function, δ_{ij}, relates the element-wise multiplication in Fourier space to a diagonal matrix operator, and † is the conjugate transpose.

For the Kaiser-Squires inversion in configuration space, the $\mathbf{A}$ and $\mathbf{A}^\dagger$ matrices are not diagonal, and therefore are slower to compute. The discretisation of the underlying smooth

shear field into finite configuration space makes the property $\mathbf{A}\mathbf{A}^\dagger = \mathbf{I}$ inexact. As a result of these factors, we choose to implement the Kaiser-Squires reconstruction in Fourier space.

The shear due to lensing is much smaller than the shape noise, and not all places on the sky contain usable galaxies. Both the shape noise and the random sampling of background galaxies propagate error through this noisy reconstruction. Binning the shear measurements into larger pixels can reduce the shape noise per pixel and ensure that there are no empty pixels, but this comes at a loss of the small scale information and cannot deal with masks or the edges of the survey.

A smoothing filter is applied to the Kaiser-Squires reconstruction to reduce the noise. This will similarly lose any small scale structure, and especially suppress peaks in the convergence. In this work, matching Chang et al. (2015), we smooth the Kaiser-Squires maps with a Gaussian kernel. The standard deviation scale, σ_{smooth}, of this Gaussian kernel is free to be chosen, where $\sigma_{smooth} = 0$ corresponds to standard, unsmoothed Kaiser-Squires.

2.2.3 Wiener Filter

2.2.3.1 Theory

The Wiener filter is the linear minimum-variance solution to linear problems of the type in equation 2.10, where the noise is uncorrelated. The Wiener filter reconstruction (Lahav et al. 1994, Zaroubi et al. 1995) is given by

$$\begin{aligned} \kappa_W &= \mathbf{W}\gamma \\ \mathbf{W} &= \mathbf{S}_\kappa \mathbf{A}^\dagger \left[\mathbf{A}\mathbf{S}_\kappa \mathbf{A}^\dagger + \mathbf{N}\right]^{-1} . \end{aligned} \tag{2.17}$$

Here $\mathbf{S}_\kappa$ and $\mathbf{N}$ are the signal and noise covariance matrices respectively, which are $\langle \kappa\kappa^\dagger \rangle$ and $\langle \mathbf{n}\mathbf{n}^\dagger \rangle$ for this problem.

This filter is the linear minimum-variance solution, as $\mathbf{W}$ is a linear operator that minimises the variance

$$\langle (\mathbf{W}\gamma - \kappa)^\dagger (\mathbf{W}\gamma - \kappa) \rangle . \tag{2.18}$$

If the chosen prior on κ does not constrain the reconstruction, so that $\mathbf{S}_\kappa^{-1} \to 0$ (Simon et al., 2009), or if the data are noise-free, $\mathbf{N} = 0$, then the linear minimum variance filter becomes the Kaiser-Squires reconstruction. Setting $\mathbf{S}_\kappa^{-1} \to 0$ is equivalent to removing the signal

prior in the following Bayesian framework.

From a different starting point, for the Wiener posterior we begin by assuming a Gaussian likelihood (Jasche and Lavaux, 2015)

$$Pr(\gamma|\kappa) = \frac{1}{\sqrt{(\det 2\pi \mathbf{N})}} \exp\Big[-\frac{1}{2}(\gamma - \mathbf{A}\kappa)^{\dagger}\, \mathbf{N}^{-1}(\gamma - \mathbf{A}\kappa)\Big]\,, \tag{2.19}$$

where it is assumed that $\mathbf{N}$ is known and the noise is both uncorrelated and Gaussian, as assumed in equation 2.10. Intrinsic alignments of clustered galaxies will violate this uncorrelation condition.

The prior on the convergence is that of a Gaussian random field, which is applicable for the density field on large scales at late times,

$$Pr(\kappa|\mathbf{S}_\kappa) = \frac{1}{\sqrt{(\det 2\pi \mathbf{S}_\kappa)}} \exp\Big[-\frac{1}{2}\kappa^{\dagger}\, \mathbf{S}_\kappa^{-1}\kappa\Big]\,. \tag{2.20}$$

Using Bayes' theorem and the fact that $Pr(\gamma|\mathbf{S}_\kappa, \kappa) = Pr(\gamma|\kappa)$, the full posterior is given by

$$\begin{aligned} Pr(\kappa|\mathbf{S}_\kappa, \gamma) &= \frac{Pr(\gamma|\kappa)Pr(\kappa|\mathbf{S}_\kappa)}{Pr(\gamma)} \\ &\propto \frac{1}{\sqrt{(\det 2\pi \mathbf{S}_\kappa)}} \frac{1}{\sqrt{(\det 2\pi \mathbf{N})}} \times \\ &\quad \exp\Big[-\frac{1}{2}\kappa^{\dagger}\mathbf{S}_\kappa^{-1}\kappa - \frac{1}{2}(\gamma - \mathbf{A}\kappa)^{\dagger}\mathbf{N}^{-1}(\gamma - \mathbf{A}\kappa)\Big] \\ &\propto \exp\Big[-\frac{1}{2}(\kappa - \mathbf{W}\gamma)^{\dagger}(\mathbf{S}_\kappa^{-1} + \mathbf{A}\mathbf{N}^{-1}\mathbf{A}^{\dagger})(\kappa - \mathbf{W}\gamma)\Big]\,, \end{aligned} \tag{2.21}$$

where $\mathbf{W}$ is the Wiener filter, so the *maximum a posteriori* (MAP) solution is that of the Wiener reconstruction.

The choice of Gaussian prior is physically motivated for the large, linear scales of the density field (see section 2.1); alternative prior distributions can be used to give different κ posterior distributions that can be maximised or from which samples can be drawn (Schneider et al., 2017). Recent work by Böhm et al. (2017) proposes the use of a Log-Normal prior distribution. This appears to fit the κ distribution from simulations (figure 2.6) and data (Clerkin et al., 2017) better than Gaussian, but, unlike the Wiener filter, lacks an analytic MAP solution.

If the aim of the reconstruction is to infer cosmology from the non-Gaussian com-

ponent of the density field, the Wiener filter may not be the ideal method for mass map recovery. The small scale modes with less power are often suppressed, losing the peak structure. Qualitatively it can be thought of as either the Gaussian prior being inappropriate or as the linear filter being insufficient.

2.2.3.2 Implementation

Using the exact Fourier space property $\tilde{\mathbf{A}}^{-1} = \tilde{\mathbf{A}}^{\dagger}$ we rewrite equation 2.17 as

$$\begin{aligned} \tilde{\kappa}_W &= \tilde{\mathbf{A}}^{-1}\tilde{\mathbf{S}}_{\gamma}\left[\tilde{\mathbf{S}}_{\gamma}+\tilde{\mathbf{N}}\right]^{-1}\tilde{\gamma} \\ &= \tilde{\mathbf{A}}^{-1}\tilde{\gamma}_W \\ &= \tilde{\mathbf{A}}^{\dagger}\tilde{\gamma}_W \ , \end{aligned} \tag{2.22}$$

where we have used $\tilde{\mathbf{A}}\tilde{\mathbf{S}}_{\kappa}\tilde{\mathbf{A}}^{\dagger} = \langle \tilde{\mathbf{A}}\tilde{\kappa}\tilde{\kappa}^{\dagger}\tilde{\mathbf{A}}^{\dagger}\rangle = \langle \tilde{\gamma}\tilde{\gamma}^{\dagger}\rangle = \tilde{\mathbf{S}}_{\gamma}$. This shows that applying the Wiener filter to the shear to recover γ_W and then applying the Kaiser-Squires inversion in Fourier space is equivalent to directly calculating the Wiener filter of the convergence.

In configuration space, the noise covariance matrix is given by

$$\left[\mathbf{N}\right]_{ij} = \frac{2\sigma_{\varepsilon}^2}{p_i}\delta_{ij} \ , \tag{2.23}$$

where p_i is the galaxy count per pixel. Empty pixels in the masked region have infinite variance, absorbing the mask into a special case of the Wiener filter denoising.

The signal properties for a Gaussian random field are constrained entirely by the mean and the signal covariance matrix, which in harmonic space is identical to the power spectrum. The cosmological principle implies that the angular distribution of a field on the sky is statistically isotropic, so the angular power spectrum, C_{ℓ}, can contain all the 2-point statistical information. The angular power spectrum of the physical shear E-mode shear signal is defined as

$$\begin{aligned} C_{\ell,E} &= \frac{1}{2\ell+1}\sum_{m=-\ell}^{+\ell}\langle |a_{\ell m,E}|^2\rangle \\ &= C_{\ell,\kappa} \ , \end{aligned} \tag{2.24}$$

where $a_{\ell m}$ are the spherical harmonic coefficients and the brackets $\langle\ \rangle$ average over realisa-

tions of the signal. The second equality assumes the flat sky approximation for high[2] ℓ.

We generate a theoretical power spectrum using the Limber approximation with the COSMOSIS package (Zuntz et al., 2015) with our prior fiducial cosmological parameters: $\Omega_m = 0.286,\ \Omega_\Lambda = 0.714,\ \Omega_b = 0.047,\ h = 0.7,\ \sigma_8 = 0.82,\ n_s = 0.962$ and $w = -1$. We use a background galaxy distribution defined from equation 2.31, and shown in figure 2.1.

It is commonly asked whether it is reasonable to assume cosmological parameters in the map reconstruction, if the maps are then used to infer cosmological parameters. Though we assume a specific set of cosmological parameters, it would still be possible to use the maps for cosmological parameter estimation, from peak statistics for example, if the same prior is used on the simulations and the data identically. If simulations are not used, the power spectrum can be jointly inferred from the data (Jasche and Lavaux, 2015) using Gibbs sampling.

In order to generate the power spectrum in flat Fourier space, rather than on the curved sky, we again use a flat sky approximation

$$k_\theta^2 P(k_\theta) = \left(\frac{\mathcal{N}}{2\pi}\right)^2 \ell(\ell+1)C_\ell \ , \tag{2.25}$$

adapted from Loverde and Afshordi (2008), where $\mathcal{N}$ is the total number of pixels in the map, k_θ is the magnitude of the projected Fourier mode, and where we have defined our projected angular power spectrum as

$$P(k_\theta)\delta(\mathbf{k}_\theta - \mathbf{k}'_\theta) = \langle\ \tilde{\gamma}(\mathbf{k}_\theta)\ \tilde{\gamma}^\dagger(\mathbf{k}_\theta)\ \rangle \ . \tag{2.26}$$

The largest scale mode is $\ell = 20.51$, which corresponds to an angular separation of 17.55 deg.

Though the signal covariance matrix is diagonal in harmonic space (equation 2.26), and the independent noise has covariance which is diagonal in configuration space (equation 2.23), there is no natural basis in which both are sparse. Inversion of dense matrices to evaluate the Wiener filter is bypassed using the algorithm presented in Elsner and Wandelt (2013), where an additional messenger field is used to pass information between harmonic and configuration space, iteratively converging to the Wiener filter solution.

These messenger field methods were extended by Jasche and Lavaux (2015) to draw

[2]We omit a prefactor which goes as $1 - \mathcal{O}(\ell^{-2})$ for high ℓ.

Markov chain Monte Carlo (MCMC) samples from the whole Wiener posterior (equation 2.21). The first application of messenger field methods to weak lensing data was by Alsing et al. (2016a, 2017), who drew samples from the Wiener posterior and generated Wiener filtered shear (not convergence) maps from CFHTLenS data. By comparison, in this work we do not sample from the Wiener posterior; instead, we use the original messenger field algorithm of Elsner and Wandelt (2013) to calculate the Wiener filter reconstruction of the convergence map from DES SV shear data and simulations.

2.2.4 Sparsity reconstruction

2.2.4.1 Theory

Consider the coefficients α of the decomposition of a signal $\mathbf{x}$ in a representation space (or "dictionary") Φ, so that $\mathbf{x} = \Phi\alpha$. Example dictionaries include the Fourier transform or wavelet transforms. Assuming a sparse prior on the signal $\mathbf{x}$ in the dictionary Φ means that its representation α is expected to be sparse, that is, with most of the coefficients equal to 0 (Starck et al., 2015). A simple example is a cosine function signal and a Fourier transformation; in this sparse basis only two coefficients have a non-zero value (corresponding to the frequency of the cosine function).

Formally most signals cannot strictly be made sparse, and are merely compressible with a choice of an appropriate transformation, such as a wavelet transform (Starck et al. 2015, Leonard et al. 2014). For a compressible signal the magnitude-ordered sparse coefficients, α_i, are expected to have exponential decay and therefore to have a Laplace distribution (Tibshirani, 1994).

Consider a generic linear inverse problem of the form $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$. A robust estimate of the signal $\mathbf{x}$ can be recovered by solving the ("LASSO") optimisation problem

$$\arg\min_{\alpha} \; ||\mathbf{y} - \mathbf{A}\Phi\alpha||_2^2 + \lambda||\alpha||_1 \; , \tag{2.27}$$

where λ is a Lagrangian multiplier (Tibshirani, 1994). Here the first term corresponds to a χ^2 minimisation, ensuring fidelity of the signal reconstruction, while the second is the sparsity-promoting regularisation term.

We can include non-constant noise variance by weighting the first χ^2 according to the variance. If the noise variance is included in the χ^2 term, the λ value can be interpreted as

a signal-to-noise level in the transformed (e.g. wavelet) space.

The second term does not use the Euclidan l_2 norm, but instead uses the sparsity-promoting l_1 norm, defined as

$$||\alpha||_1 = \sum_i |\alpha_i| \,. \tag{2.28}$$

These methods are non-linear, so it can be difficult to derive properties analytically. With realistic simulations of the data and true signal, the value for λ can be chosen to maximise some success metric. This is analogous to selecting a theoretical power spectrum for the Wiener filter, or a smoothing scale for Kaiser-Squires.

Sparse recovery methods are non-linear and are not necessarily formulated in the Bayesian framework of the Wiener filter. If one wishes to, one could make a frequentist estimate of the error of the sparse reconstruction by propagating the noise properties of the data using bootstrapping or Monte Carlo techniques.

Alternatively, sparsity can be derived from a Bayesian[3] perspective, with the l_1 regularisation term corresponding to a Laplacian prior (Tibshirani, 1994).

2.2.4.2 Implementation/GLIMPSE

The choice of dictionary depends on the structures contained in the signal. Theory of structure formation in the universe predicts the formation of quasi-spherical halos of bound matter. It is standard practice to represent the spatial distribution of matter in halos with spherically symmetric Navarro-Frenk-White (Navarro et al., 1996) or Singular Isothermal Sphere profiles. Coefficients of Isotropic Undecimated Wavelets (Starck et al., 2015) in two dimensions are well suited to the observed convergence of a dark matter halo. The wavelet transform used in the GLIMPSE algorithm is the starlet (Starck et al., 2007), which can represent positive, isotropic objects.

The sparsity prior in the starlet basis enforces a physical model that the matter field is a superposition of spherically symmetric dark matter halos. This is not wholly correct, but is an approximation which is true for the non-linear regime in the standard model of structure formation, similarly to how the assumption of Gaussianity holds in the linear regime. On large scales, where the density field is expected to be Gaussian, the GLIMPSE sparsity prior is less appropriate.

[3]Since this work was carried out, Price et al. (2018) demonstrated a possible method whereby credible intervals (posterior-derived Bayesian error bars) can be approximately calculated at pixel level in the mass map. Sparse MCMC methods have also been proposed by for radio interferometric imaging by Cai et al. (2018).

The GLIMPSE algorithm aims to solve the optimisation problem

$$\hat{\kappa} = \arg\min_{\kappa} \; ||\mathbf{N}^{-\frac{1}{2}}\left[\gamma - \mathbf{T}^{\dagger}\hat{\mathbf{A}}\mathbf{F}\kappa\right]||_2^2 + \lambda||\omega\Phi^{\dagger}\kappa||_1 + i_{\mathrm{Im}(\kappa)=0} \; , \tag{2.29}$$

where $\mathbf{F}$ is the Fourier transform matrix, $\mathbf{T}$ is the Non-equispaced Discrete Fourier Transform (NDFT) matrix, $\hat{\mathbf{A}}$ is defined in equation 2.16, ω is a diagonal matrix of weights, and $\Phi^{\dagger}$ is the inverse wavelet transform. The indicator function $i_{\mathrm{Im}(\cdot)=0}$ (defined in section 2.6) in the final term imposes realness on the reconstruction (no B-modes). The use of NDFT allows the first term to perform a forward fitted Kaiser-Squires-like step without binning the shear data, allowing the smaller-scales to be retained in the reconstruction. The full algorithm, including the calculation of the weights, is described in section 3.2 in Lanusse et al. (2016).

Though the problem presented in equation 2.29 is an optimisation using the shear data γ, in fact it is the reduced shear (equation 2.8) that GLIMPSE uses to recover κ (Lanusse et al., 2016). As an extension, the GLIMPSE algorithm can also perform the joint reconstruction with reduced shear and flexion, a third-order weak gravitational lensing effect (Bacon et al., 2006) (although no flexion data are available for our galaxy shear catalogue).

As the prior knowledge in this reconstruction relates to the quasi-spherical clustering of bound matter, enforced through a sparsity prior in starlet space, this method should better reconstruct the smaller scale non-Gaussian structure than the Wiener filter.

2.3 Data and Simulations

2.3.1 Dark Energy Survey Science Verification Data

The shear data are from the 139 deg^2 SPT-E field of the public DES SV data. This initial test dataset was taken during an observing run before the official start of the full science survey. The galaxy catalogue comes from the SVA1 (Science Verification) Data Release[4]. Due to changes to the catalogues before final release (more galaxy shear measurements are now available to us), the catalogue used in this work is not identical to that used by Chang et al. (2015), even when the same data selections are made. All maps are therefore new, and slightly different to the previously published SV map.

[4] http://des.ncsa.illinois.edu

The photometric redshifts from five optical filters (*grizY*) were estimated using the Bayesian Photometric Redshifts (BPZ) code (Benítez 2000, Coe et al. 2006, & Bonnett et al. 2016). The final median depth estimates are $g \sim$24.0, $r \sim$23.0, $i \sim$23.0 and $z \sim$22.4 (10-σ galaxy limiting magnitude). The "background galaxies", the ones from which the shear is measured, are taken in the range $0.6 < z_{mean} < 1.2$. The z_{mean} value for each galaxy is the mean of the posterior probability distribution function (PDF) estimated using the BPZ code. The PDF for each galaxy is very broad, giving a total stacked PDF of background galaxies that extends beyond the $[0.6, 1.2]$ redshift range, as can be seen in figure 2.1.

Using the `ngmix` shape catalogue, we apply a selection of `sva1_flag` $= 0$ & `ngmix_flag` $= 0$ to obtain galaxies with a well-measured shear. The `ngmix` catalogue contains corrections to measurement bias, in the form of "sensitivities", which can be applied to a weighted ensemble of hundreds or thousands of galaxies, but which cannot be applied per galaxy (which is not ideal for mass mapping). The structure of equation 2.7 implies that a multiplicative shear bias would lead to a convergence amplitude bias. Under the assumption that multiplicative shear bias will not vary across the survey area, we correct all measured ellipticities by the same debiasing factor

$$\varepsilon_{obs,i} = \varepsilon_{measured,i} \times \bar{s}^{-1} \,, \tag{2.30}$$

where i is a galaxy index and $\bar{s}$ (≈ 0.82) is the mean sensitivity correction from all galaxies in our `ngmix`-selected catalogue. The total number of galaxies after the redshift and shape measurement selection is $1,628,663$.

For the Kaiser-Squires reconstruction, the shear measurements are binned into angular pixels in a 256×256 map, with average pixel size of 4.11 arcmin, using a sinusoidal projection with a centre at RA=71.0 deg. This is similar to the 5 arcmin pixel scale of the original Chang et al. (2015) map. The choice of central RA for Kaiser-Squires is to minimise the mask in the square projection, which is a large source of systematic error. For the Wiener filter, where the mask is taken into account, the shear measurements are also binned into angular pixels in a 256×256 map, but sinusoidally projected with a central RA=81.3 deg, to make the square maximally isotropic. The GLIMPSE algorithm does not bin the input shear measurements, but requires a pixel scale for the reconstruction, which we set as 3 arcmin using its gnomonic projection centred on RA=76.95 deg and DEC=$-$52.23 deg.

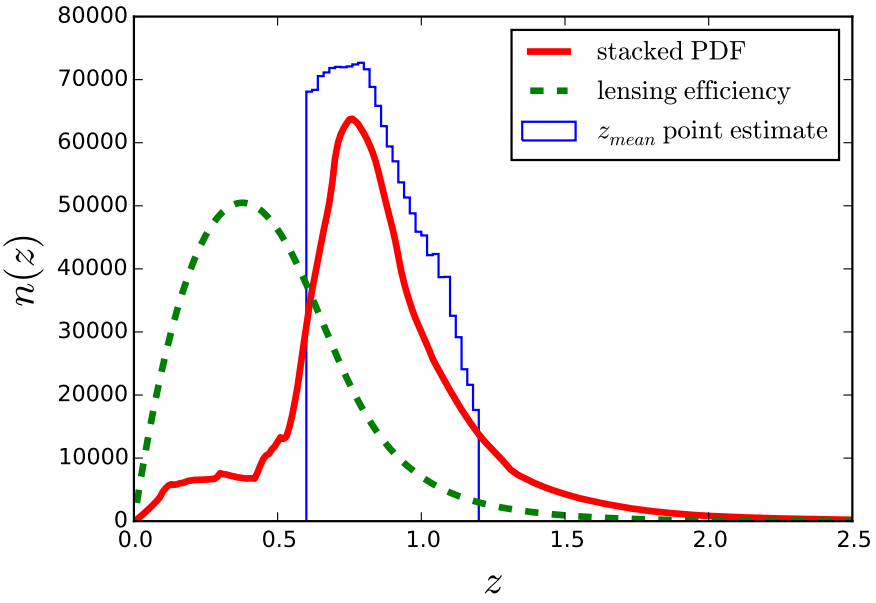

Figure 2.1: The redshift distribution from BPZ of the selected background galaxies with $0.6 < z_{mean} < 1.2$. The blue solid histogram is of the galaxies' point estimate mean redshifts in bins of $\Delta z = 0.02$. The red line is the stacked redshift probability density function (PDF) of all selected galaxies. The green dashed line is the lensing efficiency (equation 2.5) of the background galaxies.

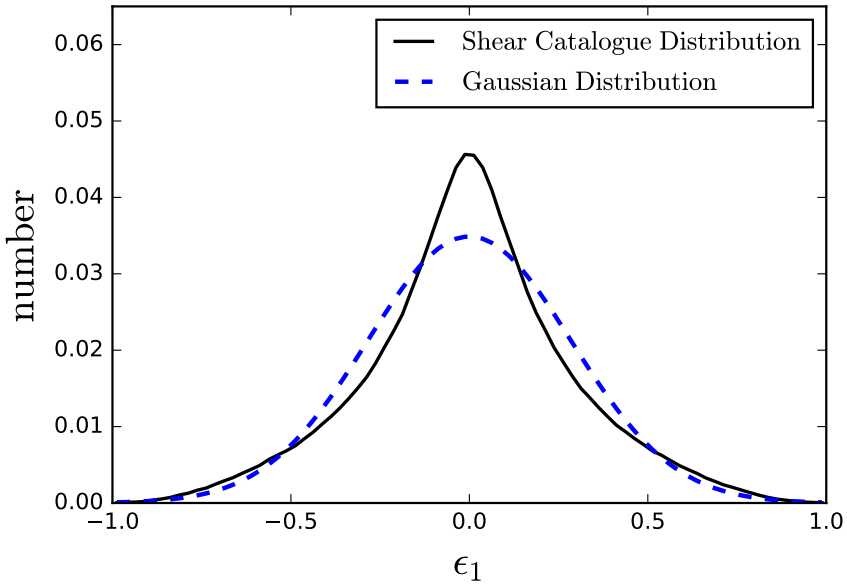

Figure 2.2: Distribution of the first component of ellipticity, ϵ_1, from the selected SV catalogue. A Gaussian distribution with the same mean and standard deviation shows that the ellipticity distribution is not a true Gaussian, though the noise per pixel will be more closely Gaussian due to the central limit theorem.

2.3.2 redMaPPer Clusters

Groups and clusters of galaxies are expected to trace the highest density regions in the foreground. They are luminous objects that correspond to regions of highly non-linear growth, where the density field has deviated from Gaussianity.

The public redMaPPer cluster catalogue (Rykoff et al., 2016) used the redMaPPer algorithm to optically identify clusters and to estimate each cluster's richness, λ_{RM}. The richness is defined as the sum of the membership probabilities over all galaxies within a scale radius (chosen to minimise the scatter in the mass-richness relation); it gives an estimate for the number of galaxies in a cluster. Cluster mass is expected to scale approximately linearly with richness. The redshift uncertainty is excellent, around $\sigma_z/(1+z) \sim 0.01$, due to the clusters containing large numbers of well modelled, red galaxies. The public redMaPPer catalogue used in this work contains only clusters with $\lambda_{RM} \geqslant 20$, so that the clusters with less certainty of detection and characterisation are not used.

2.3.3 Simulations

To compare the reconstructions between different methods, we use a simulated catalogue with a known true convergence. We use a set of N-body simulations developed for the DES collaboration and designed to be representative of the DES data (Busha et al., 2013). The simulations used are N-body light cones composed from three boxes (1400^3, 2048^3, and 2048^3 particles in boxes of comoving length 1050 Mpc/h, 2600 Mpc/h, and 4000 Mpc/h respectively). The cosmological parameters for the simulations are: $\Omega_m = 0.286$, $\Omega_\Lambda = 0.714$, $\Omega_b = 0.047$, $\sigma_8 = 0.82$, $h_0 = 0.7$, $n_s = 0.96$, $w = -1$. We apply a mask to match the SV data.

Source galaxies have randomly-assigned positions in the simulations, as correlation between the background galaxy positions and the weak lensing shear signal is expected to be negligible. The simulated catalogues contain the lensing matrix components, $\mathscr{A}_{ij}$, for each galaxy, calculated with the ray-tracing code `CALCLENS` (Becker, 2013). This provides the true κ and γ per galaxy, from which we derive the reduced shear. The shape noise due to the intrinsic ellipticities of the source galaxies, ε_s, is simulated by adding an ellipticity component to the reduced shear. Each noise realisation is generated from the data by randomly exchanging the ellipticity values between galaxies in the catalogue to remove the weak lensing signal and leave the shape noise.

We attempt to match the redshift distribution of the simulated galaxies to the observed

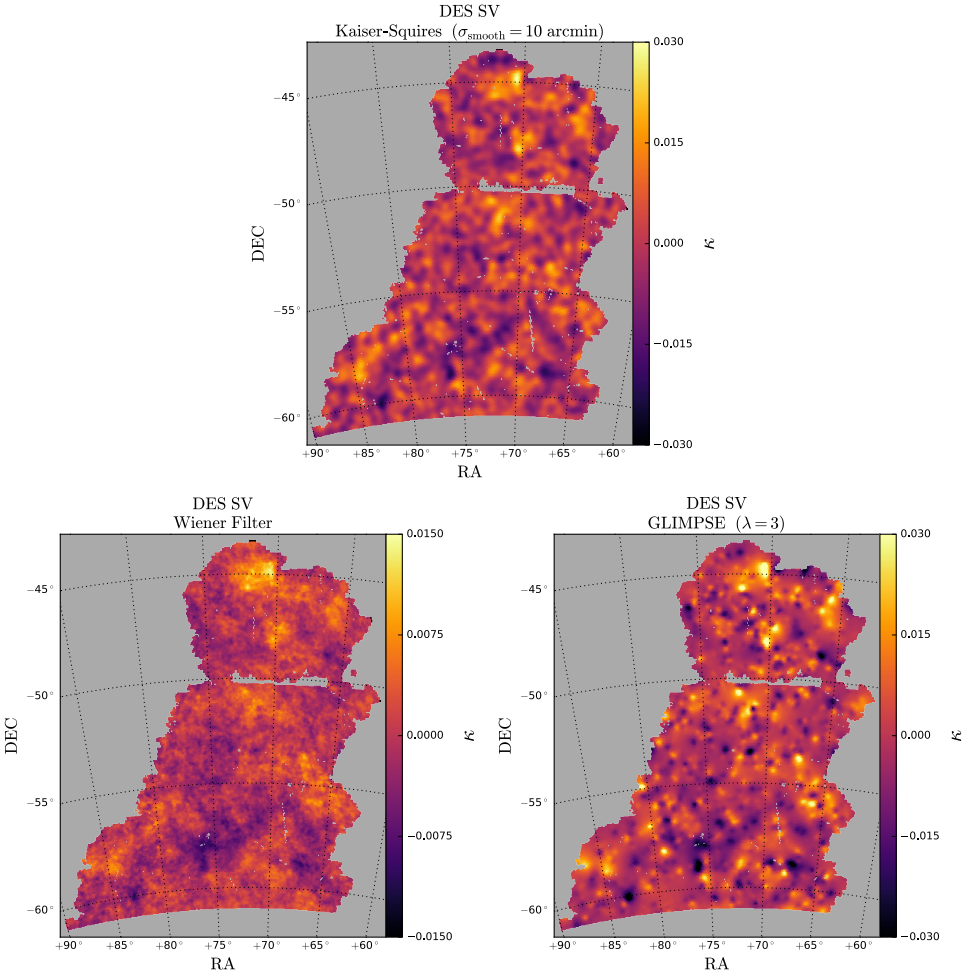

Figure 2.3: The convergence (κ) map reconstructions using the DES SV shear data with the three different methods. *Top panel*: Kaiser-Squires reconstruction with a smoothing scale $\sigma_{smooth} = 10$ arcmin. *Right panel*: The GLIMPSE reconstruction with a regularisation parameter $\lambda = 3.0$. Both tuning parameters were chosen to maximise the Pearson correlation coefficient r when tested on simulations (See section 2.4.1). *Left panel*: The Wiener filter reconstruction. **Note** that the colour scale for the Wiener filter is less than that for the other reconstructions, as the pixel values are closer to zero.

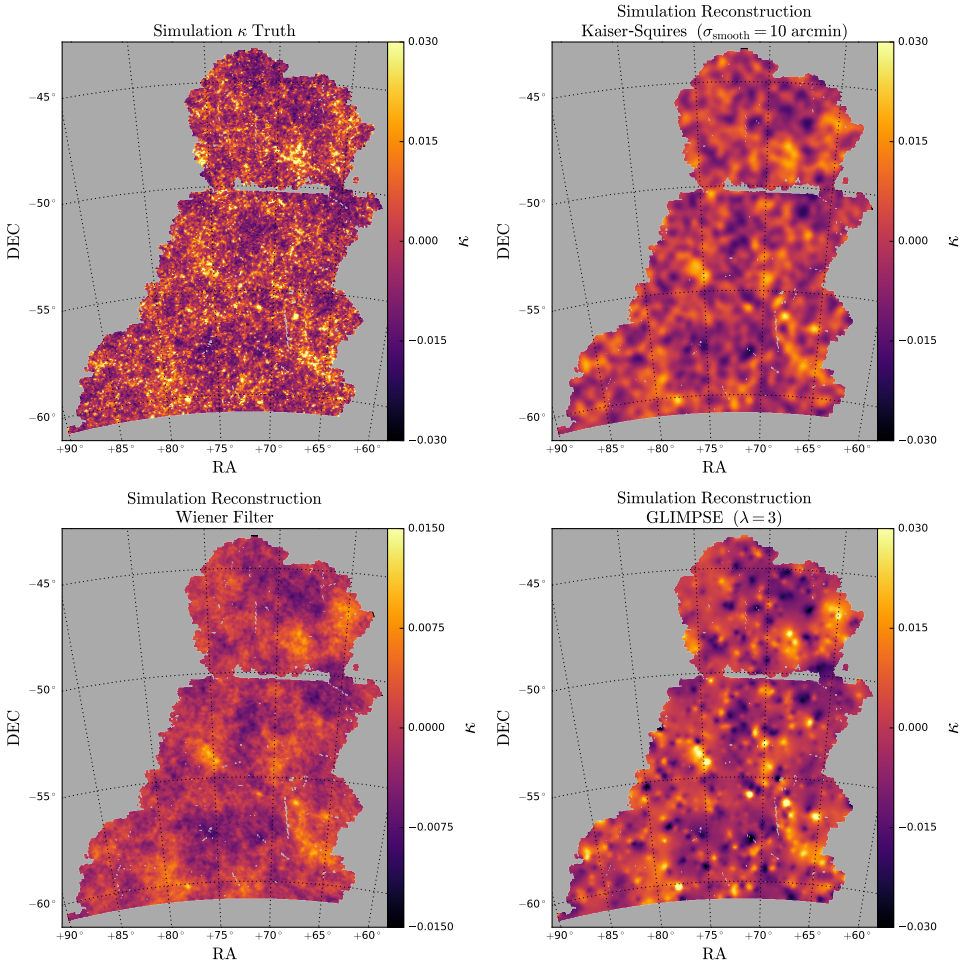

Figure 2.4: The *top left* panel is an example of a true convergence map, κ^{truth}, from simulation. The *top right* panel is the Kaiser-Squires reconstruction with a smoothing scale $\sigma_{smooth} = 10$ arcmin. The *bottom right* panel is the GLIMPSE reconstruction with regularisation parameter $\lambda = 3.0$. Both tuning parameters were chosen to maximise the Pearson correlation coefficient r when tested on simulations (see section 2.4.1). The *bottom left* panel is the Wiener filter reconstruction. **Note** that the colour scale for the Wiener filter is less than that for the other reconstructions, as the pixel values are closer to zero.

redshift distribution, $n(z)$. We use the stacked posterior probability density functions of individual galaxy redshifts from the selected data catalogue (figure 2.1), giving an estimate of the true underlying distribution. This assumes that

$$n(z) = \sum_i p_i(z) \ , \tag{2.31}$$

where $p_i(z)$ are the individual probability distributions for the galaxies from BPZ. This is not necessarily exact, due to errors in $p_i(z)$ per galaxy (Leistedt et al., 2016), but is a reasonable choice for a simulated catalogue. Using rejection sampling in bins of $\Delta z = 0.02$ we select galaxies with a probability equal to the ratio between the desired $n(z)$ from the data and the distribution in the simulation. One typical simulated catalogue contained $1,629,024$ galaxies, slightly different to the data catalogue due to the sampling scheme, but with the desired $n(z)$.

2.4 Results

To ensure that the mass map tests are consistent with different output formats, all maps were converted onto a spherical pixelisation using `HEALPix` (Górski et al., 2005). A `HEALPix` map comprises twelve subdivisions on the sphere, which are then each partitioned into `NSIDE` $\times$ `NSIDE` grids. Each pixel of a `HEALPix` supersampled `NSIDE` $= 4096$ map was filled according to the value at the corresponding RA and DEC in the reconstructed maps. The supersampled high `NSIDE` maps were then degraded to `NSIDE` $= 1024$. The true convergence maps from the simulations were directly binned from the convergence values at galaxy positions to `NSIDE` $= 1024$. For all maps the same mask is applied, where pixels with no galaxies are masked.

Figure 2.3 shows the mass map reconstructions from the SV shear data using the three different methods. An example simulation with truth and the three reconstructed maps is shown in figure 2.4. The "tuning parameters", $\sigma_{smooth} = 10.0$ arcmin for Kaiser-Squires and $\lambda = 3.0$ for GLIMPSE, are tuned to maximise the Pearson correlation coefficient r with the underlying truth when tested on simulations.

Using a suite of 10 simulations, in section 2.4.1 we calculate the Pearson correlation coefficient between the truth and the reconstruction with different methods as a test of the reconstruction's quality. In section 2.4.2, we calculate the root-mean-square error of the residuals between the truth and the reconstruction. In section 2.4.3 we calculate the variance

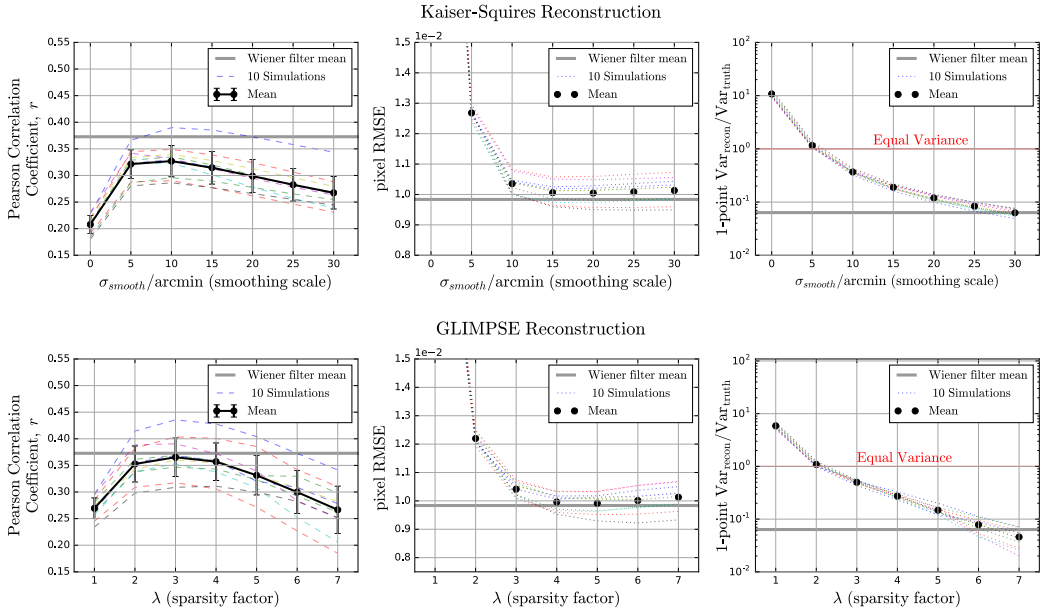

Figure 2.5: Kaiser-Squires (top) and GLIMPSE (bottom). Three different statistics comparing the true κ map and the reconstruction with 10 simulations. *Left panel*: The Pearson correlation coefficient, r (equation 2.32). The errorbar on the mean is the standard deviation of the sample. The better the reconstruction, the higher the value of r. *Middle panel*: The lower the pixel RMSE (equation 2.34), the better the reconstruction. *Right panel*: Ratio of variances between the 1-point distribution of the pixels in the reconstruction and pixels in the true map (equation 2.35).

of the 1-point distribution of the pixel values in the reconstruction and compare with the truth. In section 2.4.4 and section 2.4.5 we quantify the quality of the reconstruction of the phase and peak statistics respectively, by comparing to the simulated truth. The final result presented in section 2.4.6 compares the reconstruction from the DES SV shear data with foreground galaxy clusters from the redMaPPer catalogue (which are expected to trace non-linearities in the underlying density field).

In this work we do not use correlation functions as a test of the map reconstruction. None of the mass mapping methods here are expected to reproduce the correct correlation functions or power spectra. It is simple to show this analytically with the Wiener filter, where despite the filter giving the MAP pixel values, the pixel variance, and therefore the power spectrum, is suppressed.

2.4.1 Pixel Cross Correlation

We quantify the correlation between the true convergence from simulation and the reconstructed convergence of the simulated catalogue using the Pearson correlation coefficient. As with other metrics of success for mass map reconstruction, this can be used to tune the

sparsity λ parameter and the smoothing scale for Kaiser-Squires.

The Pearson correlation coefficient, r, between the pixels' true convergence, κ^{truth}, and the reconstruction, $\kappa^{\rm recon}$, is given by

$$r = \frac{\sum_{i=1}^{n}(\kappa_i^{truth} - \bar{\kappa}^{truth})(\kappa_i^{\rm recon} - \bar{\kappa}^{\rm recon})}{\sqrt{\sum_{i=1}^{n}(\kappa_i^{truth} - \bar{\kappa}^{truth})^2}\sqrt{\sum_{i=1}^{n}(\kappa_i^{\rm recon} - \bar{\kappa}^{\rm recon})^2}} \,, \tag{2.32}$$

where the summations are over all pixels i in the map and $\bar{\kappa}$ is the mean convergence in the map.

In the left panels of figure 2.5, the Pearson r value from 10 simulations is plotted for varying tuning parameters. Almost all of the simulations and also their mean have a maximal Pearson r value at $\sigma_{smooth} = 10.0$ arcmin for Kaiser-Squires and at $\lambda = 3.0$ for GLIMPSE.

Table 2.1 presents the mean value from the 10 simulations, where the tuning parameter is chosen to maximise r when relevant. All methods show good correlation with the underlying true convergence. Both the Wiener filter and GLIMPSE have the same highest value of $r = 0.37$, 12 per cent higher than Kaiser-Squires.

Note that the Pearson correlation coefficient as presented in equation 2.32 is invariant under a rescaling of the reconstruction. Despite the Wiener filter reconstruction having values closer to zero, the Wiener filter maps still have good correlation to the truth. This second aspect is addressed in section 2.4.3 and in the second column of table 2.1.

2.4.2 Pixel Residuals

The difference between the true convergence from simulation and the reconstruction in pixel i is defined as

$$\Delta\kappa_i = \kappa_i^{truth} - \kappa_i^{\rm recon} \,. \tag{2.33}$$

We define the root-mean-square error (RMSE) as

$$\mathrm{RMSE}(\kappa^{truth}, \kappa^{\rm recon}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\Delta\kappa_i^2} \tag{2.34}$$

where n is the number of pixels.

A smaller value of RMSE for a given method implies a better reconstruction according to this metric. It is this RMSE that the Wiener filter attempts to minimise using a linear

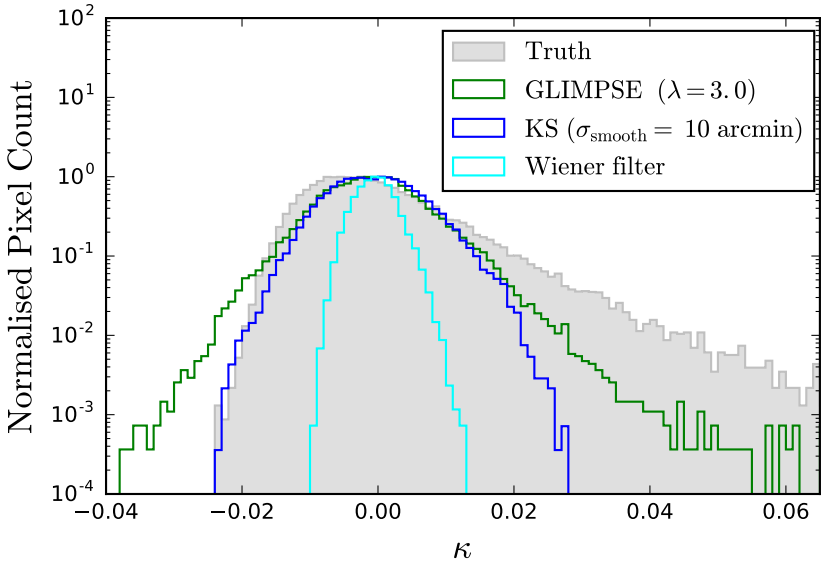

Figure 2.6: Pixel histograms (1-point distributions) for various map reconstructions from the simulated data shown in figure 2.4. The histograms are normalised such that the largest value of each is equal to one. The ratio of the variance between the reconstructions and the truth is presented in table 2.1.

filter, as defined in equation 2.18, by using an assumed signal covariance $\langle \kappa \kappa^{\dagger} \rangle$ (see section 2.2.3.1).

The centre panel of figure 2.5 shows that increasing the smoothing scale, σ_{smooth}, for Kaiser-Squires or the regularisation parameter, λ, for GLIMPSE initially reduces the pixel RMSE, but increased filtering contributes little beyond $\sigma_{smooth} = 10.0$ arcmin for Kaiser-Squires or $\lambda = 3.0$ for GLIMPSE.

The smallest mean pixel RMSE is 1.0×10^{-2} for Kaiser-Squires and 9.9×10^{-3} for GLIMPSE. The Wiener filter, whose smoothing is constrained by the prior on C_{ℓ} and which therefore cannot be tuned, has a pixel RMSE of 9.4×10^{-3}.

2.4.3 Pixel 1-Point Variance

The 1-point distribution can be thought as a histogram of the pixel values. Figure 2.6 shows an example of such a histogram (derived from the simulated truth map and reconstructions of figure 2.4).

The mean of this distribution is unconstrained by weak lensing, due to an integration constant in equation 2.7. The variance of the 1-point distribution is increased compared to the underlying truth due to shape noise in the unsmoothed Kaiser-Squires reconstruction. A reconstruction method would aim to reduce the variance of the 1-point pixel distribution to

Table 2.1: The centre column gives the average Pearson correlation coefficient r (equation 2.32) between κ^{truth} and $\kappa^{\rm recon}$ from 10 simulations. The choices of $\sigma_{smooth} = 10$ arcmin and $\lambda = 3.0$ maximise the Pearson r value. The right column gives the ratio of the pixel variance between $\kappa^{\rm recon}$ and κ^{truth} (equation 2.36).

Method	Pearson r	Variance Ratio
KS ($\sigma_{smooth} = 10$ arcmin)	0.33	3.7×10^{-1}
Wiener filter	0.37	6.3×10^{-2}
GLIMPSE ($\lambda = 3.0$)	0.37	5.0×10^{-1}

match that of the underlying truth.

We define the estimate of the variance of the 1-point distributions of the truth or reconstructed κ as

$$\begin{aligned} \mathrm{Var}_{\rm truth} &= \frac{1}{n-1}\sum_{i=1}^{n}(\kappa_i^{\rm truth} - \bar{\kappa}^{\rm truth})^2 \\ \mathrm{Var}_{\rm recon} &= \frac{1}{n-1}\sum_{i=1}^{n}(\kappa_i^{\rm recon} - \bar{\kappa}^{\rm recon})^2 \ , \end{aligned} \tag{2.35}$$

where the notation matches equation 2.34. The ratio of these variances is given by

$$\frac{\mathrm{Var}_{\rm recon}}{\mathrm{Var}_{truth}} = \frac{\sum_{i=1}^{n}(\kappa_i^{\rm recon} - \bar{\kappa}^{\rm recon})^2}{\sum_{i=1}^{n}(\kappa_i^{truth} - \bar{\kappa}^{truth})^2} \ , \tag{2.36}$$

The closer this value is to 1, the better the variance of the pixel distribution matches the truth. Using 10 simulations we can calculate this quantity for different reconstruction methods (and at different smoothing scales or λ regularisation values where relevant).

In figure 2.5 the right panel shows the result of this test for GLIMPSE and Kaiser-Squires. Both methods show a pixel distribution that has too high variance for insufficient filtering, and too low variance for over-filtering. For Kaiser-Squires, the ratio is closest to 1 at a smoothing scale of $\sigma_{smooth} = 5$ arcmin. For GLIMPSE, the ratio is closest to 1 at a sparsity regularisation value of $\lambda = 2$.

Both of these reconstruction methods have a matching variance at a smoothing parameter value less than that which maximises the Pearson correlation coefficient r. If one chose this parameter to maximise the Pearson r value, such that $\lambda = 3$ and $\sigma_{smooth} = 10$ arcmin, a good reconstruction should also have the ratio of the variances as close to 1 as possible.

The right column of table 2.1 gives the mean variance ratio from 10 simulations with the different methods. The choice of $\lambda = 3.0$ and $\sigma_{smooth} = 10$ arcmin are the tuning pa-

rameters that maximise the Pearson r value for GLIMPSE and Kaiser-Squires respectively. Though GLIMPSE and the Wiener filter reconstructions both have the same Pearson r value, the variance of the pixel values of the Wiener filter is much lower with respect to the underlying truth than is the case for GLIMPSE. This can also be seen in the reconstructions of figure 2.4, where the Wiener filter pixel values are closer to zero than the simulated true convergence.

The histogram of figure 2.6 shows, for one single example, the distributions matching what the results of the second column of table 2.1 describe. GLIMPSE outperforms the other methods at matching the variance of the underlying truth, however it still falls short. Also, all methods, including GLIMPSE, have distributions which are symmetric, unlike the asymmetric, heavy-tailed distribution of the true κ values.

Though GLIMPSE reconstructs maps with the 1-point distribution variance closest to the truth, it is also the only method to have convergence values dropping below the truth. These unphysical "negative peaks" can also be seen in the map reconstructions from data (figure 2.3) and from simulated catalogues (figure 2.4), and are likely to come from enforcement of sparsity for positive and negative wavelets equally. The physical motivation for GLIMPSE comes from a density field of superimposed halos. Though there should be no negative halos, negative wavelets are included to map the underdense regions, clearly at the expense of producing these very negative regions.

2.4.4 Phase reconstruction

The summation over all m modes at each ℓ multipole in the angular power spectrum (equation 2.24) loses all phase information; only the magnitudes are retained. This phase information corresponds to the spatial distribution of anisotropies. As the phases are dependent on the physical underlying structure, they contain information beyond what can be gained by 2-point statistics. Their retention is a well-motivated, desired property of a mass mapping reconstruction.

Inspired by Chapman et al. (2013), who use phases to test the reconstruction after foreground removal from simulated Epoch of Reonization 21-cm maps, we use the phase residual as a metric of success between our three methods.

The phase difference between the true map and the reconstruction is defined as

$$\begin{aligned}\Delta\theta_{\ell m} &= \theta^{truth}_{\ell m} - \theta^{\rm recon}_{\ell m} \\ &= \arg\left(a^{truth}_{\ell m}\right) - \arg\left(a^{\rm recon}_{\ell m}\right) , \\ &\text{where } \arg(z) = \arctan\left(\frac{\mathrm{Im}(z)}{\mathrm{Re}(z)}\right) .\end{aligned} \tag{2.37}$$

A small phase difference $\Delta\theta_{\ell m}$ between the truth and the reconstruction implies that the phase has been well reconstructed. For random variables drawn from a Gaussian distribution, this would correspond to a small standard deviation. Here, however, a Gaussian distribution would be an inappropriate choice as it assumes the data are defined on an unbounded Euclidean space.

The two dimensional data space of phase pairs, $\{\theta^{truth}_{\ell m}, \theta^{\rm recon}_{\ell m}\}$, is a torus, T^2, and the projected data space of the phase difference, $\Delta\theta_{\ell m}$, is a circle, S^1. On a circle, the maximum entropy, least informative, distribution for specified mean and variance is the von Mises (Jammalamadaka and Sengupta, 2001), which in one dimension is given by

$$Pr(\Delta\theta_{\ell m}|C,\mu) = \frac{1}{2\pi I_0(C)} \exp\left[C\cos(\Delta\theta_{\ell m} - \mu)\right] , \tag{2.38}$$

where I_0 is the modified Bessel function of order 0, and C is a concentration parameter. For $\mu = 0$, a large concentration parameter (analogous to $1/\sigma^2$) would correspond to a small dispersion in the phase reconstruction error. The aim is therefore to compare the inferred value of the concentration, C, between different mass mapping methods, with a larger value of C implying a better phase reconstruction.

By assuming that the error on the phase reconstruction is independent between phases, we can say that the phase differences, $\Delta\vec{\theta}$, are independent and identically distributed random variables, with a likelihood distribution given by

$$\begin{aligned}Pr(\Delta\vec{\theta}|C,\mu) &= \prod_{\ell m} \frac{1}{2\pi I_0(C)} \exp\left[C\cos(\Delta\theta_{\ell m} - \mu)\right] \\ &= \frac{1}{[2\pi I_0(C)]^n} \exp\left[C\sum_{\ell m}\cos(\Delta\theta_{\ell m} - \mu)\right] .\end{aligned} \tag{2.39}$$

As only the relative values of C are needed to compare different mass mapping methods, the full posterior distribution is not required. Additionally, any reasonable prior distribution,

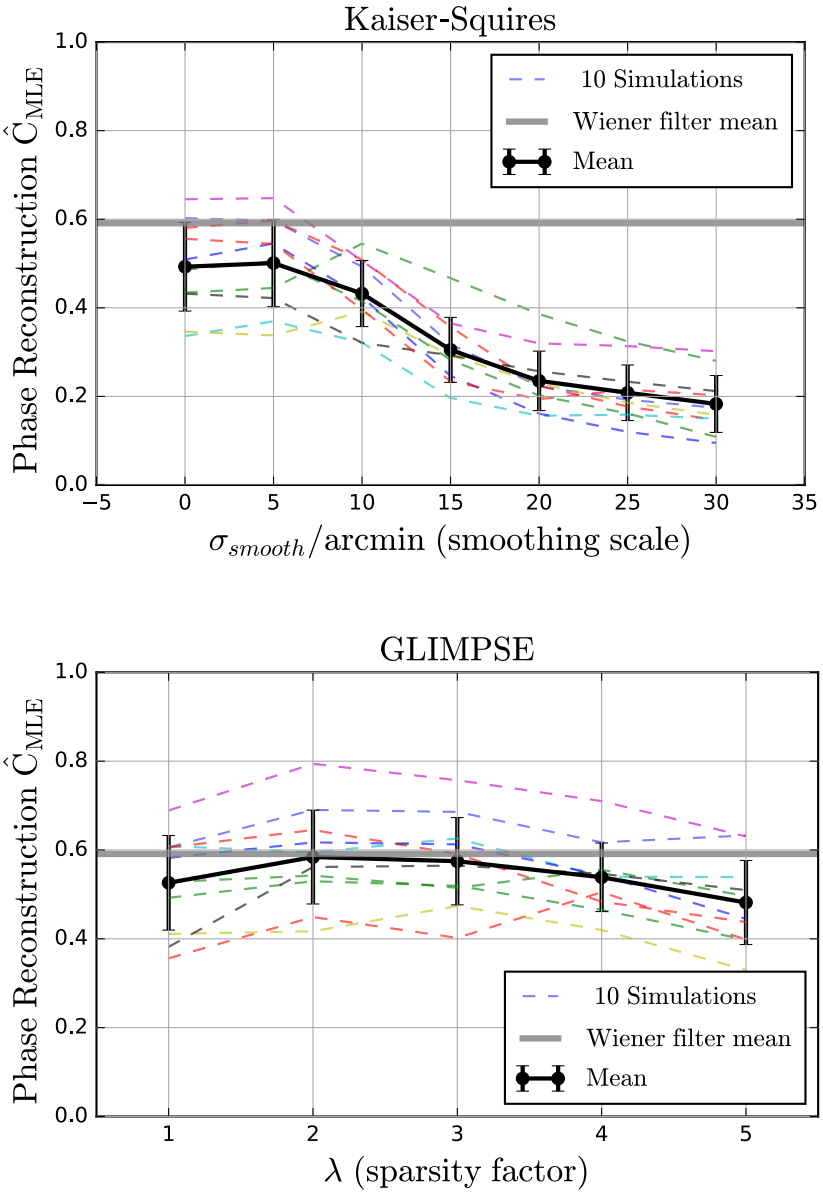

Figure 2.7: The maximum likelihood value of the concentration of the phase residual distribution, $\hat{C}_{\mathrm{MLE}}$, as described by equation 2.39. The $\hat{C}_{\mathrm{MLE}}$ values are shown for 10 different simulations and with Kaiser-Squires (*top panel*) at varying smoothing scale, σ_{smooth}, and GLIMPSE (*bottom panel*) at varying regularisation parameter λ. The phase reconstruction is best for $\sigma_{smooth} = 5$ arcmin and $\lambda = 3.0$ respectively.

$Pr(C)$, will be either flat or monotonically decreasing above zero, so the ranking of maps by the largest maximum likelihood value or maximum posterior value of C will be identical. For the purposes of this comparison the simpler maximum likelihood estimate, $\hat{C}_{\mathrm{MLE}}$, will therefore do.

We calculate the maximum likelihood values of μ and C by taking the spherical har-

Table 2.2: The mean over 10 simulations of the von Mises concentration maximum likelihood estimate, $\hat{C}_{\rm MLE}$, from phase residuals (equation 2.39).

Method	Phase reconstruction Concentration $\hat{C}_{\rm MLE}$
KS ($\sigma_{smooth} = 5$ arcmin)	0.501
Wiener filter	0.591
GLIMPSE ($\lambda = 3.0$)	0.584

monic transform of our `HEALPix` map to recover the $a_{\ell m}$ coefficients up to $\ell_{max} = 1024$, calculating the phase residual as defined by equation 2.37 between the truth and the reconstruction for each coefficient, and then maximising the likelihood (equation 2.39). The maximisation is performed using the `scipy` package BFGS algorithm (Byrd et al. 1995, Zhu et al. 1997, Morales and Nocedal 2011), using 3 random initialisation values to test for robustness.

Figure 2.7 show the results for the phase reconstruction from 10 simulations using Kaiser-Squires and GLIMPSE with varying tuning parameters. For Kaiser-Squires the mean phase reconstruction value, $\hat{C}_{\rm MLE}$, is maximised at $\sigma_{smooth} = 5.0$ arcmin. For larger smoothing scales the phase reconstruction quality drops, as phase information is lost. For GLIMPSE the mean phase reconstruction value, $\hat{C}_{\rm MLE}$, is maximised at $\lambda = 3.0$. The maximum value of $\hat{C}_{\rm MLE}$ is not particularly pronounced, and the $\hat{C}_{\rm MLE}$ values are quite stable over a range of λ.

Table 2.2 presents the mean values of $\hat{C}_{\rm MLE}$ with the best tuning parameters for the three map reconstruction methods. Both GLIMPSE and the Wiener filter do much better than Kaiser-Squires for reconstructing the phases. Though the variance from these 10 different simulations is large, the Wiener filter does slightly better than GLIMPSE, as can be seen in figure 2.7.

2.4.5 Peak Statistics

Peak statistics are a promising method for inferring cosmological parameters from data, as they access information beyond what can be inferred from 2-point correlation functions. Unlike higher order correlation functions, such as the bispectrum, peak statistics are inherently high signal-to-noise. They also probe the highly non-linear regions, where non-Gaussianity is greatest. The effect of masking is trivially taken into account by applying the identical mask to the suite of simulations used to construct a likelihood.

We cannot truly test which mass mapping method best constrains cosmology with

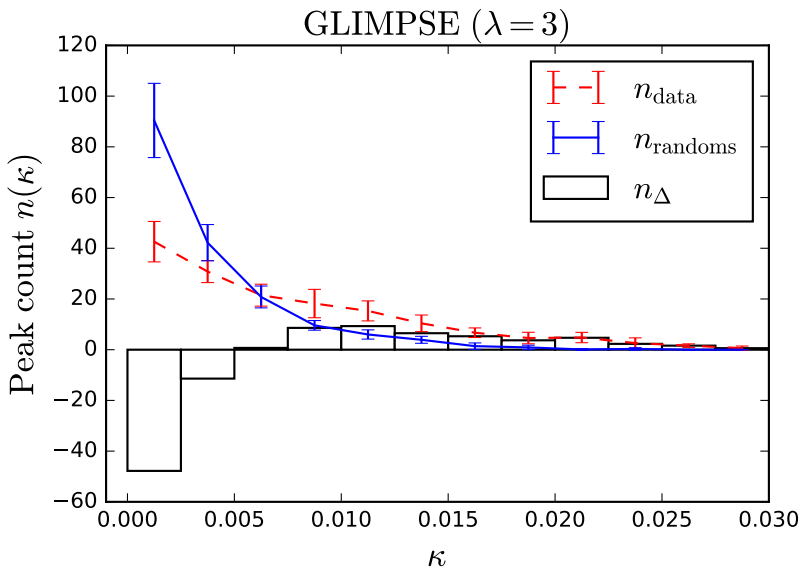

Figure 2.8: The mean $n_{\rm data}$, $n_{\rm randoms}$ and n_Δ functions with 10 simulated data catalogues and 10 random catalogues from GLIMPSE ($\lambda = 3$) reconstructions. n_Δ is defined in equation 2.40. Errorbars are standard deviation sample estimates from the 10 simulations, and are consistent with Poissonian noise. Figure 2.9 shows the signal-to-noise of n_Δ using the estimated Poissonian noise for different reconstruction methods and tuning parameters.

the statistics of density peaks without fully deriving the posterior probability distributions of cosmological parameters. It is possible to test which method returns peaks which are distinguishable from noise and at which convergence values. Distinguishing a large number of peaks from noise at high values of κ would mean the map is reconstructing the non-linear regions well.

For a given convergence map, we can define a function, $n(\kappa)$, that gives the number of peaks as a function of convergence. For a given mass reconstruction method we can compare the peaks in reconstructions from simulated data with the peaks in reconstructions from catalogues of "randoms", with shape noise but no weak lensing shear signal (equivalent to $\gamma = 0$ in equation 2.9). If a given map from data or from a simulated catalogue has the same $n(\kappa)$ as the random catalogues, then the mass mapping method used has been useless for peak statistics. On the other hand, if the map from data or simulation has a very different $n(\kappa)$ function to that from the reconstruction from the random catalogue, then the map reconstruction method has recovered "true", physical κ peaks.

In the DES SV cosmology constraints from peak statistics, Kacprzak et al. (2016) use this difference as the data vector used to constrain cosmology,

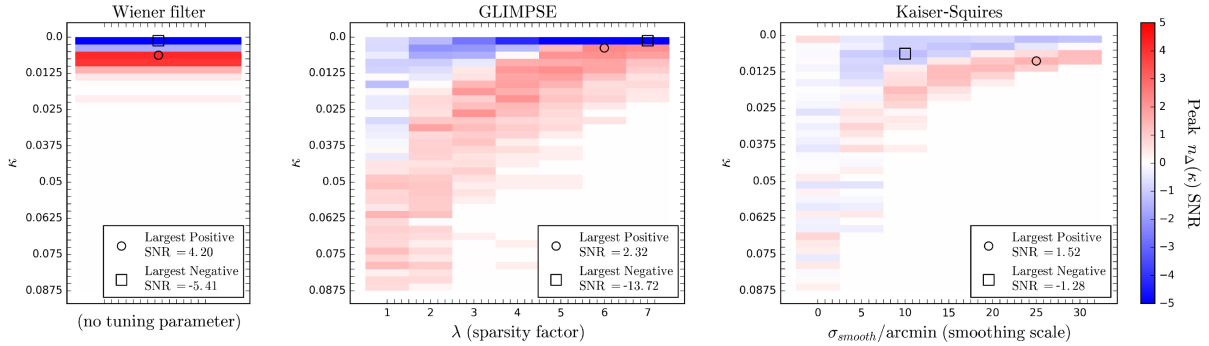

Figure 2.9: The estimated signal-to-noise (SNR) of $n_\Delta(\kappa)$ (equation 2.40) using 10 simulated data catalogues and 10 random catalogues with the three different mass mapping methods. Kaiser-Squires and GLIMPSE maps can be tuned by their respective parameters. The width of the left (Wiener filter) panel is purely nominal; it does not actually have a "flat structure", just no parameter to tune.

$$n_\Delta(\kappa_i) = n_{\rm data}(\kappa_i) - n_{\rm randoms}(\kappa_i)\ . \tag{2.40}$$

This function is far from zero at a given κ if there is a large difference between the number of peaks counted in maps reconstructed (a) from data and (b) from random catalogues.

It is reasonable to believe that the number of peaks, $n(\kappa_i)$, in the i^{th} bin, κ_i, is drawn from a Poisson distribution. The difference between two Poissonian random variables follows the Skellam distribution. Using this distribution, we expect the difference in the number peaks in maps from data and from random catalogues to have a mean given by

$$\mu_\Delta(\kappa_i) = \mu_{\rm data}(\kappa_i) - \mu_{\rm randoms}(\kappa_i)\ , \tag{2.41}$$

and a variance given by

$$\sigma^2_\Delta(\kappa_i) = \mu_{\rm data}(\kappa_i) + \mu_{\rm randoms}(\kappa_i). \tag{2.42}$$

We can therefore define a peak signal-to-noise estimate

$$\mathrm{SNR}(\kappa_i) = \frac{\mu_\Delta(\kappa_i)}{\sqrt{\sigma^2_\Delta(\kappa_i)}}\ . \tag{2.43}$$

Figure 2.8 shows $n_{\rm data}$, $n_{\rm randoms}$, and n_Δ from GLIMPSE ($\lambda = 3$) from 10 simulations and 10 random catalogues. Here we define a peak as a local maxima in the `HEALPix` map. Across different methods and smoothing parameters, the predicted variance from equa-

tion 2.42 matches well with the estimated sample variance, verifying that the peak distribution is indeed Poissonian for a given κ.

Figure 2.9 shows the peak signal-to-noise (SNR) estimates from 10 simulations and from 10 random catalogues as a function of κ and smoothing scale, for Kaiser-Squires, or λ, for GLIMPSE. As the peaks in the maps from data have higher convergence values than those from random catalogues, the SNR(κ) function is negative for low values of κ.

In the figures, the GLIMPSE reconstruction gives better signal-to-noise estimates on the peaks than does the Kaiser-Squires reconstruction. For Kaiser-Squires, the largest positive and negative signal-to-noise values are 1.52 and -1.28. For GLIMPSE, the largest positive and negative are signal-to-noise values of 2.32 and -13.72. For the Wiener filter these values are 4.20 and -5.41.

The Wiener filter therefore has the highest signal-to-noise of the peak function $n_\Delta(\kappa)$, though the κ values of these peaks are very low. As can be seen in the reconstruction from the SV data in figure 2.3, the pixel values of the Wiener filter are much closer to zero. This is reflected in the peak statistic signal-to-noise values. In the left panel of figure 2.9, the Wiener filter detects negligibly few peaks with $\kappa > 0.0125$, whereas GLIMPSE detects peaks with positive signal-to-noise up to higher values of κ. It is at these high values where the non-Gaussian information due to non-linear structure formation can be probed.

2.4.6 Foreground Clusters

Comparisons with foreground clusters of galaxies is an independent test of the mass map reconstructions, as it uses data (unlike our tests on simulations).

In figure 2.10 the redMaPPer clusters described in section 2.3.2 are overlaid on the DES SV κ map reconstructions shown in figure 2.3. The maps show good spatial correlation between the locations of the clusters and the κ peaks in the map.

The size of a cluster marker is the effective lensed cluster richness λ_{RM}^{eff}, rather than the redMaPPer cluster richness. This concept is adapted from the definition of κ_g presented in Chang et al. (2015). For a given cluster, this is defined as

$$\lambda_{RM}^{eff} = \frac{p(z)\omega(z)}{a(z)} \times \lambda_{RM} \times \frac{\langle \lambda_{RM} \rangle}{\langle \lambda_{RM}^{eff} \rangle} , \tag{2.44}$$

where z is the redshift of the cluster, $p(z)$ is the lensing efficiency at the location of the cluster (see figure 2.1), and $\omega(z)$ is the comoving distance to the cluster (so that the first term matches the integrand of equation 2.6). The final term normalises the mean, where

Table 2.3: The Pearson correlation coefficient value, r, between effective richness, λ_{RM}^{eff}, of the foreground redMaPPer clusters and the reconstructed convergence map at the location of each galaxy cluster.

Method	redMaPPer Cluster λ_{RM}^{eff} Pearson r
KS ($\sigma_{smooth} = 10$ arcmin)	0.116
Wiener filter	0.129
GLIMPSE ($\lambda = 3.0$)	0.152

$\langle\lambda_{RM}\rangle$ is the average richness over all galaxy clusters. The effective lensed cluster richness gives the richness as "seen" by the lensing effect, where clusters at the peak of the lensing efficiency should contribute more to the κ map. We therefore calculate the correlation between λ_{RM}^{eff} for each cluster and the reconstructed κ value at the cluster centre.

This method does not take into account multiple clusters overlapping in a given line of sight. In figure 2.10, many small clusters overlap on large peaks in the reconstructed κ map. The naive one-to-one correspondence between cluster and κ would mistake this for an excess of κ in the reconstruction. However, all methods will suffer equally from this assumption. A more thorough treatment of this overlapping effect is left for future work.

Table 2.3 presents the Pearson correlation coefficient r between the λ_{RM}^{eff} value of each cluster and the $\kappa^{\rm recon}$ value at the corresponding pixel. The tuning parameters for Kaiser-Squires and GLIMPSE are chosen to maximise the Pearson correlation coefficient r between the reconstruction and the truth from simulations (see section 2.4.1).

Though both GLIMPSE and the Wiener filter take into account the noise and the mask in the data, and therefore do better than Kaiser-Squires, the GLIMPSE reconstructions show higher correlation with the effective richness of the foreground clusters than do the Wiener filter reconstructions. This is no surprise, as GLIMPSE is expected to do better at reconstructing non-Gaussian κ, which would correspond to the non-linear matter structures in which clusters of galaxies form.

2.5 Conclusions

In this work we have presented convergence map reconstructions using the public DES SV shear data with three different methods: Kaiser-Squires, Wiener filter, and GLIMPSE. Kaiser-Squires is a simple inversion from shear to convergence, whereas the Wiener filter and GLIMPSE use prior knowledge about the true convergence to help regularise the recon-

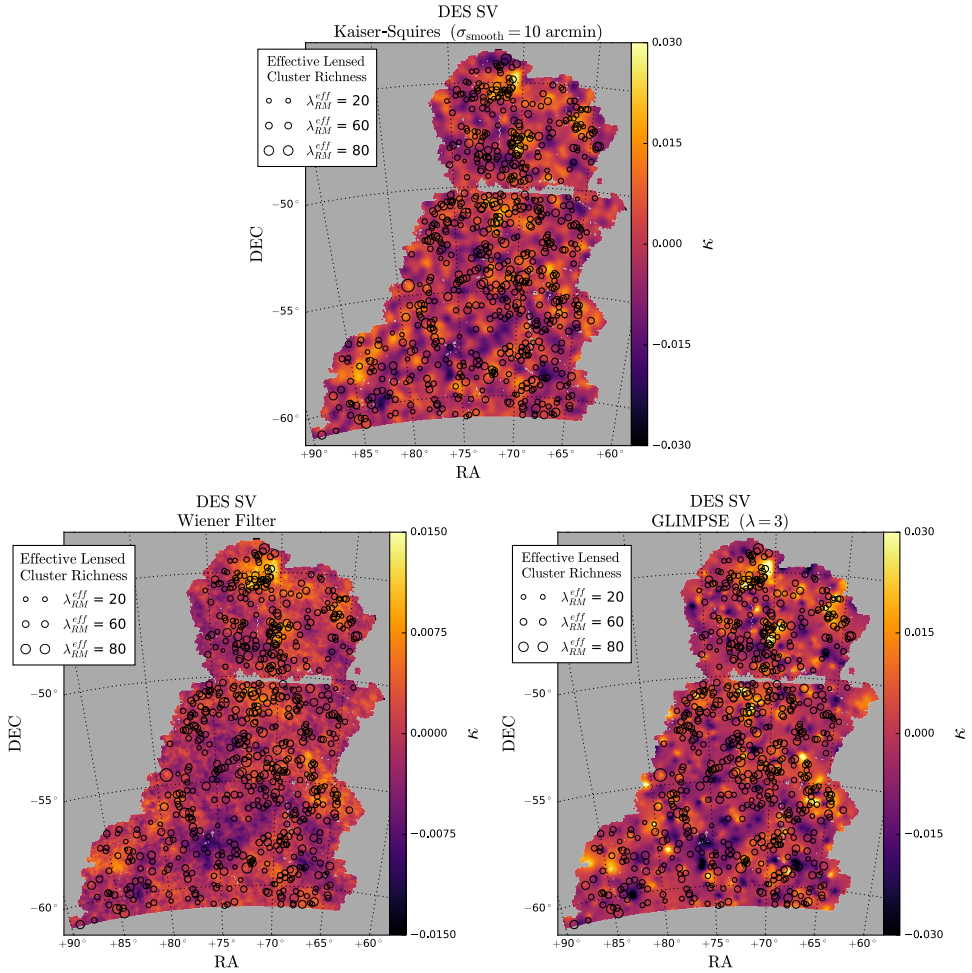

Figure 2.10: The mass map reconstruction from DES SV shear data with the three different methods, as presented in figure 2.3), with the locations of redMaPPer clusters overlaid. The size of the cluster marker is the effective richness of the cluster, as defined in equation 2.44. **Note** that the colour scale for the Wiener filter is less than that for the other reconstructions, as the pixel values are closer to zero.

struction and to reduce the effects of noise and missing data. The Wiener filter is a Bayesian MAP estimate if the signal and noise are Gaussian and the respective covariance matrices are known. The GLIMPSE method enforces a sparsity-promoting l_1 norm in a wavelet space where the wavelets represent positive, isotropic, quasi-spherical objects well. GLIMPSE is therefore expected to do well at reconstructing non-linear structures. The Wiener filter and GLIMPSE therefore aim to reconstruct different regimes: the linear and non-linear density field.

The three methods were applied to realistic simulations of the DES SV shear data, for which an underlying true convergence is known. Using these simulations we are also able to tune the Kaiser-Squires smoothing scale, σ_{smooth}, and the GLIMPSE sparsity regularisation

parameter, λ.

With these simulations we measure the Pearson correlation coefficient, r, between the truth and the reconstruction with different methods. Compared to the Kaiser-Squires reconstructions we find a 12 per cent improvement in Pearson correlation with both the Wiener filter and GLIMPSE. The tuning parameters of $\sigma_{smooth} = 10$ arcmin for Kaiser-Squires and $\lambda = 3$ for GLIMPSE maximise the Pearson correlation. We also measure the variance of the 1-point distribution of the reconstructed convergence. The Wiener filter suppresses the variance to 6.3 per cent of the truth, Kaiser-Squires to 37 per cent and GLIMPSE to 50 per cent of the truth. The tunable parameters here were those which maximised the Pearson correlation with the truth.

A large motivation for creating these maps is to reconstruct the convergence while still retaining the non-Gaussian information (which cannot be accessed with 2-point statistics such as the power spectrum). As such, we test the reconstruction of the harmonic phases, which is averaged out in the power spectrum, and the signal-to-noise of a peak statistic data vector, which is a popular probe of non-Gaussian information. The phase residuals between the truth and the reconstruction have the highest von Mises concentration with the Wiener filter ($\hat{C}_{\rm MLE} = 0.591$), with the GLIMPSE reconstruction performing comparably ($\hat{C}_{\rm MLE} = 0.584$). Both methods outperformed the Kaiser-Squires reconstruction ($\hat{C}_{\rm MLE} = 0.501$).

With realistic data vectors for peak statistics generated from simulations, the maximum signal-to-noise value was increased by a factor of 3.5 for the Wiener filter and by a factor of 9 for GLIMPSE, compared to Kaiser-Squires. The signal-to-noise of the peak statistic data vector ($n_\Delta(\kappa)$) is shown in figure 2.9, where GLIMPSE has significant signal-to-noise with high convergence peaks, where non-linearities in the underlying density field are highest. We predict these high value peaks are most useful for constraining cosmology beyond Gaussianity. In order to constrain cosmology with these different reconstruction methods, realistic simulations with different cosmological parameters or models must be used and the same reconstruction method should be applied to the simulations and data. As seen from our results, different reconstruction methods can produce convergence maps with different properties.

Finally, we switched from using simulations to instead using real observations (DES SV data). Here we measured the correlation between the reconstructed maps and the effective richness of the foreground redMaPPer clusters (this is the cluster richness as "seen" by

the lensing effect). Table 2.3 shows the results. Compared with Kaiser-Squires, the Wiener filter shows a 18 per cent increase and GLIMPSE shows a 32 per cent increase in correlation. This demonstrates with independent, cosmological data the ability of the methods to reconstruct non-linear structures.

The metrics we have used for comparing the three reconstruction methods are generic, and they have been inspired by recent applications of weak lensing mass maps to cosmological studies (e.g. Chang et al. 2016, Kacprzak et al. 2016). These metrics may not be optimal for evaluating every application of mass maps. Future studies can compare the efficiency of the three and other methods in end-to-end analyses; for example, with the estimation of cosmological parameters or identification of galaxy clusters.

Applying the Wiener filter and GLIMPSE methods to the DES Year 1 (Y1) shear catalogue would require extensions of the methods to account for the curved sky at large angular scales. The Y1 data covers ≈ 1500 deg^2 and contains $\approx 34,800,000$ galaxies, so is a large increase in data volume from DES SV. This modification has already been done with an extension of Kaiser-Squires to the sphere by Chang et al. (2017) for the Y1 DES data. These extensions would also be useful for the upcoming ≈ 5000 deg^2 DES Y3 shear catalogue.

Of future interest would be to use the Wiener filter or GLIMPSE convergence maps for scientific results, as we have shown that they reconstruct the convergence better than Kaiser-Squires according to many different metrics.

We have made our map reconstructions (as shown in figure 2.3) available at `des.ncsa.illinois.edu/releases/sva1`.

2.6 Supplementary material: indicator function

We define the indicator function $i_{\mathscr{C}}$ (as used in equation 2.29) of a set $\mathscr{C}$ as

$$i_{\mathscr{C}}(x) = \begin{cases} 0 & \text{if } x \in \mathscr{C} \\ +\infty & \text{otherwise .} \end{cases} \tag{2.45}$$

Chapter 3

Fast sampling from Wiener posteriors with Dataflow Engines

This Chapter is based on Jeffrey, Heavens, & Fortio 2018: *Fast sampling from Wiener posteriors for image data with dataflow engines.* Astronomy and computing 25 (2018): 230-237

We use Dataflow Engines (DFE) to construct an efficient Wiener filter of noisy and incomplete image data, and to quickly draw probabilistic samples of the compatible true underlying images from the Wiener posterior. Dataflow computing is a powerful approach using reconfigurable hardware, which can be deeply pipelined and is intrinsically parallel. The unique Wiener-filtered image is the minimum-variance linear estimate of the true image (if the signal and noise covariances are known) and the most probable true image (if the signal and noise are Gaussian distributed). However, many images are compatible with the data with different probabilities, given by the analytic posterior probability distribution referred to as the Wiener posterior. The DFE code also draws large numbers of samples of true images from this posterior, which allows for further statistical analysis. Naive computation of the Wiener-filtered image is impractical for large datasets, as it scales as n^3, where n is the number of pixels. We use a messenger field algorithm, which is well suited to a DFE implementation, to draw samples from the Wiener posterior, that is, with the correct probability we draw samples of noiseless images that are compatible with the observed noisy image. The Wiener-filtered image can be obtained by a trivial modification of the algorithm. We demonstrate a lower bound on the speed-up, from drawing 10^5 samples of a 128^2 image, of 11.3 ± 0.8 with 8 DFEs in a 1U `MPC-X` box when compared with a 1U server presenting 32 CPU threads. We also discuss a potential application in astronomy, to provide better dark matter maps and improved determination of the parameters of the Universe.

3.1 Introduction

Dataflow computing has recently aided the significant acceleration of many computationally-intensive and data-intensive problems. This Chapter discusses the use of Dataflow Engines (DFEs) for sampling realisations of noise-free images from the Wiener posterior distribution given noisy and incomplete data, with particular applicability to astronomy and cosmology.

The Wiener filter (Wiener, 1949) is a useful statistical tool in many image analyses, as it is a minimum variance linear filter, and moreover the filtered data are also the *maximum a posteriori* (MAP) values if the data have Gaussian signal and noise. To be more specific, if the covariance matrices of the noise and signal are known, then the Wiener filtered image has the smallest variance of any linear-filtered image. Mathematically it is straightforward to write down the expression for the Wiener-filtered image, and the covariance of compatible images, but evaluation is problematic as it involves the inversion of large matrices that are in general non-diagonal. As image datasets become larger, naive Wiener methods become unfeasible (requiring approximations such as re-binning to larger pixels or assuming white noise).

By using messenger field algorithms (described in section 3.2.2) the Wiener image and posterior can be computed feasibly, with no need to simplify the existing algorithms. Furthermore, the repeated operations inherent in drawing samples from the Wiener posterior lend themselves to efficient computation on DFEs, and we demonstrate that by a comparison with an implementation on multiple CPUs.

3.1.1 Data model

Although the typical applications of Wiener filters involve 2D image data, the formalism is general. In any case, we arrange the 2D pixel data as a list, and thus describe it by a data vector $\mathbf{d}$, and the true image is similarly described by a vector $\mathbf{s}$.

Our linear data model assumes that data $\mathbf{d}$ and true signal $\mathbf{s}$ are related by

$$\mathbf{d} = \mathbf{A}\mathbf{s} + \mathbf{n} \,. \tag{3.1}$$

where $\mathbf{n}$ is random noise, and there is a known linear operator matrix $\mathbf{A}$, which in the simplest case is just the identity matrix.

The Wiener filter $\mathbf{W}$ (Wiener 1949, Zaroubi et al. 1995) is given by

$$\mathbf{W} = \mathbf{S}\mathbf{A}^\dagger \left(\mathbf{A}\mathbf{S}\mathbf{A}^\dagger + \mathbf{N}\right)^{-1} , \tag{3.2}$$

and the Wiener filtered solution, which is the minimum-variance linearly-obtained solution for the true image, is

$$\mathbf{s}_W = \mathbf{W}\mathbf{d} . \tag{3.3}$$

In these equations, $\mathbf{S} = \langle \mathbf{s}\mathbf{s}^\dagger \rangle$ and $\mathbf{N} = \langle \mathbf{n}\mathbf{n}^\dagger \rangle$ are the signal and noise covariance matrices respectively, which are assumed to be known, and we have assumed that $\langle \mathbf{s} \rangle = \langle \mathbf{n} \rangle = 0$ for simplicity (this can easily be relaxed). The angle brackets indicate the expectation value, equal to the average over infinitely many realisations of the signal for ergodic fields. If, as we will assume, the pixel noise is uncorrelated, then $\mathbf{N}$ is diagonal in pixel space. In addition to pixel noise, missing data in a given pixel can be incorporated into the Wiener filter by setting the pixel noise variance to infinity.

As mentioned in section 3.2, the Wiener filter reconstruction, $\mathbf{s}_W$, is the linear minimum variance filter for a given $\mathbf{S}$ and $\mathbf{N}$ regardless of the statistical properties of either the signal or the noise. Note that the Wiener filtered image variance is biased low; e.g. high intensity pixels are suppressed. For Gaussian signal and noise, the Wiener filter additionally becomes the MAP estimate. In addition to computing the MAP estimate, for statistical purposes it is often useful to draw samples of maps, that are compatible with the data, with the appropriate probability. These can be used for subsequent statistical analysis of the true image, such as determining the uncertainty in a given pixel. This is discussed further in section 3.2.1.

Calculation of the Wiener filter is challenging due to the inversion of covariance matrices, which may not be diagonal, and can become prohibitively time consuming for large images, especially when one notes that for an $N \times N$ image, the matrices are $N^2 \times N^2$ in size.

In some applications the signal is statistically homogeneous, leading to a diagonal signal covariance in the Fourier/harmonic domain, which leads to a route to a solution that does not involve the inversion of large non-diagonal matrices (Elsner and Wandelt, 2013). This is not trivial, since although independent noise has a diagonal covariance matrix in pixel space, it is not diagonal in harmonic space if the dataset has varying noise variance and is thus heteroscedastic. This situation automatically arises if there are missing data,

but not only in this case. Therefore, in general there is no natural basis in which both the signal and noise covariance matrices are sparse. It is possible to take advantage of the bases in which the covariance matrices are sparse by using algorithms that employ so-called "messenger fields" (Elsner and Wandelt, 2013) to convey information between harmonic and pixel space.

The messenger field class of algorithms is highly suited to a Dataflow implementation. Using reconfigurable hardware accelerators rather than CPUs helps to deal iteratively with large volumes of data. DFEs have recently been successfully applied to a wide range of scientific problems, including geoscience (Gan et al., 2017), fluid-dynamics (Düben et al., 2015), artificial neural networks (Liang et al., 2018), quantum chemistry (Cooper et al., 2017), and genomics (Arram et al., 2015).

In section 3.2, we describe the Wiener filter in a Bayesian framework, and show how messenger fields are used to draw samples from the Wiener posterior probability distribution. In section 3.3, we describe Dataflow computing and present our implementation of the Wiener sampler. We present the results in section 3.4. In section 3.5, we describe our motivation for this work as an application to upcoming large cosmology surveys.

3.2 Theoretical Background

3.2.1 Wiener Posterior

For the linear model of equation 3.1, the Wiener filter, with $\mathbf{W}$ given by equation 3.2, is a linear operator which minimises the variance

$$V = \langle (\mathbf{Wd} - \mathbf{s})^{\dagger} (\mathbf{Wd} - \mathbf{s}) \rangle \ . \tag{3.4}$$

From a different starting point, for the Wiener posterior, we begin by assuming a Gaussian likelihood for the pixel noise[1] (Jasche and Lavaux, 2015):

$$Pr(\mathbf{d}|\mathbf{s},\mathbf{N}) = \frac{1}{\sqrt{(\det 2\pi \mathbf{N})}} \exp\Big[-\frac{1}{2} (\mathbf{d} - \mathbf{As})^{\dagger} \mathbf{N}^{-1} (\mathbf{d} - \mathbf{As}) \Big] \ . \tag{3.5}$$

Assuming that the prior on the signal is that of a Gaussian random field,

$$Pr(\mathbf{s}|\mathbf{S}) = \frac{1}{\sqrt{(\det 2\pi \mathbf{S})}} \exp\Big[-\frac{1}{2} \mathbf{s}^{\dagger} \mathbf{S}^{-1} \mathbf{s} \Big] \ , \tag{3.6}$$

[1] We can also argue that if only the covariance and the mean is known, the Gaussian distribution is most appropriate to assume, as it is the maximum entropy distribution.

then using Bayes' theorem and the fact that $Pr(\mathbf{d}|\mathbf{S},\mathbf{s},\mathbf{N}) = Pr(\mathbf{d}|\mathbf{s},\mathbf{N})$, the full Wiener posterior can be found:

$$\begin{aligned} Pr(\mathbf{s}|\mathbf{S},\mathbf{N},\mathbf{d}) &= \frac{Pr(\mathbf{d}|\mathbf{s},\mathbf{N})Pr(\mathbf{s}|\mathbf{S},\mathbf{N})}{Pr(\mathbf{d}|\mathbf{N})} \\ &= \frac{1}{\sqrt{(\det 2\pi\mathbf{S})}}\frac{1}{\sqrt{(\det 2\pi\mathbf{N})}}\exp\Big[-\frac{1}{2}\mathbf{s}^\dagger\mathbf{S}^{-1}\mathbf{s} - \frac{1}{2}(\mathbf{d}-\mathbf{A}\mathbf{s})^\dagger\mathbf{N}^{-1}(\mathbf{d}-\mathbf{A}\mathbf{s})\Big] \\ &\propto \exp\Big[-\frac{1}{2}(\mathbf{s}-\mathbf{W}\mathbf{d})^\dagger(\mathbf{S}^{-1}+\mathbf{A}^\dagger\mathbf{N}^{-1}\mathbf{A})(\mathbf{s}-\mathbf{W}\mathbf{d})\Big]\,. \end{aligned} \tag{3.7}$$

Here we see that the *maximum a posteriori* (MAP) solution is indeed that of the Wiener reconstruction, $\mathbf{s} = \mathbf{W}\mathbf{d}$.

If we can handle the large matrices, realisations of the true underlying signal image $\mathbf{s}$ can be drawn from the posterior distribution $Pr(\mathbf{s}|\mathbf{S},\mathbf{d})$. The expected mean of these samples is the Wiener-filtered image. Drawing samples from the Wiener posterior clearly also suffers from the need to invert large matrices with no natural sparse basis.

Progress can be made for signal images with statistical properties that are independent of pixel position $\mathbf{x}$ (i.e. statistically homogeneous signals), for in this case, the discrete Fourier transform of the image $\mathbf{s}_\mathbf{x}$,

$$\tilde{s}_\mathbf{k} = \sum_\mathbf{x} s_\mathbf{x} e^{-i\mathbf{k}\cdot\mathbf{x}} \tag{3.8}$$

has a diagonal covariance matrix,

$$\langle \tilde{s}_\mathbf{k}\tilde{s}^*_{\mathbf{k}'}\rangle = P(k)\delta_{\mathbf{k}\mathbf{k}'} \tag{3.9}$$

and $\delta_{\mathbf{k}\mathbf{k}'}$ is a Kronecker delta for the discrete 2D wavenumbers $\mathbf{k}$ and $\mathbf{k}'$, and $P(k)$ is the power spectrum, which depends only on the magnitude $k \equiv |\mathbf{k}|$. The covariance matrix $\mathbf{S}$ for the signal is diagonal, with entries given by the appropriate $P(k)$.

3.2.2 Messenger Fields

The messenger field approach splits the problem into two, performing some operations in harmonic space and some in pixel space, transferring the information using an extra field, $\mathbf{t}$, called the messenger field, whose covariance matrix is diagonal in both spaces. The method takes advantage of the diagonal signal covariance matrix in harmonic space and the diagonal

noise covariance matrix in pixel space, such that no matrices need to be inverted in a basis in which they are dense.

This field is defined to have zero mean and a covariance matrix proportional to the identity matrix, $\langle \mathbf{t}\mathbf{t}^\dagger \rangle \propto \mathbf{I}$, which will always be diagonal in both harmonic and pixel spaces. The Markov Chain Monte Carlo (MCMC) algorithm used in Jasche and Lavaux (2015) is a method which uses the messenger field to draw samples from the Wiener posterior, without inversion of non-diagonal covariance matrices, requiring instead repeated Fourier transforms and inverse Fourier transforms. The algorithm is presented in Algorithm 1. In the limit of large numbers of iterations, this unconditionally converges to drawing samples from the desired distribution.

A sufficient number of samples from the Wiener posterior probability distribution can characterize the statistical properties of the underlying signal given some data.

Algorithm 1 Messenger Field Wiener Sampler: an iterative method to draw sample signal images from a Wiener posterior distribution using messenger fields (Jasche and Lavaux, 2015)

1: **procedure** SAMPLER
2: for t_i in $\mathbf{t}$:
3: $t_i = \mu_i^t + \sqrt{(\sigma_i^t)^2}\, G(0,1)$
4: $\hat{\mathbf{t}} = \mathscr{F}_{2D}(\mathbf{t})$
5: for $\hat{s}_k$ in $\hat{\mathbf{s}}$:
6: $\hat{s}_k = \mu_k^s + \sqrt{(\sigma_k^{\hat{s}})^2}\, G(0,1)$
7: $\mathbf{s} = \mathscr{F}_{2D}^{-1}(\hat{\mathbf{s}})$
8: Return $\mathbf{s}$
9: GOTO line 2
10: **end procedure**

Definitions:

- $\mu_i^t = \frac{T_i}{T_i A_i^2 + \bar{N}_i} A_i d_i + \frac{\bar{N}_i}{T_i A_i^2 + \bar{N}_i} s_i$ if $A_i^2 > 0$
- $(\sigma_i^t)^2 = \frac{T_i \bar{N}_i}{T_i A_i^2 + \bar{N}_i}$ if $A_i^2 > 0$
- $\mu_k^{\hat{s}} = \frac{\hat{S}_k}{\hat{S}_k + \hat{T}_k} \hat{t}_k$
- $(\sigma_k^{\hat{t}})^2 = \frac{\hat{S}_k \hat{T}_k}{\hat{S}_k + \hat{T}_k}$
- $\mathbf{T} = \min\big((\mathbf{A}^{-1})^\dagger \mathbf{N} (\mathbf{A}^{-1})\big)\mathbf{I}$
- $\bar{\mathbf{N}} = \mathbf{N} - \mathbf{A}^\dagger \mathbf{T} \mathbf{A}$
- $G(0,1)$ is a zero-mean Gaussian random variate with unit variance.
- $\mathscr{F}_{2D}$ is the 2D Fourier transform and $\mathscr{F}_{2D}^{-1}$ its inverse.

By replacing the random variates in Algorithm 1 with zero ($G(0,1) \to 0$), the iteration

outputs converge to the (unsampled) Wiener filter reconstruction (equation 3.2). This was first described by Elsner and Wandelt (2013) in the first use of messenger fields. With this small change the code provides Wiener-filtered images, rather than samples from the Wiener posterior.

For each calculation of a Wiener filter or sample drawn from the Wiener posterior, $\mathcal{O}(n^3)$ operations are required for a length n data vector using a naive approach. Using the messenger field algorithm, this reduces to $\mathcal{O}(n \log n)$ for covariance matrices that are diagonal in their respective domains. The naive approach is bottlenecked by the matrix inversion and the messenger field approach is bottlenecked by the harmonic/Fourier transform.

3.3 DFE Implementation

3.3.1 DFE System

The standard computing paradigm in the present day still follows the outline of the von Neumann model, often called Control Flow. In a standard setup a Central Processing Unit (CPU) carries out computational operations with data and instructions provided by memory, usually Random Access Memory (RAM). Data and instructions are iteratively passed between memory and CPU.

Dataflow Engines (DFEs) use reconfigurable hardware rather than CPUs to represent a static description of an algorithm with deep hardware pipelines consisting of a series of standard arithmetic and logic operations. DFEs do not need to continually get new instructions from the memory (Pell et al., 2013). They are therefore intrinsically parallel. The Wiener sampling problem described above has a high volume of data with highly deterministic computation (few "if" statements), so is well suited to DFEs.

Unlike standard CPU-based High Performance Computing (HPC) platforms, DFEs can be reconfigured on occasion to the need of a given algorithm or dataset. For the cost of an initial build time ($\mathcal{O}$(hours)), the speed and efficiency at runtime is improved. These systems allow greater flexibility with memory, data type, and clock frequency.

For example, higher clock frequencies can lead to shorter run time of the compute kernels instantiated on the DFE. This can yield faster execution if the algorithm is compute bound. However, for higher clock frequencies it becomes more difficult to build the reconfiguration bitstream, so the clock frequency can be chosen optimally for a given algorithm.

The CPU code for managing a DFE can be written in `C` or `C++` and runs on a host (a traditional control flow machine). For the DFE, the software is written in Java-like code,

which is compiled into the reconfiguration file for the hardware chip. This turns the DFE into a problem-specific hardware accelerator.

Once reconfigured, the DFE accepts data streaming and compute action calls launched by the host CPU code. A single DFE is a PCIe card that can either be available locally on a CPU server or be mounted in a Maxeler `MPC-X`: a CPU-free 1U server appliance hosting up to 8 DFEs, which is connected to host CPU servers by an infiniband network. Each DFE carries a chip with large amount of reconfigurable logic and on-chip resources (e.g. a Field-Programmable Gate Array, FPGA) with up to 96 GB of on-board DRAM storage. The MAX4 generation cards available to the authors are MAX4 Maia DFEs with Altera Stratix V FPGA and 48 GB of DRAM. Integration of CPU and DFE codes is done by the dedicated compiler as described in Kos et al. (2015).

Dataflow Engines allow user-friendly control over the features of the underlying hardware, so the hardware description can be optimally designed and built for the algorithm at hand. This can lead to large speed-ups at runtime compared to the same algorithm's implementation on a comparative CPU platform.

Time to complete a task is also only one metric of performance among other metrics. Lower clock frequencies mean that DFEs use less power than conventional CPU machines (Gan et al., 2017). Usually FPGAs use an order of magnitude less power than CPUs (Liang et al., 2018). In many applications, it is therefore more cost efficient to use DFEs as it allows more science per Watt.

Another commonly used and increasingly popular alternative to CPU hardware are graphics processing units (GPU), which gain acceleration for vectorized problems using "single instruction, multiple data" (SIMD) architectures and high-clock frequency (Liang et al., 2018). However, they are disadvantaged by their high energy cost. CPUs are more efficient than GPUs, and, as discussed, FPGAs are in turn more efficient than CPUs. GPUs additionally do not benefit from the flexibility that allows reconfigurable DFEs to tailor to a specific algorithm. Their hardware cannot be optimally designed for a given problem.

3.3.2 Implementation

We show the steps taken to generate a typical simulated dataset with the desired properties in figure 3.2. To simulate underlying signals $\mathbf{s}$, we generate realisations of square, two-dimensional images, which are in this case real, zero-mean, Gaussian random fields with known power spectra. The real and imaginary parts of $s_{\mathbf{k}}$ are each drawn randomly from

Gaussian distributions with variance $P(k)/2$, and reality of the signal is enforced by $s^*_{\mathbf{k}} = s_{-\mathbf{k}}$. We simulate square signal maps with 128^2 pixels.

The image, and therefore the vectors $\mathbf{k}$ and $\mathbf{x}$, are two dimensional, so the transforms employs a 2D DFT. In practice the fast Fourier transform (FFT) algorithm (Cooley and Tukey, 1965) is used to evaluate the coefficients.

The datasets are generated according to the linear model of equation 3.1. For simplicity, we do not apply the linear operator (setting $\mathbf{A} = \mathbf{I}$), though this could be included for a given application. The first panel of figure 3.2 shows an initial power spectrum, $P(k)$, from which we generate our real, Gaussian field as the signal map.

The noise is independent between pixels and is drawn from a Gaussian distribution where the noise variance varies across the data. We assume that the noise variance is known. We mask some of the pixels to represent missing data. The Wiener filter and the Wiener posterior treat the missing data as a special case of infinite noise. Infinite noise variance, in the region of the missing data, is set to be 10^8, as an effective infinity.

On both CPU and DFE, we implement the messenger field algorithm (Algorithm 1) to draw samples of signal from the Wiener posterior (equation 3.7), using 5 different datasets at each iteration. This reflects Alsing et al. (2017), where multiple chains were run in parallel to test convergence. In figure 3.1, the value of the same pixel in 5 independent chains with different initial values can be shown to converge after a sufficient number of iterations. The period during which the chains have not converged is known as burn-in, and using these samples reduces the influence of the initial starting point. Subsequent points are not converged immediately, therefore it is essential to have multiple chains, to check convergence and improve statistics.

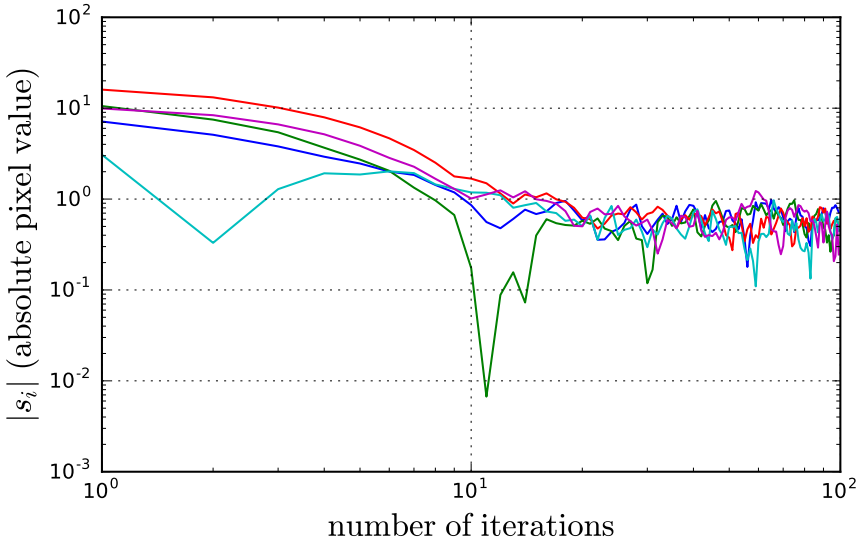

Figure 3.1: The absolute value of the same pixel at each iteration in 5 independent MCMC chains of the messenger field algorithm. The initial value of each pixel is different, to show convergence after a sufficient number of iterations.

On DFEs it is possible to instantiate fixed point and IEEE-like floating point arithmetic units of an arbitrarily chosen bitwidth, with greater flexibility beyond the standard options of single- or double-precision floating-point. Reducing the bitwidth of the number representation results in less reconfigurable logic spent on single arithmetic operations. This allows the user to instantiate more arithmetic units to fit the budget of reconfigurable space available on the chip, which may be used to implement more complex logic, or to replicate the computational pipeline; the latter reduces time to solution due to increased parallelism, but at the cost of reduced precision. In the implementation presented in this Chapter, we use single-precision floating point format on both the DFE and on the CPU, to compare more easily the results.

The CPU code, written in `C`, uses a Box-Muller transformation to generate pairs of normally-distributed random variates for use in the algorithm. This custom-written implementation was shown to be consistently faster than the `std` C++ Gaussian random number generator in unit tests. Our implementation is slightly faster as we only ever generate one pair of zero mean and unit variance Gaussian random numbers at each iteration. The DFE uses the Gaussian random number generator from a dedicated dataflow library[2]. This small difference changes the overall time measurement little, as the fraction of time spent gener-

[2] MaxPower (`maxeler.com/mymaxeler` requires Maxeler account)

ating random numbers is small in this algorithm.

The 2D FFT from the `FFTW3` package (Frigo and Johnson, 2005) was used for the CPU code, optimised with Advanced Vector Extensions (AVX2) available on the CPU hardware (see section 3.4.2). A dedicated dataflow FFT library was used for the DFE.

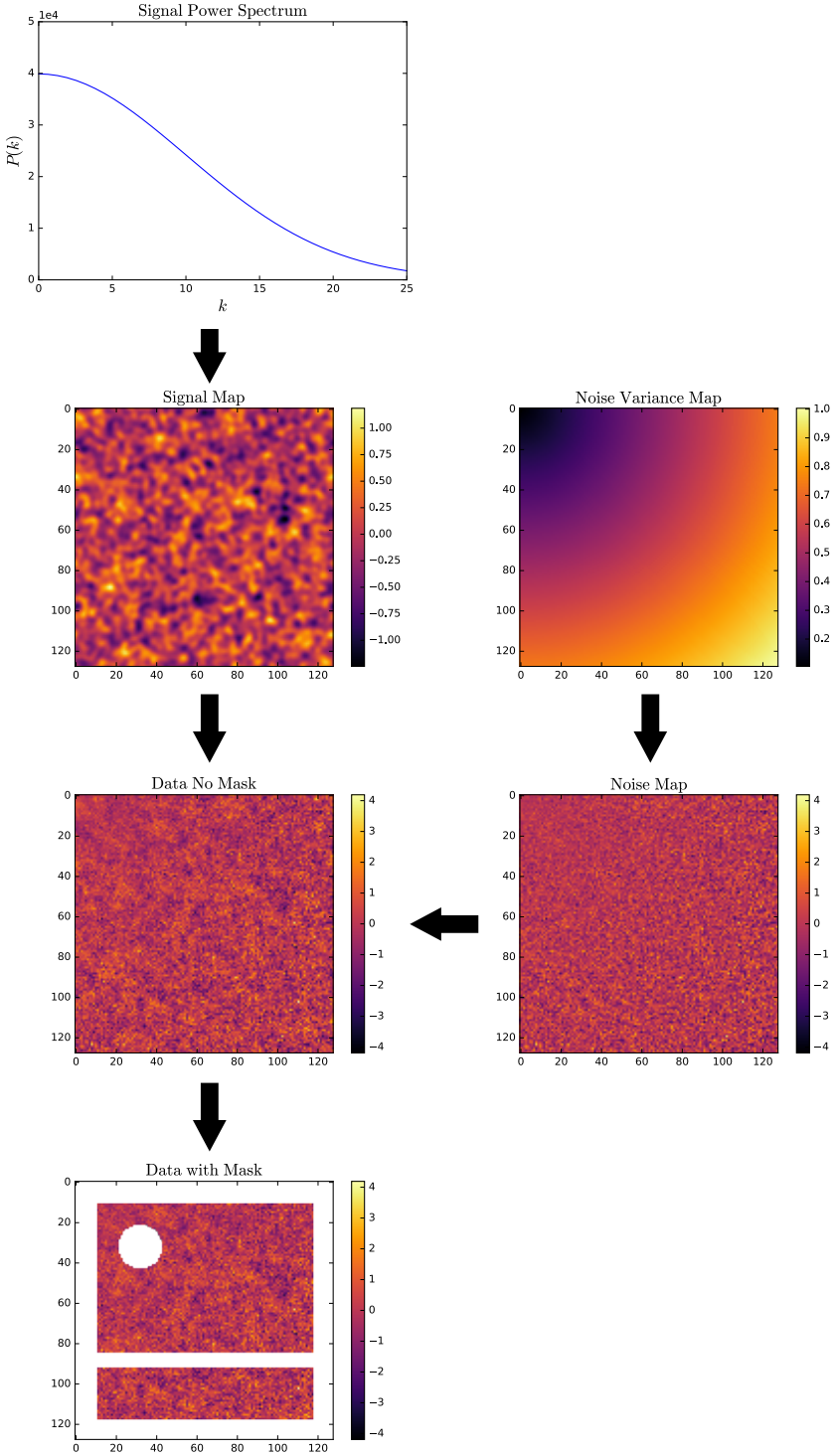

Figure 3.2: This figure shows our data model, and gives an example realisation of a simulated dataset. We begin with a ***Signal Power Spectrum***, $P(k)$, from which we generate a real, Gaussian random field as a ***Signal Map***. We then take a ***Noise Variance Map***, whose values vary across the data, from which we generate a ***Noise Map*** of Gaussian, independent pixel noise. The noise is added to the signal to generate the ***Data No Mask***. We mask pixels representing missing data in ***Data with Mask***.

3.4 Results

3.4.1 Wiener posterior properties

As described in section 3.2.2, samples from the Wiener posterior sampling algorithm without the random variates converge to the Wiener filter solution (equation 3.2). In the left panel of figure 3.3 the Wiener filter reconstruction from this method is shown for the data generated in figure 3.2. By doing this we tested that the CPU and DFE outputs are identical up to computational precision.

A second test also provided the DFE with a vector of random Gaussian variates, shared with the CPU, where the output samples from the Wiener posterior were shown to be the same within computational precision.

In the centre panel of figure 3.3, the mean of the 10^5 samples from the Wiener posterior can be seen. By comparing to the Wiener filtered image in the left panel, one can see that the Wiener filtered solution is indistinguishable from the mean of the samples from the Wiener posterior, as expected. Due to sample variance, the mean of samples from Wiener posterior is not exactly equal to the Wiener filter, though for an infinitely large number of samples it would be.

In figure 3.3, the variance of the same 10^5 samples can be seen in the right panel. The variance in the region of missing data is high, as expected, but constrained by the signal covariance. The structure of the variance of the samples matches the structure of the noise variance map (see figure 3.2) as expected.

By drawing sufficient samples from the full posterior probability distribution, the code can characterise it very well, not just providing its mean and covariance.

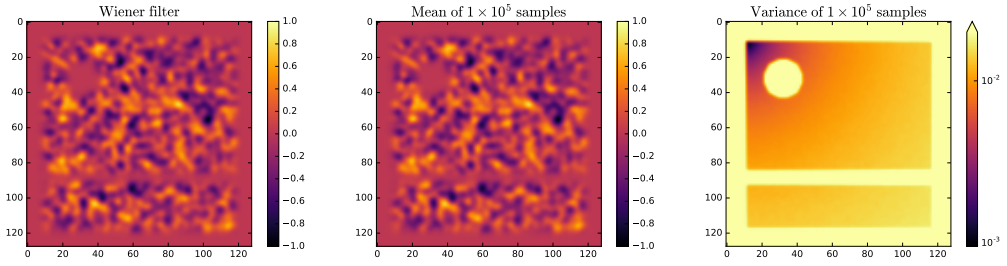

Figure 3.3: The input data, signal power spectrum and noise variance are shown in figure 3.2. *Left panel*: the left panel is the Wiener filter solution, where the random variates are not included in Algorithm 1. *Centre panel*: Mean of 10^5 samples from the Wiener posterior distribution evaluated using Algorithm 1. *Right panel*: Variance of the same set of samples.

3.4.2 CPU vs. DFE, Speed

We compare the speed of the CPU+DFE implementation of the Wiener sampler (Algorithm 1) to the pure CPU implementation. Both were run on an `Intel(R) Xeon(R) E5 – 2650 v2 @ 2.60GHz` server (2 sockets, 8 dual-thread cores per socket) presenting 32 CPU threads, which is connected to a `MPC-X` node at the STFC Hartree Centre. A single `MPC-X` box contains 8 `MAX4` (`Maia`) DFEs. The clock frequency for the DFE implementation was chosen to be 200 MHz.

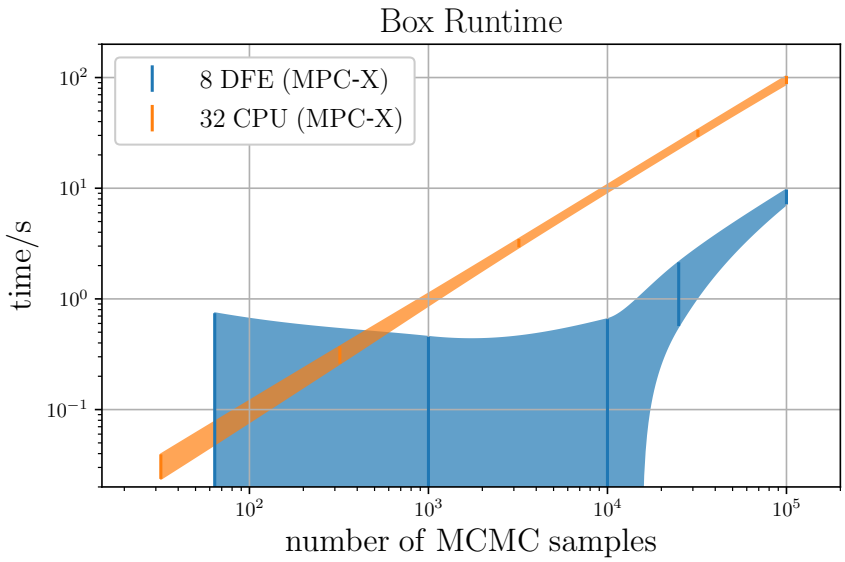

Figure 3.4: Time taken to run the sampling algorithm for a given number of iterations, where each iteration returns a sample from each of the 5 chains. Each data point is the mean of 10 runs (with the DFE data overhead removed), and 1σ error bars with Akima (1970) interpolation for the error envelope. 32 CPU threads vs 8 DFEs, run in parallel on `MPC-X`.

The sampler was timed for increasing number of iterations on both the CPU and the DFE, up to 10^5 iterations. Samples of 5 images are returned at each iteration. The time was measured from the CPU from the start to the end of the algorithm's execution. At each number of iterations, the algorithm was repeated 10 times and the measured times averaged.

Figure 3.4 shows the time to perform the algorithm for a 1U `MPC-X` with 8 DFEs against a 1U server presenting 32 CPU threads. The time as a function of number of iterations is linear for both CPU and DFE. The DFE has an initial overhead (with an average of 4.0 seconds) as the data is loaded onto the hardware, which is removed from the DFE time. Errors are obtained from 10 runs of the code. For the low run times, the DFE times have larger error-bars than the CPU, due to larger variance in the DFE data loading time; the relative effect of this decreases with longer running times.

In order to parallelise the problem, we run independent MCMC chains. We measure the time to generate a given number of samples by running the MCMC on 32 memory independent CPU threads for the CPU-only code. For the CPU+DFE code, we measure the time to generate a given number of samples on 8 DFEs by splitting the work across 8 CPU threads.

In this work we have one CPU thread orchestrating one DFE. In future work we need to support an N:1 ratio of N CPU threads served by a single DFE. This will help to utilise all the CPU computational capacity as well as all DFEs. Therefore, the speed-up of the CPU+DFE implementation in this work is a lower bound — here there is scope for considerable further acceleration.

From each independent MCMC chain some number of initial samples are unusable due to burn-in and are discarded. As each MCMC chain (run in parallel CPU threads) must discard the same number of initial samples, running 32 chains gives 32 times more unusable samples than a single chains with the same number of iterations. The 32 parallel CPUs will therefore have to discard four times more samples than 8 parallel DFE-accelerated CPU threads due to burn-in. This is also a reason why the time measurement from this `MPC-X` box parallel test should be interpreted as a lower bound on the potential speed up from DFEs.

We measure the lower bound on the parallel DFE speed-up to be 11.3 ± 0.8, where we have again used 10 time measurements to estimate the error.

3.5 Potential Applications to Cosmology

In this section we discuss some potential use cases in cosmology, although the algorithm and implementation are general and could be used in a number of contexts.

3.5.1 Power Spectrum Inference

A common problem is to extract information from the power spectrum, $P(k)$, of an underlying field, $\mathbf{s}$, as defined in equation 3.9, and an extension of the DFE code can allow this. For a zero-mean Gaussian random field, the power spectrum contains all the statistical information that defines the field. The specific aim for power spectrum inference is to calculate the posterior probability distribution of the power spectrum given a set of data.

The standard model of cosmology predicts that the density field of the early universe will be a Gaussian random field, which persists for large cosmological scales in the late universe. Estimating the power spectrum is therefore a standard tool in many cosmological

analyses with different datasets, including the early universe through Cosmic Microwave Background (CMB) radiation data (Planck Collaboration et al., 2016). A posterior probability distribution of the power spectrum of the density field in turn leads to posterior probability distributions for the cosmological parameters. These usually include, but are not limited to: the matter density Ω_m, the Dark Energy density Ω_Λ, the Dark Energy equation of state parameter w, and the Hubble parameter H_0.

The Bayesian hierarchical inference models described by Jasche and Lavaux (2015), Alsing et al. (2016b), and Alsing et al. (2017) to infer the posterior distributions of either the power spectrum or cosmological parameters in addition to samples from the field; see Fig.3.5. With Gibbs sampling, samples of both the power spectrum and the image are drawn, keeping the other temporarily fixed. For a given power spectrum, large numbers of samples from the posterior probability distribution of the underlying signal **s** can be drawn efficiently using Dataflow Engines, leading to better constraints on the power spectrum and hence on cosmological parameters.

3.5.2 Cosmological Mass Mapping

As discussed in Chapter 2, due to the local curvature of spacetime by the matter, images of distant galaxies are deformed by the inhomogeneous matter distribution along the line of sight. This is called gravitational lensing. Any matter can contribute to the lensing effect, making it a direct probe of non-visible dark matter (Kaiser and Squires, 1993). Reconstructing this density field facilitates the study of the dark matter physics, its relationship with visible matter, and can provide novel approaches to extract additional cosmological information.

The model describing this process when in the linear regime, known as weak gravitational lensing, is fully described by our linear model in equation 3.1. The data **d** are images where the pixel values are the mean of galaxy shapes within that pixel. The signal **s** is a weighted, projected density field[3] in the foreground of the observed galaxies. The pixel noise, due to the intrinsic random galaxy shapes, is approximately Gaussian. The density field in the late universe on large cosmological scales is also approximately Gaussian.

From data with these properties, the large-scale density field from weak lensing shape measurements can be principally recovered with a Wiener filter. As shown in Chapter 2 the messenger field Wiener filter algorithm is applied to Dark Energy Survey gravitational

[3]The weighted, projected density field in mass mapping is called convergence and is denoted by κ.

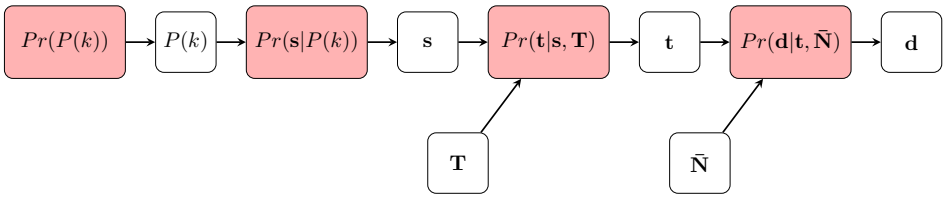

Figure 3.5: The Bayesian hierarchical forward model as described by Alsing et al. (2017) for signal image, $\mathbf{s}$, and power spectrum, $P(k)$, inference using the messenger field, $\mathbf{t}$. The work described in this Chapter uses Dataflow Engines can focus on the nodes of this network that do not include the power spectrum: the power spectrum is assumed and kept constant, and samples of the signal image are drawn.

lensing data to generate a mass map image of the underlying density field. The Wiener filter method has also long been an established tool for reconstructing the underlying density field using only galaxy positions (Lahav et al., 1994), rather than using lensing data. Obtaining a large number of samples of the Wiener distribution, as is described in this work, then gives a posterior probability distribution of the density field in each pixel.

3.5.3 Future Data Requirements for Cosmology

With current cosmic shear data[4], Alsing et al. (2017) were able to use CPUs to generate samples from the posterior probability distributions of the underlying cosmic shear signal images and the power spectrum, using the Bayesian hierarchical model shown in Fig. 3.5. 10 chains were run in parallel to a length of 10^5 samples.

Current and future cosmic shear surveys DES (DES Collaboration et al., 2017a), LSST (LSST Dark Energy Science Collaboration, 2012), and Euclid (Amendola et al., 2018)) expect orders of magnitude of increase in data volume. The European Space Agency project Euclid expects to observe over 10^9 galaxies usable for cosmic shear, compared to $\sim 3 \times 10^6$ with the CFHTLenS data used by Alsing et al. (2017). This leap in data size requires novel computational approaches to previously tractable problems. Here, Dataflow Engines can provide a solution.

3.6 Discussion

We have demonstrated a speed-up of at least 11.3 ± 0.8 for generating 10^5 samples of the Wiener posterior of possible images compatible with an observed noisy image of 128^2 pixels, using 8 DFEs in a 1U `MPC-X` box and comparing with a 1U server presenting 32 CPU

[4]Cosmic shear is the spin-2 complex field manifested as the coherent distortion of galaxy shapes due to gravitational lensing. It is a function of a linear projection into 2D of the 3D density field.

threads.

Future extensions could be to include the full Bayesian hierarchical model shown in figure 3.5, to further exploit the increased speed afforded to us by the Dataflow approach. This would lead to better constraints on the inferred cosmological parameters through samples of the power spectrum.

For data requirements of future cosmological surveys it would be useful to Wiener filter and draw samples of the Wiener posterior from data which have more than of 128^2 pixels per image. The image size in this work is constrained by the size of an FFT problem that fits within the fast FMEM on-chip memory ($\sim$6MB). We expect that future versions of the dataflow FFT library will provide the option to use off-chip memory (48GB) as an FFT buffer. We could then expect to be able to Fourier transform images of size $2^{15} \times 2^{15}$. This would increase the scientific applicability of a single DFE dramatically.

Implementing large scale Bayesian methods for cosmological parameter estimation on Dataflow Engines is a promising solution to the problem of increasingly large datasets from future surveys. This implementation of a Wiener sampler has broad application for inference or de-noising from any images or dataset with similar properties to those described here.

Chapter 4

Inference with theoretical predictions from noisy simulations

This Chapter is based on Jeffrey & Abdalla 2019: *Parameter inference and model comparison using theoretical predictions from noisy simulations.* accepted Monthly Notices of the Royal Astronomical Society, doi.org/10.1093/mnras/stz2930

When inferring unknown parameters or comparing different models, data must be compared to underlying theory. Even if a model has no closed-form solution to derive summary statistics, it is often still possible to simulate mock data in order to generate theoretical predictions. For realistic simulations of noisy data, this is identical to drawing realisations of the data from a likelihood distribution. Though the estimated summary statistic from simulated data vectors may be unbiased, the estimator has variance which should be accounted for. We show how to correct the likelihood in the presence of an estimated summary statistic by marginalising over the true summary statistic in the framework of a Bayesian hierarchical model. For Gaussian likelihoods where the covariance must also be estimated from simulations, we present an alteration to the Sellentin-Heavens corrected likelihood. We show that excluding the proposed correction leads to an incorrect estimate of the Bayesian evidence with JLA data. The correction is highly relevant for cosmological inference that relies on simulated data for theory (e.g. weak lensing peak statistics and simulated power spectra) and can reduce the number of simulations required.

4.1 Introduction

It is increasingly common, especially in cosmological surveys, to attempt to make inferences from data $\mathbf{d}$ using theory summary statistics μ that can be obtained only from simulations.

One example, currently popular in cosmology, is weak lensing peak statistics (Dietrich and Hartlap 2010; Kacprzak et al. 2016; Peel et al. 2017; Shan et al. 2017; Martinet et al. 2017). Peak statistics broadly aim to use the number of density peaks in the cosmological matter distribution to constrain cosmological parameters and models. The number of peaks in the density field (or weak lensing signal) is the result of highly non-linear structure formation and large-volume dark matter simulation are often needed to generate the theoretical number. The number of peaks in a given simulation is stochastic due to cosmic variance and then further sources of noise are added to simulate realistic observed data.

The data $\mathbf{d}$ are not the raw maps or catalogues, but the observed/derived summary statistics. For example, $\mathbf{d}$ would be the observed number of peaks in a weak lensing mass map reconstruction (see Kacprzak et al. 2016, or Chapter 2).

It was recently noted by Sellentin and Heavens (2016) that the common practice of debiasing a covariance matrix estimated from simulations of mock data (Hartlap et al., 2007) is insufficient. The sampling distribution of this estimated covariance should be incorporated into the likelihood distribution and, therefore, into the posterior distributions of the inferred parameters. Failure to do so leads to biased and overly-optimistic inferences.

In this work we note that, as with the estimated covariance described by Sellentin and Heavens (2016), an unbiased estimated summary statistic $\hat{\mu}$ is nevertheless itself a random variable, drawn from a sampling distribution with associated variance. If unaccounted for, this will lead to inaccurate parameter inference and misleading model comparison results.

In a Bayesian hierarchical framework, we present how to derive the posterior distribution of parameters by using a corrected likelihood distribution which takes into account that the theoretical predictions are based on noisy simulations.

An alternative to our presented hierarchical approach is to use likelihood-free inference (LFI) methods (e.g. Alsing et al. 2018, Leclercq 2018). In general LFI methods assume that the likelihood is unknown, and simulations are used to estimate the resulting posterior distribution conditional on data. However, if simulations are expensive and we believe

we know the naive likelihood[1] then LFI methods would be unnecessary and require more simulations, due to the increased number of degrees of freedom in the model. Instead, we show how to directly construct a Bayesian hierarchical model containing the free parameters of the known naive likelihood.

In section 4.2 we show how to marginalise over possible summary statistics μ to derive the likelihood $P(\mathbf{d} \mid \hat{\mu})$, the probability of the data $\mathbf{d}$ conditional on the estimated summary statistic $\hat{\mu}$. In section 4.3, we consider the case in which the naive likelihood is Gaussian, and derive the corrected likelihood distribution in the presence of both known and unknown (estimated) covariance matrices. In section 4.4, we use a one-dimensional toy model to demonstrate the effect of estimated summary statistics; we show that the corrected likelihood distribution matches samples generated from the toy model. In section 4.5, we use the public JLA supernovae data to show the effect of estimating summary statistics from simulations, using draws from the known likelihood as mock simulations.

4.2 Likelihood-based inference with simulations

4.2.1 Posterior and likelihood

Using the example of weak lensing peak statistics, we would wish to evaluate the posterior distribution of parameters of interest θ (e.g. Ω_m, σ_8 etc.) conditional on our observed data $\mathbf{d}$ (the number of peaks in a weak lensing map), which is given by Bayes' theorem:

$$P(\theta|\mathbf{d}) = \frac{P(\mathbf{d}|\theta)\, P(\theta)}{P(\mathbf{d})} \,. \tag{4.1}$$

If we were able to directly model the summary statistic $\mu(\theta)$ (that is, if we could calculate the number of weak lensing peaks for given parameters θ), then we could substitute $\mu(\theta)$ into the likelihood:

$$P(\mathbf{d}|\theta) = P(\, \mathbf{d} \mid \mu(\theta)\,) \,. \tag{4.2}$$

Of course, this is not possible in many cases. We must rely on estimates $\hat{\mu}_i$ of the true, but unattainable, summary statistic, with simulations run at position i in parameter space with parameters θ_i.

The correct representation of Bayes' theorem for the posterior distribution is

[1] Assuming we could condition on the true summary statistics which can be calculated $P(\mathbf{d}|\mu(\theta))$.

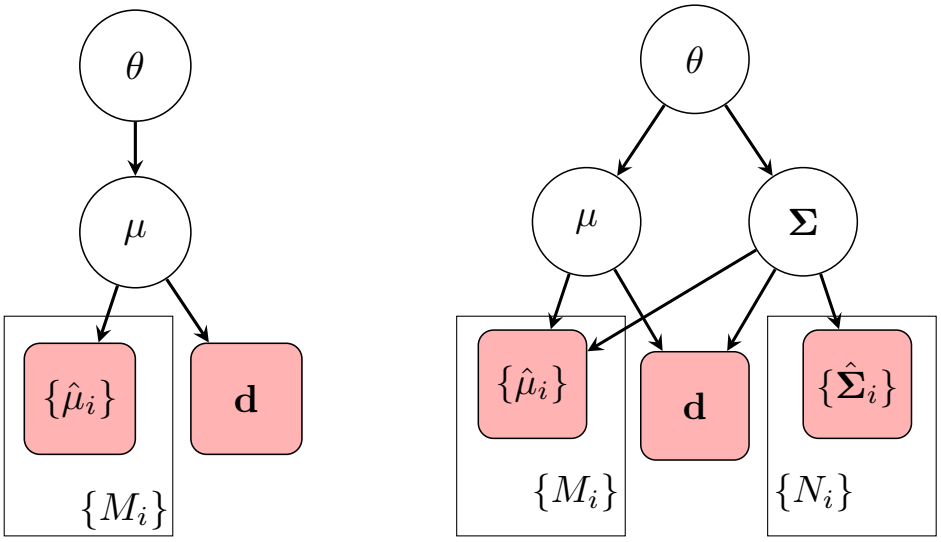

Figure 4.1: Probabilistic graphical representation in plate notation of Bayesian hierarchical models from sections 4.2.2 & 4.3.1 (*left panel*) and section 4.3.2. (*right panel*). Shaded nodes are "observed', either from experimental data ($\mathbf{d}$) or from simulations ($\hat{\mu}$ and $\hat{\Sigma}$). Each plate (rectangular box) includes the amount of data associated with the variable, for example each $\hat{\mu}_i$ (run at position i in parameter space) comes from M_i simulations.

$$P(\theta|\mathbf{d},\{\hat{\mu}_i\}) = \frac{P(\mathbf{d}|\{\hat{\mu}_i\},\theta)\ P(\theta)}{P(\mathbf{d})} \quad , \tag{4.3}$$

or equivalently

$$P(\theta|\mathbf{d},\{\hat{\mu}_i\}) = \frac{P(\mathbf{d},\{\hat{\mu}_i\}|\theta)\ P(\theta)}{P(\{\hat{\mu}_i\})\ P(\mathbf{d})} \quad , \tag{4.4}$$

keeping in mind that, as observed quantities, the observed summary statistic $\mathbf{d}$ and the statistics from the simulations $\{\hat{\mu}_i\}$ are independent and, therefore, separable. The brackets $\{\}$ represent the set of simulations run over the positions i in parameter space.

The factors $P(\{\hat{\mu}_i\})$ and $P(\mathbf{d})$ are both Bayesian evidence terms for the observed data and observed simulations. As constants, they can be largely ignored.

4.2.2 Likelihood correction

At position i in parameter space a set of M simulated data are generated. As it is not possible to calculate[2] the summary statistic μ, we estimate it from M simulated data realisations $\mathbf{d}_{\rm sim}$. The estimate is often the mean

[2]that is, $\mu(\theta)$ cannot be calculated for a given θ (e.g. peak statistics)

$$\hat{\mu}(\theta_i) = \frac{1}{M}\sum_{j=0}^{M-1} \mathbf{d}_{sim,j}(\theta_i) \ , \tag{4.5}$$

where each $j_{\rm th}$ data realisation is independent. In some cases, the summary statistic may not be the expectation of the likelihood $P(\mathbf{d}|\mu)$, so the estimator $\hat{\mu}$ would be another function of the simulated data (not the mean).

Even if $\hat{\mu}$ is an unbiased estimate (that is, $\langle\hat{\mu}\rangle = \mu$), it is often mistakenly assumed that

$$P(\ \mathbf{d}\ |\ \hat{\mu}_i\ ) = P(\ \mathbf{d}\ |\ \mu(\theta_i)\ ) \ , \tag{4.6}$$

which is not generally correct.

The correct likelihood to be used for parameter inference is with the estimated $\{\hat{\mu}_i\}$ from simulations as used in equation 4.3 or equation 4.4. This can be rewritten as a marginalisation over the unknown true summary statistic

$$\begin{aligned} P(\mathbf{d}\ |\ \{\hat{\mu}_i\},\ \theta) &= \int P(\mathbf{d}\ ,\ \mu\ |\ \{\hat{\mu}_i\},\ \theta)\ \mathrm{d}\mu \\ &= \int P(\mathbf{d}\ |\mu,\ \theta)\ P(\mu\ |\ \{\hat{\mu}_i\},\ \theta)\ \mathrm{d}\mu \\ &= \frac{1}{P(\{\hat{\mu}_i\})}\int P(\mathbf{d}\ |\mu,\ \theta)\ P(\{\hat{\mu}_i\}|\ \mu,\ \theta)\ P(\mu|\theta)\ \mathrm{d}\mu\ . \end{aligned} \tag{4.7}$$

or alternatively

$$P(\mathbf{d},\{\hat{\mu}_i\}|\ \ \theta) = \int P(\mathbf{d}\ |\mu,\ \theta)\ P(\{\hat{\mu}_i\}|\ \mu,\ \theta)\ P(\mu|\theta)\ \mathrm{d}\mu \ , \tag{4.8}$$

which is the same up to a constant evidence factor $P(\{\hat{\mu}_i\}$, as would be expected from equations 4.3 and 4.4. Which of the previous two distributions one wishes to think of as the traditional *likelihood* is somewhat academic, as once they are included into Bayes' theorem (equation 4.3 or 4.4), the posterior is the same.

Adding the corrected likelihood into equation 4.3 gives the posterior distribution for the parameters of interest

$$P(\theta|\mathbf{d},\{\hat{\mu}_i\}) = \frac{P(\theta)}{P(\mathbf{d})\ P(\{\hat{\mu}_i\})}\ \int P(\mathbf{d}\ |\mu)\ P(\{\hat{\mu}_i\}|\ \mu)\ P(\mu|\theta)\ \mathrm{d}\mu\ . \tag{4.9}$$

Here and in what follows, we drop explicit dependence on θ in terms like $P(\{\hat{\mu}_i\}|\ \mu,\ \theta)$, and

distributions will only condition on the previous layer in the hierarchical model (figure 4.1), as is a common practice (e.g. Leistedt et al. 2018). One can nevertheless keep in mind that probability distributions with μ is also conditioned on θ.

Equation 4.9 is the most general form, but each factor in the final integral can be evaluated for certain forms of the naive likelihood ($P(\mathbf{d} \mid \mu)$) and chosen priors:

- The first factor is the naive likelihood that would be used if the summary statistic μ could be calculated.

- The second factor is a sampling distribution of $\hat{\mu}$. If our simulated datasets $\mathbf{d}_{sim,j}$ for a given set of parameters θ_i are independent and realistic, then each is an independent and identically distributed (i.i.d.) draw from the naive likelihood distribution. Assuming we know the naive likelihood, it is usually possible to evaluate the sampling distribution $P(\hat{\mu} \mid \mu)$.

- The final factor of equation 4.7 is the prior on the summary statistic μ conditional on position in parameter space. This is not assigning prior probabilities to the values of the parameters themselves but on to possible forms of the summary statistic.

  If we believe, for example, that μ should vary smoothly in parameters space, this could be enforced through a smoothness prior $P(\mu|\theta)$. This can be compared to emulation methods (e.g. Heitmann et al. 2009, Bird et al. 2019, Jennings et al. 2019), where μ is estimated from simulations using a smoothing prior, either explicitly (Gaussian processes) or implicitly (machine learning). However, the uncertainties are not generally included in the final posterior distribution in a principled hierarchical way as described here.

  Conversely, if we claim to know nothing (or very little) about μ *a priori*, then we might consider a uniform prior (section 4.3.1).

4.2.3 Bayesian Hierarchical Model

The model described in the previous section is hierarchical and has a network of parameters related by conditional probabilities. Specific probability distributions of interest, such as the posterior probability distribution of the parameters, are evaluated by appropriate use of Bayes' theorem and marginalisation.

The left panel of figure 4.1 shows the probabilistic graphical representation of the hierarchical model pertaining to equation 4.9. This graphical representation may make the

logical steps of the previous section clearer, especially allowing us to see where conditions can be dropped (e.g. $P(\mathbf{d}\,|\mu,\,\theta) = P(\mathbf{d}\,|\mu)$).

4.3 Gaussian Naive Likelihood

For many cosmological analyses, the data $\mathbf{d}$ are assumed to have Gaussian noise and are, therefore, drawn from a Gaussian likelihood. In this section we derive the corrected likelihood for the case with known covariance (section 4.3.1) and unknown (estimated) covariance (section 4.3.2).

A Gaussian likelihood is usually an approximation, as there are likely to be some sources of non-Gaussian noise. It may be a very good approximation however. By the central limit theorem it may be the correct distribution in some limit of large numbers. For example, in a survey to measure the matter power spectrum $\mathrm{P}(k)$, if the galaxies are a Poisson process, then for modes that average many galaxies (high k) the likelihood is approximately Gaussian. Similarly, if weak lensing peaks are Poissonian, the binned counts of peaks will be approximately Gaussian for large numbers of observed peaks.

For cases where the naive likelihood is non-Gaussian and one wishes to calculate the corrected likelihood (conditional on an estimated summary statistic $\hat{\mu}$), one should evaluate equation 4.7 analytically or numerically.

4.3.1 Known Covariance

Consider the case where the naive likelihood (the first factor in the integral of equation 4.9) is a Gaussian/normal distribution, such that

$$P(\mathbf{d} \mid \mu(\theta)\,) = \mathscr{N}(\mathbf{d},\,\mu,\,\Sigma)\;\;, \tag{4.10}$$

and the covariance Σ is assumed known.

In this case, the sampling distribution (the second factor in the integral of equation 4.9) is

$$P(\hat{\mu}_i \mid \mu_i,\,\Sigma) = \mathscr{N}(\hat{\mu}_i,\,\mu_i,\,\frac{1}{M}\Sigma)\;\;, \tag{4.11}$$

at position i in parameter space.

If we do not wish to enforce any prior knowledge about μ, it seems reasonable to use the Jeffreys' prior (Jeffreys 1946, Jeffreys 1998) as an objective prior distribution for μ,

which, for our Gaussian likelihood (equation 4.10), is uniform

$$P(\mu|\theta) = P(\mu) \propto 1 \quad . \tag{4.12}$$

This flat prior on μ, if unbounded, is formally improper. However, the resulting (posterior-like) distribution, $P(\mu_i \mid \hat{\mu}_i,\, \Sigma)$, is Gaussian and therefore a true probability distribution.

With these distributions (equation 4.10-4.12) for a Gaussian naive likelihood, we perform the marginalisation integration (equation 4.7),

$$\begin{aligned} P(\mathbf{d} \mid \hat{\mu}_i,\, \Sigma) &= \frac{1}{P(\{\hat{\mu}_i\})} \int \mathscr{N}(\mathbf{d},\, \mu,\, \Sigma)\, \mathscr{N}(\hat{\mu}_i,\, \mu,\, \frac{1}{M}\Sigma)\, \mathrm{d}\mu \\ &\propto \int \mathscr{N}(\mathbf{d},\, \hat{\mu}_i, \Big(\frac{M+1}{M}\Big)\Sigma)\, \mathscr{N}(\mu,\, \mathbf{x},\, \mathbf{X})\, \mathrm{d}\mu \ , \end{aligned} \tag{4.13}$$

where $\mathbf{x}$ is a certain function of $\{\mathbf{d},\, \hat{\mu},\, \Sigma,\, M\}$ (but not μ) and $\mathbf{X}$ is a certain function of $\{\Sigma,\, M\}$[3]. The first factor can be brought outside the integral. The integration over μ of the second factor, which is a normal distribution, evaluates to one, which removes the dependence on $\mathbf{x}$. Normalising the resulting distribution with respect to $\mathbf{d}$ gives the corrected likelihood:

$$P(\mathbf{d} \mid \hat{\mu}_i,\, \Sigma) = \mathscr{N}(\mathbf{d},\, \hat{\mu}_i,\, \Big(\frac{M+1}{M}\Big)\Sigma) \ . \tag{4.14}$$

For summary statistics ($\hat{\mu}_i = \hat{\mu}(\theta_i)$) estimated from simulations (run with θ_i parameters), where the likelihood distribution for data $\mathbf{d}$ conditional on the true (but unknown) μ_i is Gaussian with known covariance Σ, and with an objective Jeffreys' prior on μ, then equation 4.14 is the corrected form of the likelihood. It is this corrected likelihood that should be used for parameter inference.

In this case, the corrected likelihood has the same Gaussian form as the naive likelihood, but with a scaled covariance. At first glance, this scaling could be mistaken for Bessel's correction for an unbiased estimate of the sample variance; however, here we actually know the covariance Σ and the added scaling comes from uncertainty in our estimate $\hat{\mu}$.

[3]This can be shown by completing the square, and is listed in Bromiley (2003) and Petersen and Pedersen (2012)

4.3.2 Unknown Covariance

In the previous section, we assumed that the summary statistic μ was estimated from simulations, but the covariance Σ was known. This situation is unlikely and it is foreseeable that both the summary statistic and covariance would also be estimated from simulated data.

The estimate of the covariance from N independent data simulations is given by

$$\hat{\mathbf{S}} = \frac{1}{N-1}\sum_{i=0}^{N-1}(\mathbf{d}_{sim,i} - \bar{\mathbf{d}}_{sim})(\mathbf{d}_{sim,i} - \bar{\mathbf{d}}_{sim})^{\dagger}\ , \tag{4.15}$$

where $\dagger$ is the conjugate transpose, and $\bar{\mathbf{d}}_{sim} = \frac{1}{N}\sum_{i=0}^{N-1}\mathbf{d}_{sim,i}$ is the just the estimated summary statistic $\hat{\mu}$ as given by equation 4.5 (but with N simulations, not M).

For the case where the summary statistic μ is not estimated from simulations, Sellentin and Heavens (2016) calculate the corrected likelihood

$$P(\mathbf{d}\,|\mu,\,\hat{\mathbf{S}},\,N) \propto \int P(\mathbf{d}\mid\mu,\,\Sigma)P(\hat{\mathbf{S}}\mid\Sigma)P(\Sigma)\,\mathrm{d}\Sigma\ . \tag{4.16}$$

For a Gaussian naive likelihood $P(\mathbf{d}\mid\mu,\,\Sigma)$ the distribution of the estimated covariance $P(\hat{\mathbf{S}}\mid\Sigma)$ is Wishart. With these distributions and a Jeffreys' prior for Σ, the resulting Sellentin-Heavens corrected likelihood is given by

$$P(\mathbf{d}\mid\mu,\,\hat{\mathbf{S}}) = \frac{\Gamma\left(\frac{N}{2}\right)|\hat{\mathbf{S}}|^{-1/2}}{\Gamma\left(\frac{N-p}{2}\right)\left[\pi(N-1)\right]^{\frac{p}{2}}}\left[1+\frac{(\mathbf{d}-\mu)^{\dagger}\hat{\mathbf{S}}^{-1}(\mathbf{d}-\mu)}{N-1}\right]^{-\frac{N}{2}}\ . \tag{4.17}$$

where p is the number of elements in the data vector $\mathbf{d}$ (i.e. the dimensionality). This has the form of a multivariate t-distribution.

In the case considered in this work, represented by the right panel of figure 4.1 we are assuming that the summary statistic μ cannot be calculated, and that we must estimate $\hat{\mu}$ from simulations. The integral in equation 4.16 must then be replaced by

$$P(\mathbf{d}\,|\{\hat{\mu}_i\},\,\{\hat{\Sigma}_i\},\,\theta) = \frac{1}{P(\{\hat{\mu}_i\})P(\{\hat{\Sigma}_i\})}\times \iint P(\mathbf{d}|\mu,\Sigma)P(\{\hat{\mathbf{S}}_i\}|\Sigma)P(\{\hat{\mu}_i\}|\mu,\Sigma)P(\Sigma|\theta)P(\mu|\theta)\,\mathrm{d}\mu\,\mathrm{d}\Sigma\ . \tag{4.18}$$

This integral is the (corrected) likelihood in the posterior distribution of θ for the model shown in the right-hand panel of figure 4.1.

We note that a given sample mean $\hat{\mu}_i$ and sample covariance $\hat{\mathbf{S}}_i$ are independent, even if the simulated data used to estimate them are the same (Anderson, 2003), so that $P(\hat{\mu},\hat{\mathbf{S}}) = P(\hat{\mu})\,P(\hat{\mathbf{S}})$. If the simulated data used to evaluate $\hat{\mu}$ and $\hat{\mathbf{S}}$ are the same, one can set $N = M$ in what follows. If, as is often the case in cosmological analyses, the covariance is assumed not to vary with respect to the parameters of interest and is instead estimated once with more simulated data realisations, then N is fixed and the i indices on the estimated covariance are dropped.

Using the same distributions as described so far in section 4.3, marginalising over the unknown true summary statistic μ and covariance Σ (equation 4.18) and renormalising gives the new corrected likelihood (details in section 4.7):

$$P(\mathbf{d} \mid \hat{\mu}_i,\, \hat{\mathbf{S}}_i) = \frac{\Gamma\left(\frac{N}{2}\right)\sqrt{\frac{M}{M+1}}|\hat{\mathbf{S}}_i|^{-1/2}}{\Gamma\left(\frac{N-p}{2}\right)\left[\pi(N-1)\right]^{\frac{p}{2}}}\left[1+\frac{M}{(M+1)(N-1)}(\mathbf{d}-\mu_i)^{\dagger}\hat{\mathbf{S}}_i^{-1}(\mathbf{d}-\mu_i)\right]^{-\frac{N}{2}}. \tag{4.19}$$

This corrected likelihood gives the probability of observing the data $\mathbf{d}$ conditional on an estimated mean summary statistic $\hat{\mu}_i = \hat{\mu}(\theta_{\mathbf{i}})$ from M simulations and an estimated covariance matrix $\hat{\Sigma}_i = \hat{\Sigma}(\theta_{\mathbf{i}})$ from N simulations, where we have assumed that the naive likelihood is Gaussian and have used Jeffrey's priors on μ and Σ.

4.4 Toy Model Demonstration

As a verification and demonstration of the result given in equation 4.14, where $\hat{\mu}$ is estimated from simulated data and Σ is known, we construct a toy model. This toy model also relies on the assumed flat prior on μ and the fact that the sampling distribution is symmetric with respect to μ and $\hat{\mu}$.

Let us assume that in different realisations of an experiment, different experimenters randomly and independently generate M simulations, from which $\hat{\mu}$ is estimated according to equation 4.5. The underlying likelihood distribution for the data with known μ is Gaussian and therefore the simulated data are themselves i.i.d. draws from the Gaussian distribution (equation 4.10). Each experimenter then draws a realisation of the data $\mathbf{d}$ from

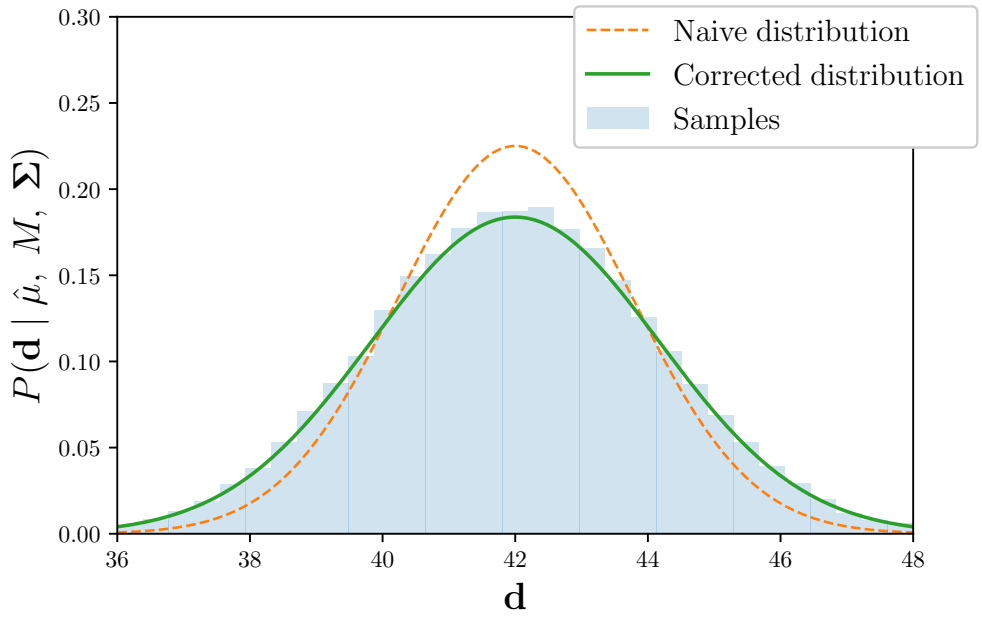

Figure 4.2: One-dimensional toy model for the naive Gaussian likelihood with known variance, where $\mu = 42$, $\Sigma = \pi \approx 3.14$, and $M = 2$. The samples (described in section 4.4) are distributed according to the corrected likelihood distribution (equation 4.14), whereas the naive likelihood distribution has reduced variance.

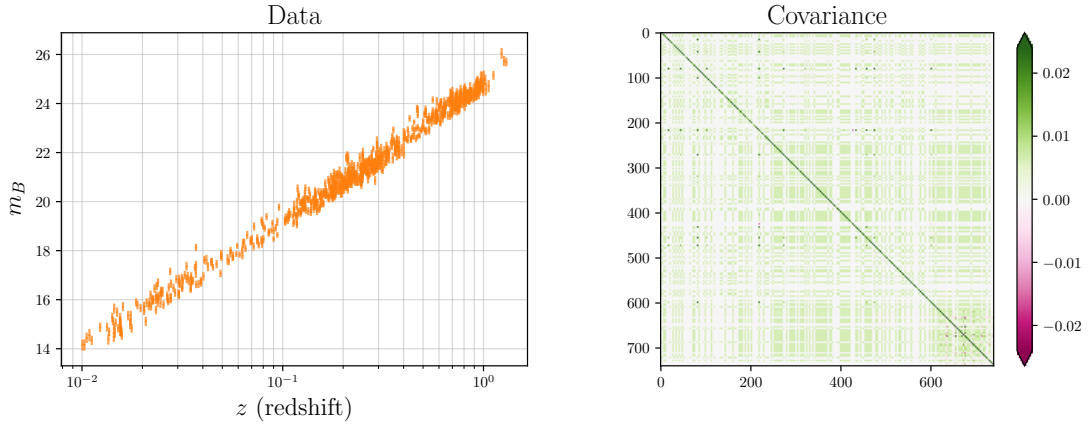

Figure 4.3: *Left panel*: The observed magnitude m_B data for 740 SN Ia. The error bars are taken as the square-root of the diagonal elements of the covariance. *Right panel*: The covariance matrix as described in section 4.5.1.

the naive Gaussian likelihood distribution with the known variance and the mean given by their estimated $\hat{\mu}$.

Though each experimenter draws their data realisation $\mathbf{d}$ from a Gaussian likelihood with the known variance Σ, the data realisations from all the experimenters will be distributed according to the corrected likelihood (equation 4.14) with variance $\frac{M+1}{M}\Sigma$.

In figure 4.2 we take a one-dimensional case where $\mu = 42$, $\Sigma = \pi \approx 3.14$, and $M = 2$. We see that the data samples from 10^5 different experiments match the corrected likelihood distribution (equation 4.14), whereas the naive likelihood distribution underestimates the variance, as expected.

4.5 JLA Supernovae Demonstration

In this section, we use public type Ia supernova (SN Ia) data[4] from the SDSS-II/SNLS3 Joint Light-Curve Analysis (JLA) (Betoule et al., 2014) as a demonstration of the corrected likelihoods described in the previous sections.

This is, of course, only a demonstration, as the summary statistic $\mu(\theta)$ (SN Ia apparent magnitudes) can actually be calculated for the model considered. We generate simulated data by drawing realisations from the known likelihood[5], and estimate $\hat{\mu}$. We can then constrain cosmological parameters using a likelihood distribution conditional on our estimated $\hat{\mu}$.

4.5.1 Data and Model

The data are observed B-band peak apparent magnitudes $\mathbf{d} = (m_{B,\text{obs},1}, m_{B,\text{obs},2}, ...)$ for 740 SN Ia over a range of redshifts up to $z = 1.3$. The supernovae also have associated light-curve stretch X_1, colour at maximum-brightness C and host stellar mass M_{stellar}, which are included in the model and covariance. The data and associated covariance are shown in figure 4.3.

We use the model from Betoule et al. (2014) where the SN Ia are standardizable candles with expected apparent magnitude

$$\begin{aligned} m_B = 5\log_{10}\left(\frac{D_L}{1\text{Mpc}}\right) + 25 \\ + M_B + \Delta M \Theta(M_{\text{stellar}} - 10^{10} M_\odot) \\ - \alpha X_1 + \beta C \,, \end{aligned} \tag{4.20}$$

where α and β are nuisance parameters for the stretch and colour respectively. M_B is the absolute magnitude of the host with a correction term ΔM depending on M_{stellar} (where Θ is the Heaviside function). We take a flat wCDM Universe, with luminosity distance

[4] `supernovae.in2p3.fr/sdss_snls_jla/ReadMe.html`

[5] Alsing et al. (2018) and Leclercq (2018) take a similar approach to demonstrate likelihood-free methods.

$$D_L = \frac{c(1+z)}{H_0} \int_0^z \frac{\mathrm{d}z'}{\sqrt{\Omega_m (1+z')^3 + (1-\Omega_m)(1+z')^{3(w+1)}}} \ , \tag{4.21}$$

where c is the speed-of-light in a vacuum, H_0 is the Hubble parameter, Ω_m is the matter density parameter, and w is the equation of state for dark energy.

In our demonstration, the parameters of interest $\theta = (\Omega_m, w, M_B)$ are allowed to vary. For simplicity, we fix the other parameters: $H_0 = 70.0$ km s^{-1} Mpc^{-1}, $\Delta M = -0.05$, $\alpha = 0.1256$, $\beta = 2.6342$. The covariance is calculated according to Betoule et al. (2014) and is also fixed.

4.5.2 Likelihood and Priors

We assume a Gaussian likelihood, where the log-likelihood is given by:

$$\ln P(\mathbf{d}|\mu, \Sigma) = -\frac{1}{2}\ln\left[\det\left(2\pi\Sigma\right)\right] - \frac{1}{2}(\mathbf{d}-\mu)^\dagger \Sigma^{-1} (\mathbf{d}-\mu) \ , \tag{4.22}$$

where the data and covariance are those described in section 4.5.1 (and shown in Fig. 4.3), and our summary statistic μ is given in equation 4.20.

For simplicity we take uniform priors in the ranges: $0.05 \leq \Omega_m \leq 0.6$, $-1.5 \leq w \leq -0.3$, $-19.15 \leq M_B \leq -18.95$.

Simulations for this demonstration are run on a regular grid of shape $[12, 13, 11]$ (for Ω_m, w, M_B) spanning the prior range. The regular grid is a particularly poor choice to sample the posterior distribution when simulations are expensive. However, this is a demonstration, and for real-world analysis many better sampling schemes are available (including latin hypercubes and grid transformations to better sample the expected posterior distribution).

Once the posterior is evaluated at these grid positions, the parameter space is upsampled to a $[48, 52, 44]$ grid. The new grid positions are evaluated by interpolating the posterior distribution from the original grid using a radial basis function 'thin plate' spline (Duchon 1976; Bookstein 1989; Jones et al. 2001). This spline interpolation worked particularly well in avoiding edge effects or artefacts around points when we compared their results with those of more poorly-performing simple polynomial splines.

4.5.3 Results

First, let us imagine three different experimenters, who, despite having access to the same data (described above), run their own independent simulations to estimate the summary statistics $\hat{\mu}(\theta)$ on the grid in parameter space. This results in different $\hat{\mu}_i$ for $i = 1, 2, 3$.

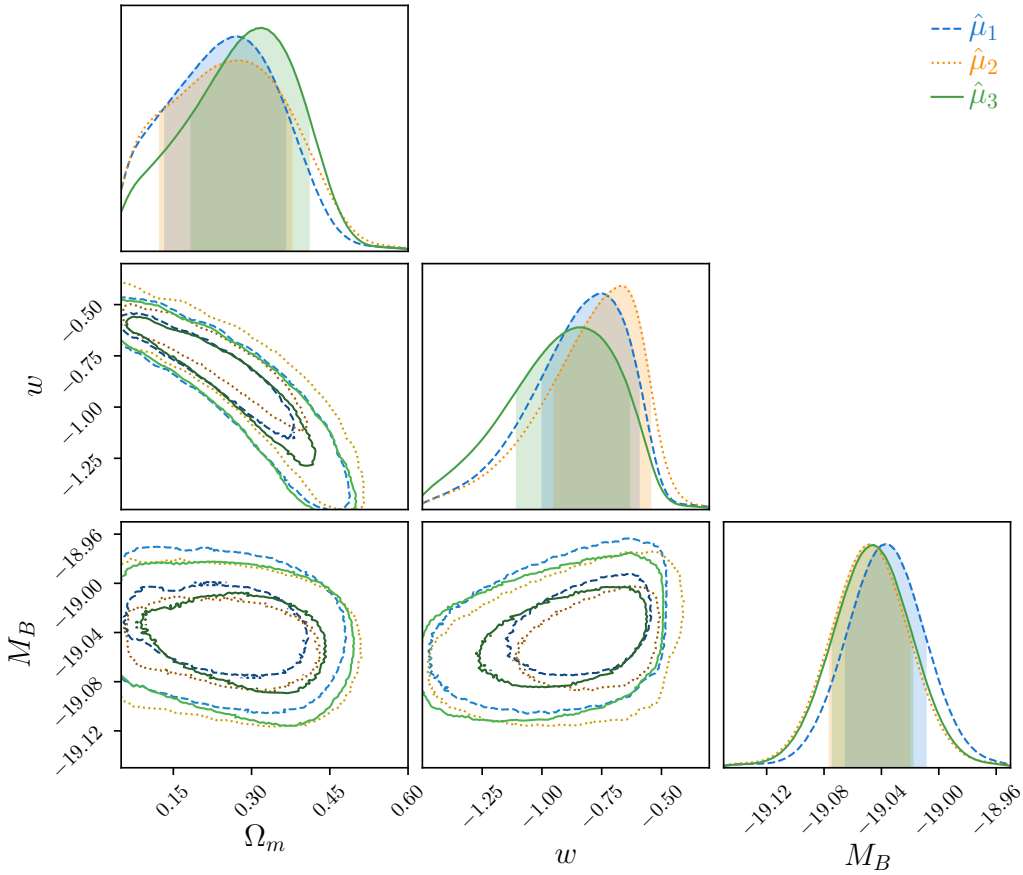

Figure 4.4: JLA posterior distribution for Ω_m, w, and M_B (described in section 4.5.2) using three independent estimates $\hat{\mu}$ with $M = 2$ simulations per position in parameter space. This uses the naive Gaussian distribution (equation 4.10) without the correction (equation 4.14), and the contours are therefore optimistically reduced.

Evaluating the posterior distribution using the naive Gaussian likelihood set-up described in section 4.5.2, and using $M = 2$ simulations per parameter grid position, results in the three posterior distributions in figure 4.4. The three different experimenters have three different posterior distributions due to their different $\hat{\mu}_i$.

Having different individual posterior distributions is in itself is not a problem. If different experiments have different data but the same underlying parameters, their resulting posterior distributions will look different, and will quantify their own individual uncertainty in the parameters. However, this variance of the data has been taken into account, and will be reflected in each posterior distribution. In the case of different $\hat{\mu}_i$ in figure 4.4, the fact that $\hat{\mu}$ was a random draw from a distribution (just like the data) has not been taken into account. As they have ignored the resulting correction to their likelihood, each experimenter will be overly optimistic about their own inferences.

In figure 4.5, the posterior distribution has been calculated using the likelihood cor-

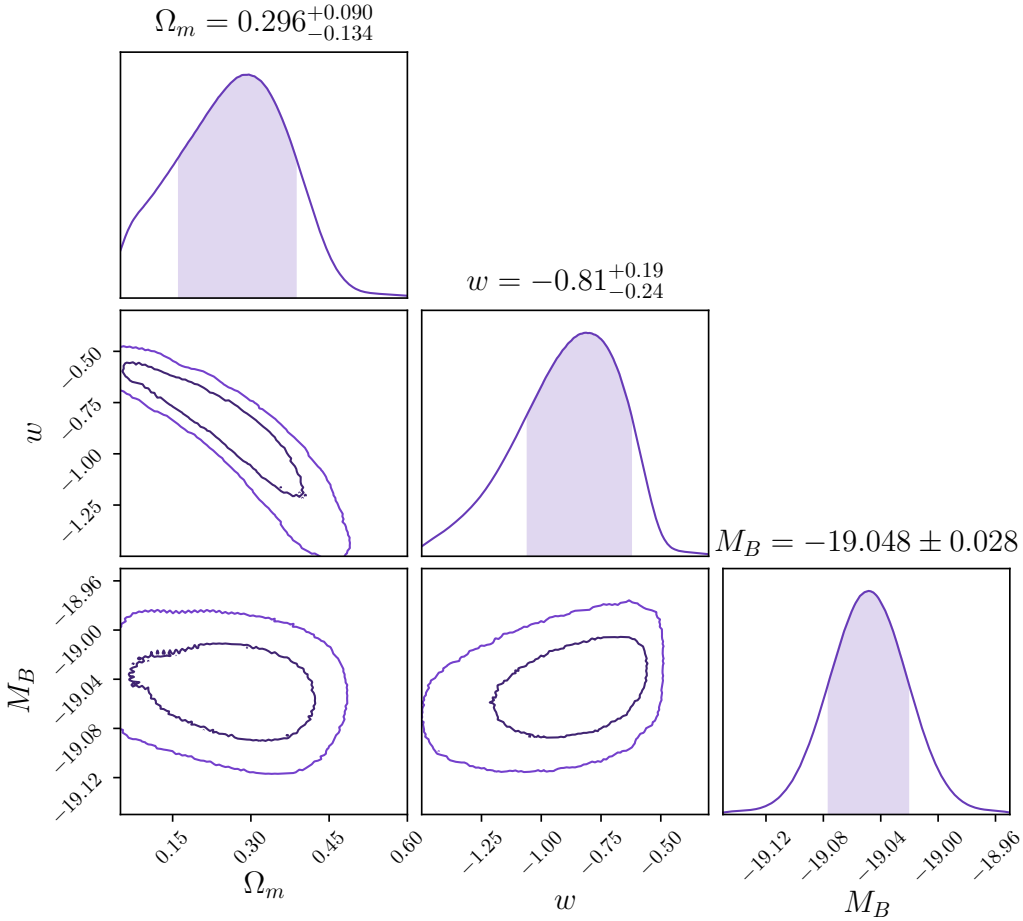

Figure 4.5: JLA posterior distribution for Ω_m, w, and M_B (described in section 4.5.3) using the corrected likelihood (equation 4.14) with $M = 2$ (we set $\hat{\mu} = \mu$ in this example for clarity). The contours are broader than in figure 4.4 (25 per cent increase in the parameter covariance determinant) as this likelihood takes into account that the estimated summary statistic is a draw from a sampling distribution.

rection (equation 4.14), which takes into account the variance in $\hat{\mu}_i$ (using the μ summary statistic for clarity). Using ChainConsumer (Hinton, 2016), we measure a 25 per cent increase in the determinant of the parameter covariance with the corrected likelihood. The resulting posterior distribution is slightly broader, reflecting the added uncertainty in the inference.

4.5.4 Model Comparison

The comparison of different theoretical models using the data in a Bayesian framework is usually done by calculating the Bayesian evidence:

$$P(\mathbf{d}|\text{Model}) = \mathscr{Z} = \int P(\mathbf{d}|\theta,\, \text{Model})\; P(\theta|\, \text{Model})\; \mathrm{d}\theta\;. \tag{4.23}$$

Two models can be compared by evaluating the Bayes factor:

$$K = \frac{\mathscr{Z}_1}{\mathscr{Z}_2} = \frac{P(\mathbf{d}|\text{Model}_1)}{P(\mathbf{d}|\text{Model}_2)} = \frac{P(\text{Model}_1|\mathbf{d})}{P(\text{Model}_2|\mathbf{d})} \frac{P(\text{Model}_2)}{P(\text{Model}_1)} \,. \tag{4.24}$$

If one has no reason to believe a given model more than another *a priori*, then the second factor (the ratio of the prior distributions) equals one. In this case, the Bayes factor becomes a ratio of the model probabilities (conditional on the data).

We evaluate the Bayesian evidence (equation 4.23) for the three parameter case for the JLA analysis described in section 4.5.2 with the uncorrected naive Gaussian likelihood and the corrected likelihood (equation 4.14) with $M = 2$. Calculating the Bayes factor gives

$$\begin{aligned} \ln K_{\hat{\mu}_1} &= 46 \\ \ln K_{\hat{\mu}_2} &= 31 \\ \ln K_{\hat{\mu}_3} &= 30 \,, \end{aligned} \tag{4.25}$$

for $\hat{\mu}_1$, $\hat{\mu}_2$, and $\hat{\mu}_3$ respectively, where

$$\ln K = \ln \mathscr{Z}_{\text{corrected}} - \ln \mathscr{Z}_{\text{uncorrected}} \tag{4.26}$$

using the corrected likelihood to evaluate the evidence $\mathscr{Z}_{\text{corrected}}$ and the naive, uncorrected likelihood for $\mathscr{Z}_{\text{uncorrected}}$.

As a check, after increasing the number of evaluated grid points by a factor of 4 we still calculate the same K values. Additionally, we calculate $\ln K$ using a different cosmological parametrisation, sampling scheme and data (section 4.8) and get similar results.

For all three, the corrected likelihood is more than a factor of $\exp[30]$ more probable than the uncorrected. This is further validation of the corrected likelihood; the model (i.e. the corrected likelihood) shows a better goodness-of-fit. Furthermore, if one were using an estimated summary statistic, but not using the corrected likelihood, one's belief in a model would be incorrect by this factor.

This effect would be particularly harmful if comparing two models, where it is possible to calculate μ for the first, but $\hat{\mu}$ is estimated from simulations for the second. Using the same Gaussian likelihood for both, without the correction for the second, would lead one to incorrectly favour the first model.

In figure 4.6, the log Bayes factors $\ln K$ (equation 4.25) are shown for the three esti-

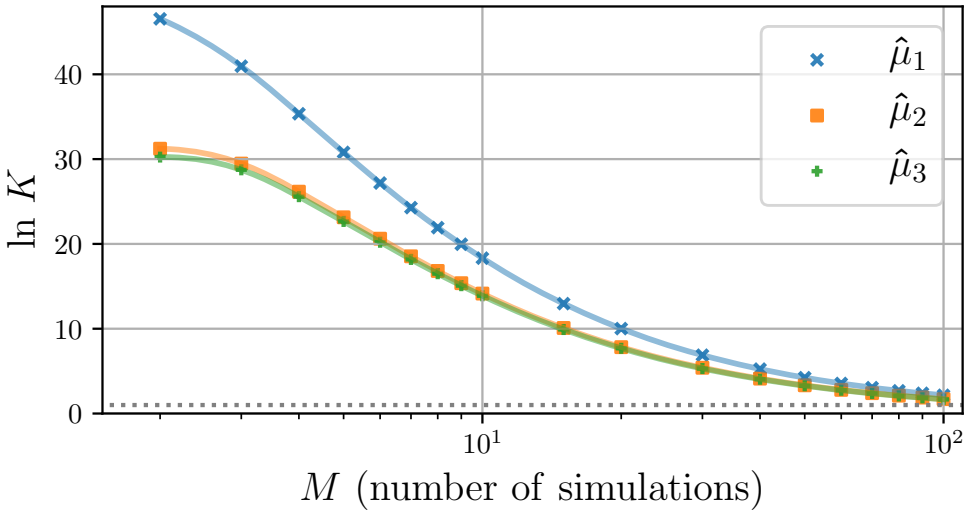

Figure 4.6: The log Bayes factor $\ln K$ as a function of the number of simulations M for the three estimated summary statistics $\hat{\mu}_i$ (using cubic spline interpolation between evaluated points).

mated summary statistics $\hat{\mu}_i$ as a function of the number of simulations M. As the number of simulations increase, the error incurred by using the uncorrected likelihood decreases. The value asymptotes to, but will never reach, $\ln K = 0$. For even a large number of simulations (large M), the error is not negligible. For many analyses, rather than running 10^2 expensive simulations, it would be better to use the corrected likelihood and avoid this error.

4.6 Discussion & Conclusions

In this work, we have shown how to take the sampling distribution of estimated summary statistics, $\hat{\mu}$, into account for parameter inference in a cosmological context. For situations where the naive likelihood is Gaussian, we have evaluated this correction (by marginalising over the unknown μ) for the case with known covariance (equation 4.14) and estimated covariance (equation 4.19).

We have validated the corrected likelihood with a toy model (section 4.4). Using JLA SN Ia data, we have demonstrated the effect of the corrected likelihood on cosmological parameter inference. For model selection, in our simple three-parameter inference demonstration, we show that the log Bayesian evidence $\ln K$ will be incorrect by a factor of over 30 if the uncorrected likelihood is not used.

In the era of DES (DES Collaboration et al., 2017b), LSST (LSST Science Collaboration et al., 2009) and Euclid (Amendola et al., 2016), cosmological analyses will have access to large cosmological datasets. Sole reliance on 2-point statistics in the linear regime

will be tantamount to wasting data which is rich in cosmological information. However, many summary statistics (μ) that access information beyond these 2-point statistics in the linear regime cannot be calculated analytically and need realistic simulations to be run to estimate $\hat{\mu}$.

A typical approach has been to run an excessive number of simulations at each position in parameter space, such that the variance of $\hat{\mu}$ in negligible. This approach has diminishing returns, as variance asymptotically tends to zero as $\frac{1}{M}$. This effectively aims to increase the number of simulations M until the sampling distribution $P(\hat{\mu} \mid \mu)$ can be effectively viewed as Dirac delta function (a limit it will, of course, never reach). Accepting a small increase in the resulting parameter constraints and correcting the likelihood for this sampling distribution means that fewer simulations have to be run.

If one does not wish to take the sampling distribution into account, one might use "cheap" simulations where it is possible to run enough that one effectively reaches the Dirac delta function limit. This has two potential pitfalls: firstly, the limit is never truly reached, which may affect the inferred parameters or model comparison results; secondly, cheap simulations are likely to be less realistic. It is far better to have slightly broader posterior distributions and to have used reliable simulations, than to have tighter constrains on parameters that are biased due to unreliable simulations.

The approach taken in this Chapter requires the acceptance that simulations are not "free". Simulations are increasingly an essential part of analyses. Like data, reliable simulations are often expensive in terms of time and resources and are, therefore, an acceptable contribution to the uncertainty of inferred parameters.

4.7 Supplementary material: full derivation

Here the full derivation leading to equation 4.19 is given, starting from the integration equation 4.18,

$$
\begin{aligned}
&P(\mathbf{d}|\{\hat{\mu}_i\},\ \{\hat{\mathbf{S}}_i\}) \propto \\
&\quad \iint P(\mathbf{d}|\mu,\Sigma) P(\{\hat{\mathbf{S}}_i\}|\Sigma) P(\{\hat{\mu}_i\}|\mu,\Sigma) P(\Sigma|\theta) P(\mu|\theta)\ \mathrm{d}\Sigma\ \mathrm{d}\mu \\
&\quad \int \left[\int P(\mathbf{d}|\mu,\Sigma) P(\{\hat{\mu}_i\}|\mu,\Sigma) P(\mu|\theta)\ \mathrm{d}\mu \right] P(\{\hat{\mathbf{S}}_i\}|\Sigma) P(\Sigma|\theta) \mathrm{d}\Sigma\ .
\end{aligned}
\tag{4.27}
$$

The first factor in the integral, $P(\mathbf{d}|\mu,\Sigma)$, is the Gaussian naive likelihood, given by equation 4.10. The final factor in the first integral is a uniform prior, $P(\mu) \propto 1$, as described in section 4.3.1. The second line uses the fact that $\hat{\mu}$ and $\hat{\mathbf{S}}$ are independent (Anderson, 2003). We can first perform the integration over μ, which, using the result from section 4.3.1, gives

$$P(\mathbf{d}\,|\hat{\mu}_i,\,\hat{\mathbf{S}}_i) \propto \int \mathscr{N}(\mathbf{d},\hat{\mu}_i,\Big(\frac{M+1}{M}\Big)\Sigma)P(\hat{\mathbf{S}}_i|\Sigma)P(\Sigma|\theta)\mathrm{d}\Sigma\,. \tag{4.28}$$

The second factor is the Wishart distribution

$$\begin{aligned} P(\hat{\mathbf{S}}_i|\Sigma,N) &= \mathscr{W}(\hat{\mathbf{S}}_i|\Sigma/(N-1),N-1) \\ &= \frac{|\hat{\mathbf{S}}_i|^{\frac{N-p-2}{2}}\exp\left[-\frac{N-1}{2}\mathrm{Tr}\left[\Sigma^{-1}\hat{\mathbf{S}}_i\right]\right]}{2^{\frac{p(N-1)}{2}}|\Sigma/(N-1)|^{\frac{N-1}{2}}\Gamma_p\left(\frac{N-1}{2}\right)}\,, \end{aligned} \tag{4.29}$$

and the third factor is the Jeffreys prior for Σ

$$P(\Sigma) \propto |\Sigma|^{-\frac{p+1}{2}}\,, \tag{4.30}$$

as described by Sellentin and Heavens (2016).

With these factors, we can rewrite equation 4.28, to give

$$\begin{aligned} &P(\mathbf{d}\,|\hat{\mu}_i,\,\hat{\mathbf{S}}_i) \\ &\quad\propto \int \mathrm{d}\Sigma\,|\Sigma|^{-\frac{N+p+1}{2}}\,\exp\Big[-\frac{1}{2}\mathrm{Tr}\big(\Sigma^{-1}\phi_i\big)\Big] \\ &\quad\propto |\phi_i|^{-N/2}\,, \end{aligned} \tag{4.31}$$

where

$$\phi_i = (N-1)\,\hat{\mathbf{S}}_i + \mathbf{v}_i\mathbf{v}_i^\dagger\,, \tag{4.32}$$

and

$$\mathbf{v}_i = (\mathbf{d}-\hat{\mu}_i)\sqrt{\frac{M}{M+1}}\,. \tag{4.33}$$

Using the identity $|\mathbf{A}+\mathbf{v}\mathbf{v}^\dagger| = |\mathbf{A}||1+\mathbf{v}^\dagger\mathbf{A}^{-1}\mathbf{v}|$ and normalising gives the new corrected likelihood in equation 4.19. The result is discussed in section 4.3.2.

4.8 Supplementary material: Pantheon

In section 4.5.4, the Bayes factors were calculated by evaluating the integral (equation 4.23) numerically on the parameter grid described in section 4.5.2. This resulted in values of $\ln K > 30$ for the corrected vs. uncorrected likelihood using the JLA data with the parameters described. Here we briefly describe how we verified the magnitude of this effect using 1048 Pantheon SN Ia (Scolnic et al., 2018) with a different set of parameters and a different method to evaluate the Bayesian evidence.

We allow four parameters to vary: q_0 (deceleration), j_0 (jerk), M_B, and $h = H_0/[100.0 \text{ km s}^{-1} \text{ Mpc}^{-1}]$. Deceleration and jerk are the parametrisation of a third order Taylor expansion of the scale factor $a(t)$ (Visser, 2004). The priors are uniform in the ranges: $-1 < q_0 < 1$, $-2 < j_0 < 2$, $-20 < M_B < -18$, and $0.4 < h < 1.2$. To evaluate the posterior distribution and Bayesian evidence we use PLINY (Rollins, 2015), which is a nested sampler (Skilling, 2004), and has been shown to be accurate when compared with known closed-form Bayesian evidence results.

The Bayes factor K is also differently defined here, as we set $\mu = \hat{\mu}$, inverting the result. The evaluated value $\ln K \approx \frac{1}{43}$, validates our results from section 4.5.4.

Chapter 5

Deep learning dark matter map reconstructions

This Chapter is based on Jeffrey et al. *Deep learning dark matter map reconstructions from DES SV weak lensing data.* submitted to MRNAS Letters, arXiv:1908.00543

We present the first reconstruction of dark matter maps from weak lensing observational data using deep learning. We train a convolution neural network (CNN) with a Unet based architecture on over 3.6×10^5 simulated data realisations with non-Gaussian shape noise and with cosmological parameters varying over a broad prior distribution. We interpret our newly created DES SV map as an approximation of the posterior mean $P(\kappa|\gamma)$ of the convergence given observed shear. Our DeepMass[1] method is substantially more accurate than existing mass-mapping methods. With a validation set of 8000 simulated DES SV data realisations, compared to Wiener filtering with a fixed power spectrum, the DeepMass method improved the mean-square-error (MSE) by 11 per cent. With N-body simulated MICE mock data, we show that Wiener filtering with the optimal known power spectrum still gives a worse MSE than our generalised method with no input cosmological parameters; we show that the improvement is driven by the non-linear structures in the convergence. With higher galaxy density in future weak lensing data unveiling more non-linear scales, it is likely that deep learning will be a leading approach for mass mapping with Euclid and LSST.

[1] github.com/NiallJeffrey/DeepMass

5.1 Introduction

The evolving cosmological density field is rich in information about the cosmological model of the Universe, its unknown parameters, and cosmic web-dependent astrophysics. Though the largest fraction of the density is invisible dark matter, the gravitational lensing effect of galaxies can be used to infer fluctuations in the total foreground matter distribution. Accurate "mass maps" will be essential for the science goals of the upcoming LSST survey and the ESA Euclid mission.

The maps considered in this Chapter are of the two-dimensional convergence, κ, a weighted projection of the matter density field in the foreground of the observed galaxies. Recovering the convergence from the measured galaxy shapes, known as observed shear $\gamma_{\rm obs}$ in the weak lensing regime, is an ill-posed inverse problem, troubled by survey masks (missing data) and galaxy "shape noise".

A typical principled approach to reconstructing more accurate mass maps in the presence of noisy, masked shear data is to use physically motivated priors. In Chapter 2, it was shown that using either Gaussian priors or "halo model" sparsity priors for κ improved the accuracy of the reconstructions with Dark Energy Survey Science Verification (DES SV) data. Implemented methods include using log-normal (Böhm et al., 2017) priors or E-mode priors (Mawdsley et al., 2019).

However, all of these priors take functional forms that only approximate the true object of interest, the prior on the convergence field $P(\kappa|\mathscr{M})$ (with model assumptions $\mathscr{M}$). These approximations are necessary because we cannot represent the probability distribution of the non-linear density field in closed form. For example, we cannot characterise it uniquely in terms of its moments (Carron and Szapudi, 2017). Even if the true, unapproximated prior were available, evaluation via direct calculation would likely be intractable.

Fortunately we can still draw realisations of convergence maps from the prior distribution $P(\kappa)$ in the form of simulations, which provides opportunity to a new generation of methods based on deep learning. Such an approach has been simultaneously proposed by Shirasaki et al. (2018), where a conditional adversarial network was used to learn a mapping from noisy convergence maps to an estimate of the noise-free convergence.

In this work, we propose a deep learning method to estimate the posterior mean of the convergence map from observed weak lensing shear measurements. In section 5.4 we demonstrate our method on simulations and DES SV data.

5.2 Weak gravitational lensing

5.2.1 Shear and convergence

Given a distribution of source galaxies n in radial comoving distance ω, the convergence at position $\vec{\phi}$ on the sky is given by a weighted integral of the density

$$\kappa(\vec{\phi}) = \frac{3H_0^2\Omega_m}{2}\int_0^\infty \Big[\int_0^\omega \mathrm{d}\omega' \frac{\omega'(\omega-\omega')}{\omega}\frac{\delta(\vec{\phi},\omega')}{a(\omega')}\Big] n(\omega)\mathrm{d}\omega \,, \tag{5.1}$$

where H_0 is the present value of the Hubble parameter, a is the cosmological scale factor, Ω_m is the matter density parameter, and δ is the overdensity.

We express the linear data model in matrix notation,

$$\gamma = \mathbf{A}\kappa + \mathbf{n} \,, \tag{5.2}$$

where $\mathbf{n}$ is a vector of noise per pixel. The matrix operator acting on the convergence $\mathbf{A}\kappa$ is the shear contribution due to lensing (Bartelmann and Schneider, 2001). In this formulation, the elements γ are the complex shear measurements binned into angular pixels in a two-dimensional image format.

We do not take into account the second order effects of reduced shear (Schneider and Seitz, 1995b), flexion (Bacon et al., 2006) or intrinsic alignments (Kirk et al., 2015). However, the deep learning approach taken in this Chapter is extremely flexible; as long as an effect can be modelled and included in the training data, it will be taken into account in the mass map reconstruction. This is not generally true of other methods. For example, flexion requires reformulations of methods (e.g. Lanusse et al. 2016). Additionally, noise per pixel is invariably approximated as Gaussian, which we do not assume in our deep learning approach.

5.2.2 Previous mapping approaches

The original mass mapping approach by Kaiser and Squires (1993) was a direct deconvolution. In practice Kaiser-Squires (KS) inverts the matrix $\mathbf{A}$ in Fourier space, where the matrix is diagonal. As this deconvolution is across a finite space, the edges of the data and internal masks introduce artefacts. KS is further troubled by the noise term in equation 5.2, which it does not take into account.

In a Bayesian framework we may wish to consider the posterior distribution of the convergence κ conditional on the observed shear γ

$$P(\kappa|\gamma,\mathscr{M}) = \frac{P(\gamma|\kappa,\mathscr{M})\,P(\kappa|\mathscr{M})}{P(\gamma|\mathscr{M})} \ , \tag{5.3}$$

The denominator $P(\gamma)$ is a Bayesian evidence term conditional on model $\mathscr{M}$. The first factor of the numerator is the likelihood $P(\gamma|\kappa,\mathscr{M})$, which encodes our noise model. The second term is the prior $P(\kappa|\mathscr{M})$, a possible selection of which was discussed in section 5.1.

If we believe a realisation of the convergence κ is a realisation of Gaussian random field, then the form of $P(\kappa)$ would be Gaussian. If the noise per pixel is Gaussian then the likelihood is also Gaussian, which results in a posterior distribution with both the mean and maximum given by the Wiener filter:

$$\hat{\kappa}_{\rm w} = \mathbf{W}\gamma = \mathbf{S}_\kappa \mathbf{A}^\dagger \left[\mathbf{A}\mathbf{S}_\kappa \mathbf{A}^\dagger + \mathbf{N}\right]^{-1} \gamma \, , \tag{5.4}$$

where $\mathbf{S}_\kappa = \langle \kappa\kappa^\dagger \rangle$ and $\mathbf{N} = \langle \mathbf{n}\mathbf{n}^\dagger \rangle$ are the signal and noise covariance matrices respectively (see Chapters 2 and 3 for implementation). The signal covariance in harmonic space is diagonal for isotropic fields. On the sphere, its elements are given by the κ power spectrum, $C_\kappa(\ell)$.

This Gaussian distribution is only approximately true for large scales where Gaussianity persists from the early Universe. On smaller scales, non-Gaussianity grows due to non-linear structure formation, which results in the cosmic web of the late Universe.

5.3 Deep learning maps

5.3.1 Convolution neural networks

We take a standard deep learning approach. We seek an approximation $\mathscr{F}_\Theta$ to the function that maps the pixelised shear to the convergence map

$$\hat{\kappa} = \mathscr{F}_\Theta(\gamma) \ , \tag{5.5}$$

where the parameters of the function Θ are to be learned (Goodfellow et al., 2016). We learn these parameters by minimising a mean-square-error (MSE) cost function

$$J(\Theta) = ||\mathscr{F}_\Theta(\gamma) - \kappa_{\rm true}||_2^2 \ , \tag{5.6}$$

evaluated on a set of training data which consists of pairs of realistic shear and "truth" (noise-free) convergence maps. If the training data "truth" maps are drawn from a prior dis-

tribution $P(\kappa)$, and the corresponding noisy shear map is drawn from the likelihood $P(\gamma|\kappa)$, this MSE cost function corresponds to $\mathscr{F}_\Theta(\gamma)$ being a mean[2] posterior estimate (Jaynes, 2003), such that $\hat{\kappa}$ is approximating:

$$\hat{\kappa} = \mathscr{F}_\Theta(\gamma) = \int \kappa \, P(\kappa|\gamma) \, \mathrm{d}\kappa \ . \tag{5.7}$$

We use a deep convolution neural network (CNN) to approximate the function $\mathscr{F}_\Theta$, where the parameters Θ are primarily elements of learned filters in convolutional layers. CNNs are particularly suited for two-dimensional image or one-dimensional time series data with translation invariant features in the underlying signal.

The CNN is a series of iteratively computed layers. At a given layer j the signal $\mathbf{x}_j$ is computed from the previous layer $\mathbf{x}_j = \rho \mathbf{M}_j \mathbf{x}_{j-1}$, with linear operator (e.g. convolution) $\mathbf{M}_j$ and nonlinear *activation* function ρ (LeCun et al. 1990, Mallat 2016). The output of a layer is sometimes called a *feature map*.

Due to their additional layers, deep architectures are often able to learn features with greater complexity than shallow architectures and therefore can better approximate the target function. For a general overview of deep learning and neural networks we recommend Goodfellow et al. (2016).

5.3.2 DeepMass architecture

Our DeepMass architecture is based on the Unet (Ronneberger et al., 2015), which has a so-called *expanding path* and *contracting path*. The DeepMass contracting path differs from the original Unet: usually convolutions and activation are followed by a *max pooling* operation to downsample the images, whereas we use *average pooling* (Géron, 2017). With each downsampling operation, the images decrease in resolution, but the 3×3 filters cover more angular size of the image. The convolution after a pooling operation therefore has a receptive field that covers larger physical features in the convergence κ map.

There are similarities between Unet architectures and sparse recovery methods. These consider representations where the solution is sparse and employ transforms which are fixed (e.g. Fourier, wavelets) or learned from data, and optimisation is solved using proximal theory (Starck et al., 2015). The Unet expanding and contracting path are very similar to synthesis and analysis concepts in sparse representations. This has motivated the use of wavelets to implement the Unet average pooling and the expanding path (Ye et al., 2018;

[2]The mean posterior is not generally the *maximum a posteriori*

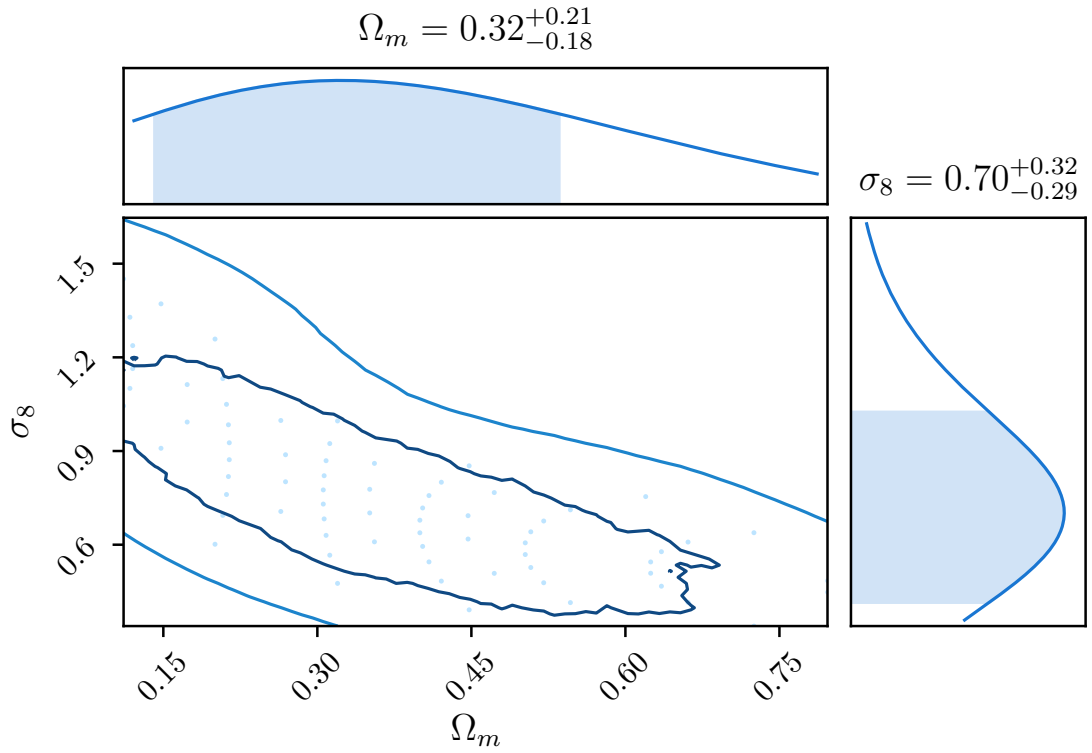

Figure 5.1: Prior range of cosmological parameters Ω_m and σ_8 of the training data. Simulations were run at the marked points.

Han and Ye, 2018). There are nevertheless significant differences: Unets can learn rich sets of features (corresponding to sparse dictionaries) from large training datasets, and the CNN implementation of non-linearity.

We differ from the original Unet by not using padding in the convolutional layers, as the edge of our data mask is already many pixels away from the edge of the square image. This choice means that output of a convolution has the same image dimensions as the input.

The full architecture and code can be seen online (DeepMass[†]), and a graphical model of the architecture can be seen in section 5.6. We have added *Batch Normalization* layers (Ioffe and Szegedy, 2015) after each convolutional layer; without this, training often became stuck in local minima of the cost function with respect to the parameters Θ. For all layers, except for the final, we use the rectified linear unit (ReLU) activation. In the final layer we use a sigmoid function, which forces the output to be between 0 and 1 (inputs and outputs are correspondingly rescaled).

For simplicity and memory efficiency, we aimed to work with real (32-bit) numbers, thus necessitating an initial operation acting on the complex shear γ. The best results came from using a fixed Wiener filter operation before the first convolution (rather than KS, as

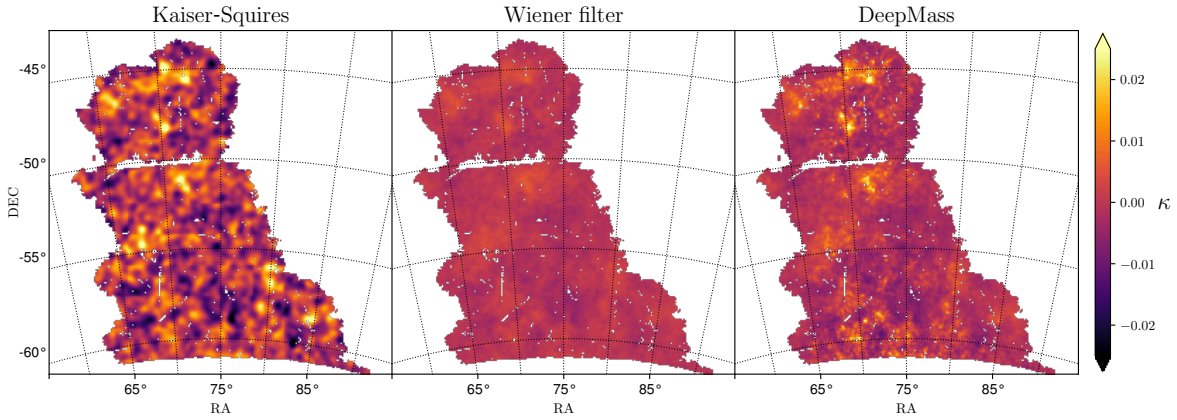

Figure 5.2: Convergence κ reconstruction from DES SV observational data with: KS, Wiener filtering, and DeepMass.

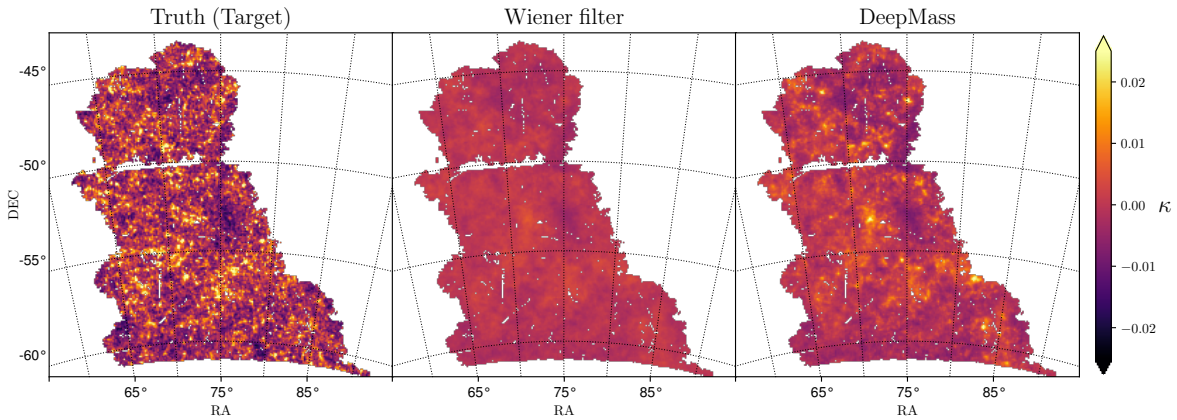

Figure 5.3: Example L-PICOLA validation simulation (*left*) and the corresponding Wiener (*centre*) and DeepMass (*right*) reconstructions.

might be expected). This is equivalent to the first layer having $\mathbf{M}_{j=0} = \mathbf{W}$ and $\rho = 1$, with no free parameters. We could also interpret the Unet after the initial Wiener operation as $\mathscr{G}_\Theta$ where $\mathscr{F}_\Theta(\gamma) = \mathscr{G}_\Theta(W(\gamma))$. The Wiener filter used a power spectrum with cosmological parameters σ_8 and Ω_m fixed at the mean of the marginal posterior distributions from DES Y1 analysis (Abbott et al., 2018). The flat sky power spectrum was an average of 102 power spectra of projected patches.

5.3.3 Training data

5.3.3.1 L-PICOLA simulations

The training data is derived from 74 independent dark matter simulations, with each simulation covering an octant of the sky. The simulations used a standard flat ΛCDM cosmological model with $H_0 = 70$ km Mpc^{-1}s^{-1}. The scalar spectral index and baryon density were fixed at $n_s = 0.95$ and $\Omega_b = 0.044$ respectively. The values of Ω_m and the amplitude parameter

σ_8 are distributed on a non-Euclidean grid with distances between points giving a density according to our prior $P(\sigma_8, \Omega_m)$ as shown in figure 5.1. Weak lensing constraints are most sensitive to combinations of this pair of parameters, so we avoid overfitting to a single cosmology by varying them in the training data.

To generate a convergence map from a simulation, the matter particles were binned using the HEALPix (Górski et al., 2005) pixelisation of the sphere with NSIDE=2048 in comoving radial shells of 50 Mpc/h. The density ρ map in a given redshift was converted into an overdensity $\delta = \rho/\bar{\rho} - 1$ using the average density in the shell $\bar{\rho}$. The convergence was calculated per pixel using equation 5.1. We wish to have the $n(z)$ in the lensing kernel match the DES SV data (section 5.4.1), which we approximate by summing the individual posterior redshift distributions per galaxy from the BPZ photometric redshift code (Coe et al., 2006). The convergence maps were downgraded to NSIDE=1024.

The dark matter simulations are generated using the L-PICOLA code (Howlett et al., 2015), which is based on the COLA (Tassev et al., 2013) algorithm. This uses a combination of second-order Lagrangian perturbation theory (2LPT) and a Particle-Mesh (PM) which requires fewer time steps than "full" N-body (e.g. Gadget Springel 2005) and therefore can generate simulations more quickly. This allows more training data to be generated in a given amount of compute time.

We used a 1250 Mpc$/h$ comoving simulation box, 768^3 particles, and a 1536^3 grid. A $z < 1.6$ lightcone was generated with up to four box replicates, using 30 time steps from $z = 20$. The initial conditions used Eisenstein and Hu 1999 for the linear matter power spectrum.

The drawback of this approach is the accuracy of the dark matter distribution. The finite spatial resolution and fewer timesteps used by the COLA method particularly affects small distance scales. Our experiments have shown a suppression of the L-PICOLA power spectrum at scales of $\ell > 700$ of order 10 per cent (relative to NICAEA[3] (Kilbinger et al., 2009) theory), as is expected with COLA methods. We correct the power of the L-PICOLA convergence by estimating the smooth part of the $C_\kappa(\ell)$ using a polynomial order 1 Savgol filter with window size 91 for each convergence map and reweighting spherical harmonics. Using the ratio of NICAEA and only the smooth part of the measured simulation power spectrum ensures that the natural fluctuations inherent in $C_\kappa(\ell)$ for a given realisation are

[3] nicaea.readthedocs.io

preserved.

5.3.3.2 Training images

From the 74 independent HEALPix convergence maps over an octant of the sky, we generate 376,684 DES SV mock data realisations. A given realisation is generated from the HEALPix convergence map by randomly choosing a position on the sphere, applying a uniform random rotation between 0 and 360 deg, and extracting a square patch using a gnomonic projection with 256^2 pixels of size 4.5^2 arcmin^2. If the generated image has pixels outside the octant, it is rejected. The rotation step is not to make the reconstruction rotation invariant, which happens naturally as $P(\kappa)$ is isotropic by the cosmological principle, but it is to augment the training data and learn $\mathscr{F}_\Theta$ better.

From the projected square κ convergence map, the complex noise-free shear map is generated using the **A** matrix from equation 5.2. The mask is applied and a random shape noise map is added. The noise map is generated by randomly shuffling the positions of galaxies in the original catalogue; this keeps the density of galaxies the same, but destroys the coherent lensing signal. This way we forward model the non-Gaussian noise inherent in the data (something that other methods do not do).

5.4 Results

5.4.1 Dark Energy Survey SV data

DES is a ground based photometric galaxy survey, observing in the southern sky from the 4m Blanco telescope in Chile with five photometric filters (Flaugher et al., 2015). The SV (A1) data[4] come from an initial run of 139 deg^2, but with depth of approximately that of the full 6 year survey (Chang et al., 2015). We make a redshift cut of $0.6 < z_{\rm mean} < 1.2$, where $z_{\rm mean}$ is the mean of the z posterior per galaxy. Data selection choices match Chapter 2, though some maps appear different due to changes in pixel size and flat-sky projection.

In figure 5.2 we apply KS, Wiener filtering, and the trained DeepMass CNN. Kaiser-Squires uses a 10 arcmin Gaussian smoothing as in Chapter 2. The Wiener filtering uses a power spectrum with Ω_m and σ_8 at the mean of their respective marginal posterior distributions from the Year 1 DES cosmology result (Abbott et al., 2018). The DeepMass CNN was trained using the Adam optimiser with a learning rate = 1×10^{-5} for 20 epochs (retraining

[4] http://des.ncsa.illinois.edu

over the full training set). The final Wiener and DeepMass maps were smoothed with a Gaussian of $\sigma = 2.25$ arc min (half pixel size) to remove very small scale artefacts arising from the HEALPix projection.

The DeepMass reconstruction clearly shows more non-linear structure than the Wiener filter. Individual peaks, which are suppressed by Wiener filtering, are resolved by DeepMass.

5.4.2 Validation on simulations

Out of the original generated training images (section 5.3.3), 8000 were kept and not used for training to be used as validation. One such example can be seen in figure 5.3, with the corresponding Wiener filter and DeepMass reconstructions. As with the reconstruction from observational data, DeepMass can be seen to recover the non-linear (cosmic-web) structure better than Wiener filtering. Compared to Wiener filtering, the MSE over all 8000 maps is improved using DeepMass by 11 per cent.

In the work of Chapter 2, using a "halo-model" sparsity prior did not outperform Wiener filtering in terms of MSE, so we expect DeepMass MSE to outperform GLIMPSE. However, MSE minimisation relates just to the posterior mean, so alternative metrics (e.g. constraints from peak statistics) remain to be explored.

Using 18 non-overlapping mock DES SV data from the MICE (Fosalba et al., 2015) simulations we apply a Wiener filter with an optimal power spectrum calculated using the known cosmological parameters (not available in real data applications). Nevertheless, without using the known cosmological parameters as input, DeepMass still recovers maps with an average of 2 per cent better MSE.

Furthermore, the MICE cosmological parameters lead to a relatively low power and fewer high variance structures above the signal-to-noise level. The largest improvement over Wiener filtering comes when there are more non-linear (non-Gaussian) structures. With the same MICE simulations, restricting ourselves to pixels where the truth is greater than two standard deviations from the mean $\kappa > 2\sigma$, compared to Wiener filtering, DeepMass improves the MSE by 8 per cent. As is to be expected, therefore, DeepMass improves over Wiener filtering due to its ability to reconstruct the non-linear structures in the cosmological signal.

5.5 Conclusion

With DeepMass, we have presented a deep learning method to reconstruct convergence κ maps from shear measurements. With DES SV, we have shown the mass map reconstruction with deep learning from observational data.

By training with simulations over a broad prior distribution of cosmological parameters, we have a generalised method which needs no input cosmological parameters. This method has shown substantial improvement over Wiener filtering both qualitatively (by eye) and quantitatively (11 per cent MSE reduction on the validation data). The flexible approach also takes into account non-Gaussian noise in the weak lensing data. As our simulated training data are samples drawn from the prior $P(\kappa)$, the approach has a principled Bayesian interpretation, without the need for evaluation of closed-form priors.

The quality of the reconstruction with these initial experiments, and its flexibility, makes the deep learning approach a preeminent candidate for mass mapping with future weak lensing surveys.

5.6 Supplementary material: DeepMass architecture

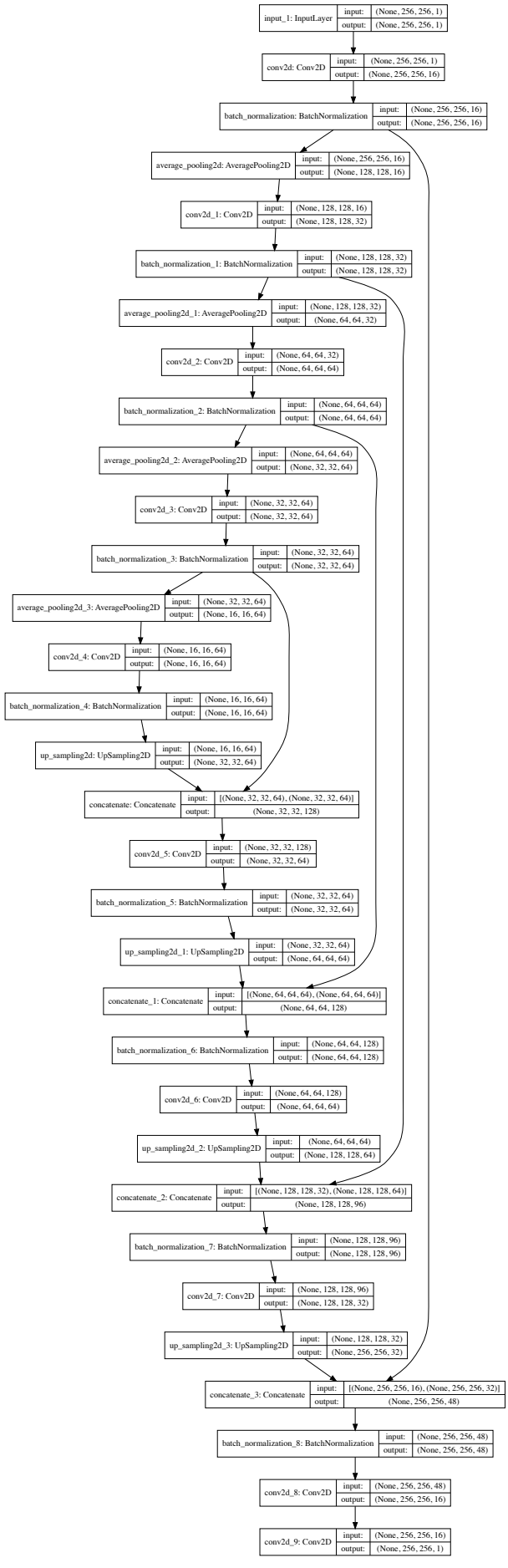

Figure 5.4: The final, full DeepMass architecture for 256× 256 images.

Chapter 6

Conclusions and future challenges

6.1 Thesis summary

In this section, I will summarise the work described in Chapters 2 to 5. In section 6.2 I will outline some future research that could follow from the work in the previous chapters and its relationship with other concurrent results from the field.

In Chapter 2, I compared three mass mapping methods with closed-form priors using DES SV data and simulations. The Wiener filter and GLIMPSE offered substantial improvements over smoothed KS with a range of metric.

Both the Wiener filter and GLIMPSE convergence reconstructions show a 12 per cent improvement in Pearson correlation with the underlying truth from simulations. To compare the mapping methods' abilities to find mass peaks, I measured the difference between peak counts from simulated ΛCDM shear catalogues and catalogues with no mass fluctuations (a standard data vector when inferring cosmology from peak statistics); the maximum signal-to-noise of these peak statistics is increased by a factor of 3.5 for the Wiener filter and 9 for GLIMPSE. With simulations I measured the reconstruction of the harmonic phases; the phase residuals' concentration is improved 17 per cent by GLIMPSE and 18 per cent by the Wiener filter. The correlation between reconstructions from data and foreground redMaPPer clusters is increased 18 per cent by the Wiener filter and 32 per cent by GLIMPSE.

In Chapter 3, I demonstrated how the Wiener filter (one of the above methods) computation can be sped up by an order of magnitude using Dataflow Engines (DFE), which are FPGA-based reconfigurable hardware. I demonstrated a lower bound on the speed-up, from drawing 10^5 samples of a 128^2 image, of 11.3 ± 0.8 with 8 DFEs in a 1U `MPC-X` box when compared with a 1U server presenting 32 CPU threads.

In Chapter 4, I presented a Bayesian hierarchical model which takes into account added uncertainty introduced when noisy simulations are used to generate theoretical predictions. The correction achieved with this model is highly relevant for cosmological inference that relies on simulated data for theory (e.g. weak lensing peak statistics and simulated power spectra) and can reduce the number of simulations required. The results of this work have been implemented by independent groups in the cosmology community working with different datasets.

In Chapter 5, I demonstrated how mass map reconstructions can be improved using deep learning techniques trained on simulations. Compared with Wiener filtering, the (publicly available) DeepMass method was shown to improve mean-square-error by over 10 per

cent. I presented an interpretation of our newly created DES SV map as an approximation of the posterior mean $P(\kappa|\gamma)$ of the convergence given observed shear. The quality of the reconstruction with these initial experiments, and its flexibility, makes the deep learning approach a preeminent candidate for mass mapping with future weak lensing surveys.

Many of the techniques described in this thesis, which have often been demonstrated using DES data and simulations, will be applied to the next generation of surveys, Euclid and LSST, over a much larger sky area and with billions of galaxies.

6.2 Further work: likelihood-free inference with dark matter maps

In this section I will briefly outline some of the steps that would follow from the work in the previous chapters and which I plan to take in the near future. This will also provide some connections to developments in the field that have occurred simultaneously to the research presented in this thesis.

6.2.1 Motivation

In Chapter 4, I provided a framework to evaluate marginal posterior distributions of unknown cosmological parameters using noisy simulations to model the theory predictions. This relies on knowing the forms of the conditional probabilities, $P(\mathbf{d}|\mu)$ (e.g. the likelihood of measuring a power spectrum given a theoretically predicted power spectrum). In many cases these are not actually known, and forms are used which are approximate under certain conditions (e.g. C_ℓ is Gaussian for large ℓ).

The example of peak statistics was used as a situation in which we cannot predict the summary statistics from direct theoretical calculation, so we must rely on simulations. The corrected likelihood was calculated in Chapter 4 for the case of a flat prior and Gaussian naive likelihood. As previously discussed, these distributions can be changed depending on the statistical model, which may necessitate sampling or approximate inference methods. Whatever the method used to calculate the eventual result, we still need to question why we can assume that the naive likelihood for peak statistics is Gaussian (e.g. Kacprzak et al. 2016).

One common justification might be as follows. We could model the peak counts as Poissonian, as we did in Chapter 2. We could then use a standard result of the central limit theorem (see Sivia and Skilling 2006), which states that for large expected values $\bar{X}$, the

Poisson distributed variables X can be approximated as being Gaussian distributed. This justification presents three potential problems:

- If we have only a few peaks in a $\Delta\kappa$ bin, which will be the case for high κ, this approximation breaks down; a symmetric Gaussian will assign non-zero probability to negative counts.
- The variance of a Poissonian must vary with the mean, so we would need to calculate the covariance at every evaluated position in parameter space. This will require many more simulations than the approach where we assume the covariance matrix to be fixed.
- The original justification of the Poisson distribution is itself only an approximation.

The final problem is perhaps the worst. If an observable is the result of non-linear physics for which no closed-form expression exists, it is unlikely that its stochastic properties will be known either. However, if data simulations are realistic, they should include these properties, which we may hope to learn.

6.2.2 Compression and likelihood-free inference

Let us assume a data vector with a single element d_0 and a single unknown parameter θ_0. We can additionally assume that we do not know the likelihood, but can still simulate realistic noisy realisations $d_{\mathrm{sim},0}$ (as we could in Chapter 4). If the simulations are run at points in parameters space that are distributed according to a prior $P(\theta_0)$, then the density of points in the $d_{\mathrm{sim},0}$ - θ_0 plane corresponds to the joint distribution $P(d_{\mathrm{sim},0}, \theta_0)$.

The pyDELFI[1] (Alsing et al., 2018) method uses deep neural networks to estimate the probability density of the posterior $P(\theta_0|d_{\mathrm{obs},0})$ evaluated from the cloud of points in $\{d_{\mathrm{sim},0}, \theta_0\}$ space in the presence of the observed data $d_{\mathrm{obs},0}$ (see Lueckmann et al. 2018, Papamakarios et al. 2018, Alsing et al. 2019).

The joint or conditional densities can in principle be estimated in arbitrarily high dimension, though that would require more training data (simulations). If the dimension of the problem is reduced by compressing the data vector, the number of simulations required for an accurate estimate of the posterior distribution is also reduced.

Consider some likelihood $\mathscr{L} = P(\mathbf{d}|\theta)$. We can saturate the information inequality (equivalent to preserving the Fisher information) with

[1] Density Estimation Likelihood-Free Inference

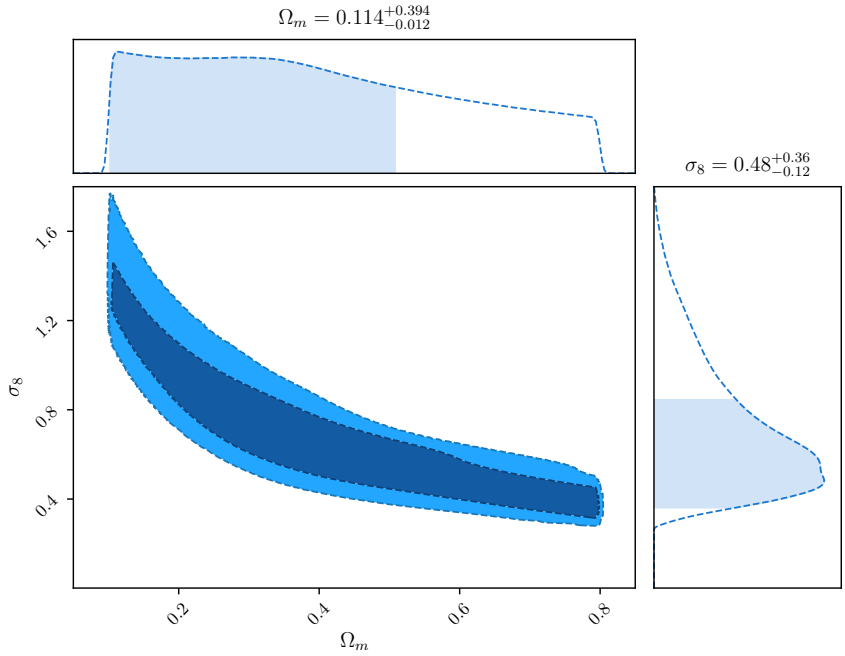

Figure 6.1: Posterior distribution for Ω_m and σ_8 from DES SV peak count data estimated using the simulations described in Chapter 5 with PYDELFI. This preliminary work is an illustrative example of likelihood-free inference using simulated training data.

$$\mathbf{t} = \nabla_\theta \mathscr{L}\Big|_{\theta_*} \quad , \tag{6.1}$$

where $\mathbf{t}$ is our compressed data vector and where the gradient is evaluated at a fiducial point in parameter space that corresponds to the maximum likelihood estimate (Alsing and Wandelt, 2018).

For a Gaussian with mean μ and fixed covariance Σ (following notation from Chapter 4), the compression evaluates to

$$\mathbf{t} = (\nabla_\theta \mu)^\dagger \Sigma^{-1} \mathbf{d} \quad . \tag{6.2}$$

This is equivalent to the MOPED[2] compression method, as proposed by Tegmark et al. (1997) and Heavens et al. (2000).

As a brief, but illustrative, example of this approach, figure 6.1 shows preliminary constraints for Ω_m and σ_8 from DES SV data in a single tomographic bin. This posterior distribution was estimated using PYDELFI with 1332 independent data realisations derived from the L-PICOLA simulations described in Chapter 5. The peak statistic data was

[2]Massively Optimised Parameter Estimation and Data compression

compressed using MOPED from 10 elements (κ bins between 0 and 0.028) to 2 elements (corresponding to the dimension of θ). The covariance matrix was calculated from 230 data realisations from around the fiducial point ($S_8 = 0.76$, $\Omega_m = 0.26$).

As would be expected from a single tomographic redshift bin (where I used the same selections as Chapter 5) of DES SV data, the parameter constraints are not particularly good[3]. As the convergence is non-Gaussian, by combining peaks and power spectrum data, we would hope to improve our constraints.

Figure 6.2 shows the individual constraints from power spectrum data, peaks count data and their combination. Unfortunately, the constraints from the joint dataset does not seem to improve very much over the individual datasets. In part, this is because the DES SV data is not very constraining, but I have also found (with tests using an assumed likelihood) that the MOPED compression is not lossless in this case. The joint compressed summary statistic has lost relevant information compared to the uncompressed joint summary statistic.

To improve over MOPED, we could solve equation 6.1, still assuming a fiducial Gaussian $\mathscr{L}$, but taking into account the $\nabla_\theta \Sigma$ terms. This non-linear compression would account for the covariance being a function of the underlying parameters, but would transform Gaussian distributed noise to non-Gaussian. If we were assuming a Gaussian form for the likelihood, this would introduce errors into our inference, but is perfectly allowed as we are learning the distribution directly using PYDELFI.

To implement this non-linear compression, we would again take a fiducial position in parameter space around which we would estimate the gradient of the covariance $\nabla_\theta \Sigma$. The gradient of the covariance requires many more simulations than the point-estimate of just the gradient. The number of simulations used for the linear MOPED compression may be insufficient for the next-order non-linear compression (i.e. it would introduce numerical errors). Even if this were numerically stable, we do not know *a priori* if the posterior distribution we are estimating can be characterised with a single well-defined peak about which we can expand (Graff et al., 2011).

However, given we are permitting non-linear data compression, which likelihood-free inference methods can easily deal with, using a deep learning compression approach is also permitted. To be able to estimate posterior distributions with likelihood-free methods, we require large training sets with simulated realistic data vectors. These can also be used to

[3] One can constrast with Kacprzak et al. (2016).

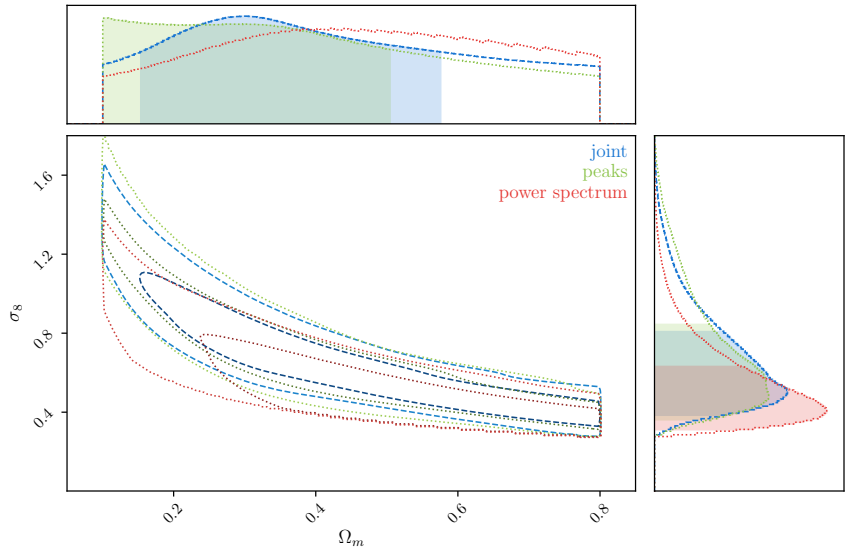

Figure 6.2: Posterior distribution for Ω_m and σ_8 from DES SV joint power spectrum and peak count data estimated using the simulations described in Chapter 5 with PYDELFI. This preliminary work demonstrates how combining peaks and power spectrum using MOPED does not significantly improve the constraints.

find a compressor function that performs better than equation 6.1, as no fiducial likelihood function needs to be assumed (e.g. Charnock et al. 2018). With this approach, we also do not need to ensure the training data for the compression step is independent, so we can take advantage of data augmentation (as described in Chapter 5).

We are therefore able to use deep learning *regression* to generate an estimate $\hat{\theta}$ of an unknown parameter and then use the same training data with likelihood-free inference to estimate a posterior distribution for that parameter with the compressed data $\mathbf{d} = \hat{\theta}$. This is quite an exciting prospect: representing uncertainties in predictions from machine learning in a principled Bayesian way.

The natural next step would be to consider using the deep learning techniques of Chapter 5 to extract a compressed statistic directly from the maps. In a recent paper, Fluri et al. (2019) use the KiDS-450 tomographic weak lensing dataset to constrain cosmological parameters using compressed statistic $\mathbf{t}$ derived from a CNN trained on simulated shear maps.

Fluri et al. (2019) take advantage of the corrected likelihood that I presented in Chapter 4. This implicitly assumes that their compressed statistic, if it could be modelled from theory, would have a Gaussian likelihood distribution. Given the series of non-linear operations applied to the data fed into the CNN, it might be difficult to justify this assumption of Gaussian noise. However, using mock data derived from the same simulations that were

used to train the CNN compressor, it would be possible to estimate the posterior distribution directly with likelihood-free inference techniques.

A greater challenge, which I will discuss in the next section, can be encapsulated in the question: "How do you know what the CNN is learning?"

6.2.3 Simulations and interpretability

In Chapter 5, I described how the convergence maps derived from the L-PICOLA simulation were rescaled in harmonic space to account for lack of power at small scales. Figure 6.3 shows, for one particular cosmology[4], that the power from L-PICOLA is suppressed in comparison to the theoretical prediction. The shaded bands represent 1 and 2σ standard deviation due to cosmic variance across the whole sky.

For the full-sky case the L-PICOLA simulation shows significant (2σ) deviation from the theoretical prediction for $\ell > 500$. This is not a particularly high multipole; in future work the resolution of the simulations can be improved to reduce the angular scale where deviations occur, but this would not guarantee that other statistics of the map are correct.

We could say that the minimum requirement for our training data used in deep learning map reconstruction (Chapter 5) would be to have convergence maps that were more realistic than a Gaussian random field. By having the L-PICOLA cosmic-web simulations as described in Chapter 5, with the corrected angular power, we more than fulfilled this requirement.

If we went on and used these simulations to construct a CNN compressed statistic, we would have to be more certain in the properties of the simulation. With the trained CNN, it is difficult to understand what the learned features correspond to.

We can imagine, as is likely, that the training maps had the incorrect 3-point statistics, $\langle \kappa^3 \rangle$. If a map-based CNN learned to include bispectrum information in the compressed statistic, as is also likely, then the final inference step would give an incorrect posterior distribution. Furthermore, even if the CNN only incorporated 2-point information $\langle \kappa^2 \rangle$ in the compressed statistic (which is unlikely), then the final posterior distribution would still be incorrect as PYDELFI would implicitly use higher-order statistics $\langle \kappa^n \rangle$ to constrain the form of the posterior.

This presents some interesting future challenges: quantifying the accuracy of simulations and interpreting what a deep learning method is really learning from the data.

[4] i.e. one set of cosmological parameters in ΛCDM.

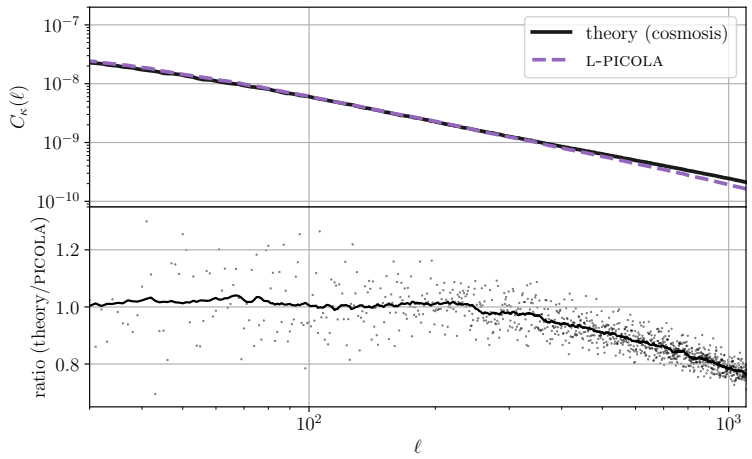

Figure 6.3: Angular power spectrum of convergence $C_\kappa(\ell)$ with multipole ℓ from theory (NICAEA) and L-PICOLA simuation. The shaded 1 and 2σ regions correspond to full-sky due to cosmic variance.

These challenges may seem difficult, but by tackling them we may be able to use dark matter maps to answer cosmological questions that traditional methods cannot answer. In the era of Euclid and LSST, this will allow us to improve cosmological parameter constraints and challenge the standard model in new ways.

6.3 Disciplines and disciples

A researcher working in cosmology, who has never spoken to a particle physicist, does not really understand the good nor the bad of the cosmologists' approach. This is an old problem, and an old example. Cosmologists and particle physicists often do, in fact, take time to discuss their respective approaches.

The often discussed (argued about) frequentist approach in particle physics, employed in Higgs particle analysis for example, is probably the result of history, philosophy, and lack of necessity. When faced with a problem of statistical interest, a particle physicist is afforded the luxury of rerunning the experiment, whereas a cosmologist struggles to rerun the Universe.

There are, of course, Bayesian particle physicists and many astronomers who would not identify as either Bayesian or frequentist. The point is to note that with a large experiment, once a course has been decided, it is difficult to change. Analysis choices becomes frozen in. The ever-present risk of stagnation in observational cosmology is probably greater for the titanic communities in high energy physics.

It is a worry that in our focussed discussions of covariance matrices and tensions, we may miss new approaches that would make covariance matrices a thing of history. We are lucky to be working during what is often called an *Artificial Intelligence revolution*. With large amounts of observed and simulated data, one might hope that a weak lensing experiment in the decades to come will not rely on what DES are currently doing.

Looking for the next brilliant step in computer learning that would revolutionise cosmology is difficult. It requires paying attention outside the field, putting effort into incomplete ideas, and comes with the risk of wasted time if nothing is found. The alternative, to keep within one's own scientific community, is more comfortable. In an inflating Universe of new ideas, an observer's horizon can easily shrink.

"*And what should they know of England$_{[,]}$ who only England know?*"

Rudyard Kipling, *The English Flag*, 1891

Bibliography

Abbott, B. P., Abbott, R., Abbott, T. D., Abernathy, M. R., Acernese, F., Ackley, K., Adams, C., Adams, T., Addesso, P., Adhikari, R. X., Adya, V. B., Affeldt, C., Agathos, M., Agatsuma, K., Aggarwal, N., Aguiar, O. D., Aiello, L., Ain, A., Ajith, P., Allen, B., Allocca, A., Altin, P. A., Anderson, S. B., Anderson, W. G., Arai, K., Arain, M. A., Araya, M. C., Arceneaux, C. C., Areeda, J. S., Arnaud, N., Arun, K. G., Ascenzi, S., Ashton, G., Ast, M., Aston, S. M., Astone, P., Aufmuth, P., Aulbert, C., Babak, S., Bacon, P., Bader, M. K. M., Baker, P. T., Baldaccini, F., Ballardin, G., Ballmer, S. W., Barayoga, J. C., Barclay, S. E., Barish, B. C., Barker, D., Barone, F., Barr, B., Barsotti, L., Barsuglia, M., Barta, D., Bartlett, J., Barton, M. A., Bartos, I., Bassiri, R., Basti, A., Batch, J. C., Baune, C., Bavigadda, V., Bazzan, M., Behnke, B., Bejger, M., Belczynski, C., Bell, A. S., Bell, C. J., Berger, B. K., Bergman, J., Bergmann, G., Berry, C. P. L., Bersanetti, D., Bertolini, A., Betzwieser, J., Bhagwat, S., Bhandare, R., Bilenko, I. A., Billingsley, G., Birch, J., Birney, R., Birnholtz, O., Biscans, S., Bisht, A., Bitossi, M., Biwer, C., Bizouard, M. A., Blackburn, J. K., Blair, C. D., Blair, D. G., Blair, R. M., Bloemen, S., Bock, O., Bodiya, T. P., Boer, M., Bogaert, G., Bogan, C., Bohe, A., Bojtos, P., Bond, C., Bondu, F., Bonnand, R., Boom, B. A., Bork, R., Boschi, V., Bose, S., Bouffanais, Y., Bozzi, A., Bradaschia, C., Brady, P. R., Braginsky, V. B., Branchesi, M., Brau, J. E., Briant, T., Brillet, A., Brinkmann, M., Brisson, V., Brockill, P., Brooks, A. F., Brown, D. A., Brown, D. D., Brown, N. M., Buchanan, C. C., Buikema, A., Bulik, T., Bulten, H. J., Buonanno, A., Buskulic, D., Buy, C., Byer, R. L., Cabero, M., Cadonati, L., Cagnoli, G., Cahillane, C., Bustillo, J. C., Callister, T., Calloni, E., Camp, J. B., Cannon, K. C., Cao, J., Capano, C. D., Capocasa, E., Carbognani, F., Caride, S., Casanueva Diaz, J., Casentini, C., Caudill, S., Cavaglià, M., Cavalier, F., Cavalieri, R., Cella, G., Cepeda, C. B., Baiardi, L. C., Cerretani, G., Cesarini, E., Chakraborty, R., Chalermsongsak, T., Chamberlin, S. J., Chan, M., Chao, S., Charlton, P., Chassand e-Mottin, E., Chen, H. Y.,

Chen, Y., Cheng, C., Chincarini, A., Chiummo, A., Cho, H. S., Cho, M., Chow, J. H., Christensen, N., Chu, Q., Chua, S., Chung, S., Ciani, G., Clara, F., Clark, J. A., Cleva, F., Coccia, E., Cohadon, P. F., Colla, A., Collette, C. G., Cominsky, L., Constancio, M., Conte, A., Conti, L., Cook, D., Corbitt, T. R., Cornish, N., Corsi, A., Cortese, S., Costa, C. A., Coughlin, M. W., Coughlin, S. B., Coulon, J. P., Countryman, S. T., Couvares, P., Cowan, E. E., Coward, D. M., Cowart, M. J., Coyne, D. C., Coyne, R., Craig, K., Creighton, J. D. E., Creighton, T. D., Cripe, J., Crowder, S. G., Cruise, A. M., Cumming, A., Cunningham, L., Cuoco, E., Dal Canton, T., Danilishin, S. L., D'Antonio, S., Danzmann, K., Darman, N. S., Da Silva Costa, C. F., Dattilo, V., Dave, I., Daveloza, H. P., Davier, M., Davies, G. S., Daw, E. J., Day, R., De, S., DeBra, D., Debreczeni, G., Degallaix, J., De Laurentis, M., Deléglise, S., Del Pozzo, W., Denker, T., Dent, T., Dereli, H., Dergachev, V., DeRosa, R. T., De Rosa, R., DeSalvo, R., Dhurandhar, S., Díaz, M. C., Di Fiore, L., Di Giovanni, M., Di Lieto, A., Di Pace, S., Di Palma, I., Di Virgilio, A., Dojcinoski, G., Dolique, V., Donovan, F., Dooley, K. L., Doravari, S., Douglas, R., Downes, T. P., Drago, M., Drever, R. W. P., Driggers, J. C., Du, Z., Ducrot, M., Dwyer, S. E., Edo, T. B., Edwards, M. C., Effler, A., Eggenstein, H. B., Ehrens, P., Eichholz, J., Eikenberry, S. S., Engels, W., Essick, R. C., Etzel, T., Evans, M., Evans, T. M., Everett, R., Factourovich, M., Fafone, V., Fair, H., Fairhurst, S., Fan, X., Fang, Q., Farinon, S., Farr, B., Farr, W. M., Favata, M., Fays, M., Fehrmann, H., Fejer, M. M., Feldbaum, D., Ferrante, I., Ferreira, E. C., Ferrini, F., Fidecaro, F., Finn, L. S., Fiori, I., Fiorucci, D., Fisher, R. P., Flaminio, R., Fletcher, M., Fong, H., Fournier, J. D., Franco, S., Frasca, S., Frasconi, F., Frede, M., Frei, Z., Freise, A., Frey, R., Frey, V., Fricke, T. T., Fritschel, P., Frolov, V. V., Fulda, P., Fyffe, M., Gabbard, H. A. G., Gair, J. R., Gammaitoni, L., Gaonkar, S. G., Garufi, F., Gatto, A., Gaur, G., Gehrels, N., Gemme, G., Gendre, B., Genin, E., Gennai, A., George, J., Gergely, L., Germain, V., Ghosh, A., Ghosh, A., Ghosh, S., Giaime, J. A., Giardina, K. D., Giazotto, A., Gill, K., Glaefke, A., Gleason, J. R., Goetz, E., Goetz, R., Gondan, L., González, G., Castro, J. M. G., Gopakumar, A., Gordon, N. A., Gorodetsky, M. L., Gossan, S. E., Gosselin, M., Gouaty, R., Graef, C., Graff, P. B., Granata, M., Grant, A., Gras, S., Gray, C., Greco, G., Green, A. C., Greenhalgh, R. J. S., Groot, P., Grote, H., Grunewald, S., Guidi, G. M., Guo, X., Gupta, A., Gupta, M. K., Gushwa, K. E., Gustafson, E. K., Gustafson, R., Hacker, J. J., Hall, B. R., Hall, E. D., Hammond, G., Haney, M., Hanke, M. M., Hanks, J., Hanna, C.,

Hannam, M. D., Hanson, J., Hardwick, T., Harms, J., Harry, G. M., Harry, I. W., Hart, M. J., Hartman, M. T., Haster, C. J., Haughian, K., Healy, J., Heefner, J., Heidmann, A., Heintze, M. C., Heinzel, G., Heitmann, H., Hello, P., Hemming, G., Hendry, M., Heng, I. S., Hennig, J., Heptonstall, A. W., Heurs, M., Hild, S., Hoak, D., Hodge, K. A., Hofman, D., Hollitt, S. E., Holt, K., Holz, D. E., Hopkins, P., Hosken, D. J., Hough, J., Houston, E. A., Howell, E. J., Hu, Y. M., Huang, S., Huerta, E. A., Huet, D., Hughey, B., Husa, S., Huttner, S. H., Huynh-Dinh, T., Idrisy, A., Indik, N., Ingram, D. R., Inta, R., Isa, H. N., Isac, J. M., Isi, M., Islas, G., Isogai, T., Iyer, B. R., Izumi, K., Jacobson, M. B., Jacqmin, T., Jang, H., Jani, K., Jaranowski, P., Jawahar, S., Jiménez-Forteza, F., Johnson, W. W., Johnson-McDaniel, N. K., Jones, D. I., Jones, R., Jonker, R. J. G., Ju, L., Haris, K., Kalaghatgi, C. V., Kalogera, V., Kandhasamy, S., Kang, G., Kanner, J. B., Karki, S., Kasprzack, M., Katsavounidis, E., Katzman, W., Kaufer, S., Kaur, T., Kawabe, K., Kawazoe, F., Kéfélian, F., Kehl, M. S., Keitel, D., Kelley, D. B., Kells, W., Kennedy, R., Keppel, D. G., Key, J. S., Khalaidovski, A., Khalili, F. Y., Khan, I., Khan, S., Khan, Z., Khazanov, E. A., Kijbunchoo, N., Kim, C., Kim, J., Kim, K., Kim, N.-G., Kim, N., Kim, Y. M., King, E. J., King, P. J., Kinzel, D. L., Kissel, J. S., Kleybolte, L., Klimenko, S., Koehlenbeck, S. M., Kokeyama, K., Koley, S., Kondrashov, V., Kontos, A., Koranda, S., Korobko, M., Korth, W. Z., Kowalska, I., Kozak, D. B., Kringel, V., Krishnan, B., Królak, A., Krueger, C., Kuehn, G., Kumar, P., Kumar, R., Kuo, L., Kutynia, A., Kwee, P., Lackey, B. D., Landry, M., Lange, J., Lantz, B., Lasky, P. D., Lazzarini, A., Lazzaro, C., Leaci, P., Leavey, S., Lebigot, E. O., Lee, C. H., Lee, H. K., Lee, H. M., Lee, K., Lenon, A., Leonardi, M., Leong, J. R., Leroy, N., Letendre, N., Levin, Y., Levine, B. M., Li, T. G. F., Libson, A., Littenberg, T. B., Lockerbie, N. A., Logue, J., Lombardi, A. L., London, L. T., Lord, J. E., Lorenzini, M., Loriette, V., Lormand, M., Losurdo, G., Lough, J. D., Lousto, C. O., Lovelace, G., Lück, H., Lundgren, A. P., Luo, J., Lynch, R., Ma, Y., MacDonald, T., Machenschalk, B., MacInnis, M., Macleod, D. M., Magaña-Sandoval, F., Magee, R. M., Mageswaran, M., Majorana, E., Maksimovic, I., Malvezzi, V., Man, N., Mandel, I., Mandic, V., Mangano, V., Mansell, G. L., Manske, M., Mantovani, M., Marchesoni, F., Marion, F., Márka, S., Márka, Z., Markosyan, A. S., Maros, E., Martelli, F., Martellini, L., Martin, I. W., Martin, R. M., Martynov, D. V., Marx, J. N., Mason, K., Masserot, A., Massinger, T. J., Masso-Reid, M., Matichard, F., Matone, L., Mavalvala, N., Mazumder, N., Mazzolo, G., McCarthy, R., McClelland, D. E., McCormick, S.,

McGuire, S. C., McIntyre, G., McIver, J., McManus, D. J., McWilliams, S. T., Meacher, D., Meadors, G. D., Meidam, J., Melatos, A., Mendell, G., Mendoza-Gandara, D., Mercer, R. A., Merilh, E., Merzougui, M., Meshkov, S., Messenger, C., Messick, C., Meyers, P. M., Mezzani, F., Miao, H., Michel, C., Middleton, H., Mikhailov, E. E., Milano, L., Miller, J., Millhouse, M., Minenkov, Y., Ming, J., Mirshekari, S., Mishra, C., Mitra, S., Mitrofanov, V. P., Mitselmakher, G., Mittleman, R., Moggi, A., Mohan, M., Mohapatra, S. R. P., Montani, M., Moore, B. C., Moore, C. J., Moraru, D., Moreno, G., Morriss, S. R., Mossavi, K., Mours, B., Mow-Lowry, C. M., Mueller, C. L., Mueller, G., Muir, A. W., Mukherjee, A., Mukherjee, D., Mukherjee, S., Mukund, N., Mullavey, A., Munch, J., Murphy, D. J., Murray, P. G., Mytidis, A., Nardecchia, I., Naticchioni, L., Nayak, R. K., Necula, V., Nedkova, K., Nelemans, G., Neri, M., Neunzert, A., Newton, G., Nguyen, T. T., Nielsen, A. B., Nissanke, S., Nitz, A., Nocera, F., Nolting, D., Normandin, M. E. N., Nuttall, L. K., Oberling, J., Ochsner, E., O'Dell, J., Oelker, E., Ogin, G. H., Oh, J. J., Oh, S. H., Ohme, F., Oliver, M., Oppermann, P., Oram, R. J., O'Reilly, B., O'Shaughnessy, R., Ott, C. D., Ottaway, D. J., Ottens, R. S., Overmier, H., Owen, B. J., Pai, A., Pai, S. A., Palamos, J. R., Palashov, O., Palomba, C., Pal-Singh, A., Pan, H., Pan, Y., Pankow, C., Pannarale, F., Pant, B. C., Paoletti, F., Paoli, A., Papa, M. A., Paris, H. R., Parker, W., Pascucci, D., Pasqualetti, A., Passaquieti, R., Passuello, D., Patricelli, B., Patrick, Z., Pearlstone, B. L., Pedraza, M., Pedurand , R., Pekowsky, L., Pele, A., Penn, S., Perreca, A., Pfeiffer, H. P., Phelps, M., Piccinni, O., Pichot, M., Pickenpack, M., Piergiovanni, F., Pierro, V., Pillant, G., Pinard, L., Pinto, I. M., Pitkin, M., Poeld, J. H., Poggiani, R., Popolizio, P., Post, A., Powell, J., Prasad, J., Predoi, V., Premachandra, S. S., Prestegard, T., Price, L. R., Prijatelj, M., Principe, M., Privitera, S., Prix, R., Prodi, G. A., Prokhorov, L., Puncken, O., Punturo, M., Puppo, P., Pürrer, M., Qi, H., Qin, J., Quetschke, V., Quintero, E. A., Quitzow-James, R., Raab, F. J., Rabeling, D. S., Radkins, H., Raffai, P., Raja, S., Rakhmanov, M., Ramet, C. R., Rapagnani, P., Raymond, V., Razzano, M., Re, V., Read, J., Reed, C. M., Regimbau, T., Rei, L., Reid, S., Reitze, D. H., Rew, H., Reyes, S. D., Ricci, F., Riles, K., Robertson, N. A., Robie, R., Robinet, F., Rocchi, A., Rolland, L., Rollins, J. G., Roma, V. J., Romano, J. D., Romano, R., Romanov, G., Romie, J. H., Rosińska, D., Rowan, S., Rüdiger, A., Ruggi, P., Ryan, K., Sachdev, S., Sadecki, T., Sadeghian, L., Salconi, L., Saleem, M., Salemi, F., Samajdar, A., Sammut, L., Sampson, L. M., Sanchez, E. J., Sandberg, V., Sandeen, B., Sand ers, G. H., Sanders, J. R.,

Sassolas, B., Sathyaprakash, B. S., Saulson, P. R., Sauter, O., Savage, R. L., Sawadsky, A., Schale, P., Schilling, R., Schmidt, J., Schmidt, P., Schnabel, R., Schofield, R. M. S., Schönbeck, A., Schreiber, E., Schuette, D., Schutz, B. F., Scott, J., Scott, S. M., Sellers, D., Sengupta, A. S., Sentenac, D., Sequino, V., Sergeev, A., Serna, G., Setyawati, Y., Sevigny, A., Shaddock, D. A., Shaffer, T., Shah, S., Shahriar, M. S., Shaltev, M., Shao, Z., Shapiro, B., Shawhan, P., Sheperd, A., Shoemaker, D. H., Shoemaker, D. M., Siellez, K., Siemens, X., Sigg, D., Silva, A. D., Simakov, D., Singer, A., Singer, L. P., Singh, A., Singh, R., Singhal, A., Sintes, A. M., Slagmolen, B. J. J., Smith, J. R., Smith, M. R., Smith, N. D., Smith, R. J. E., Son, E. J., Sorazu, B., Sorrentino, F., Souradeep, T., Srivastava, A. K., Staley, A., Steinke, M., Steinlechner, J., Steinlechner, S., Steinmeyer, D., Stephens, B. C., Stevenson, S. P., Stone, R., Strain, K. A., Straniero, N., Stratta, G., Strauss, N. A., Strigin, S., Sturani, R., Stuver, A. L., Summerscales, T. Z., Sun, L., Sutton, P. J., Swinkels, B. L., Szczepańczyk, M. J., Tacca, M., Talukder, D., Tanner, D. B., Tápai, M., Tarabrin, S. P., Taracchini, A., Taylor, R., Theeg, T., Thirugnanasambandam, M. P., Thomas, E. G., Thomas, M., Thomas, P., Thorne, K. A., Thorne, K. S., Thrane, E., Tiwari, S., Tiwari, V., Tokmakov, K. V., Tomlinson, C., Tonelli, M., Torres, C. V., Torrie, C. I., Töyrä, D., Travasso, F., Traylor, G., Trifirò, D., Tringali, M. C., Trozzo, L., Tse, M., Turconi, M., Tuyenbayev, D., Ugolini, D., Unnikrishnan, C. S., Urban, A. L., Usman, S. A., Vahlbruch, H., Vajente, G., Valdes, G., Vallisneri, M., van Bakel, N., van Beuzekom, M., van den Brand, J. F. J., Van Den Broeck, C., Vand er-Hyde, D. C., van der Schaaf, L., van Heijningen, J. V., van Veggel, A. A., Vardaro, M., Vass, S., Vasúth, M., Vaulin, R., Vecchio, A., Vedovato, G., Veitch, J., Veitch, P. J., Venkateswara, K., Verkindt, D., Vetrano, F., Viceré, A., Vinciguerra, S., Vine, D. J., Vinet, J. Y., Vitale, S., Vo, T., Vocca, H., Vorvick, C., Voss, D., Vousden, W. D., Vyatchanin, S. P., Wade, A. R., Wade, L. E., Wade, M., Waldman, S. J., Walker, M., Wallace, L., Walsh, S., Wang, G., Wang, H., Wang, M., Wang, X., Wang, Y., Ward, H., Ward, R. L., Warner, J., Was, M., Weaver, B., Wei, L. W., Weinert, M., Weinstein, A. J., Weiss, R., Welborn, T., Wen, L., Weßels, P., Westphal, T., Wette, K., Whelan, J. T., Whitcomb, S. E., White, D. J., Whiting, B. F., Wiesner, K., Wilkinson, C., Willems, P. A., Williams, L., Williams, R. D., Williamson, A. R., Willis, J. L., Willke, B., Wimmer, M. H., Winkelmann, L., Winkler, W., Wipf, C. C., Wiseman, A. G., Wittel, H., Woan, G., Worden, J., Wright, J. L., Wu, G., Yablon, J., Yakushin, I., Yam, W., Yamamoto, H., Yancey, C. C., Yap, M. J., Yu, H.,

Yvert, M., ZadroŻny, A., Zangrando, L., Zanolin, M., Zendri, J. P., Zevin, M., Zhang, F., Zhang, L., Zhang, M., Zhang, Y., Zhao, C., Zhou, M., Zhou, Z., Zhu, X. J., Zucker, M. E., Zuraw, S. E., Zweizig, J., LIGO Scientific Collaboration, and Virgo Collaboration (2016). Observation of Gravitational Waves from a Binary Black Hole Merger. *Phys. Rev. Lett.*, 116(6):061102.

Abbott, T. M. C., Abdalla, F. B., Alarcon, A., Aleksić, J., Allam, S., Allen, S., Amara, A., Annis, J., Asorey, J., Avila, S., and et al. (2018). Dark Energy Survey year 1 results: Cosmological constraints from galaxy clustering and weak lensing. *Phys. Rev. D*, 98(4):043526.

Abolfathi, B., Aguado, D. S., Aguilar, G., Allende Prieto, C., Almeida, A., Ananna, T. T., Anders, F., Anderson, S. F., Andrews, B. H., Anguiano, B., and et al. (2018). The Fourteenth Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the Extended Baryon Oscillation Spectroscopic Survey and from the Second Phase of the Apache Point Observatory Galactic Evolution Experiment. *ApJS*, 235:42.

Akima, H. (1970). A new method of interpolation and smooth curve fitting based on local procedures. *Journal of the ACM (JACM)*, 17(4):589–602.

Alam, S., Ata, M., Bailey, S., Beutler, F., Bizyaev, D., Blazek, J. A., Bolton, A. S., Brownstein, J. R., Burden, A., Chuang, C.-H., Comparat, J., Cuesta, A. J., Dawson, K. S., Eisenstein, D. J., Escoffier, S., Gil-Marín, H., Grieb, J. N., Hand, N., Ho, S., Kinemuchi, K., Kirkby, D., Kitaura, F., Malanushenko, E., Malanushenko, V., Maraston, C., McBride, C. K., Nichol, R. C., Olmstead, M. D., Oravetz, D., Padmanabhan, N., Palanque-Delabrouille, N., Pan, K., Pellejero-Ibanez, M., Percival, W. J., Petitjean, P., Prada, F., Price-Whelan, A. M., Reid, B. A., Rodríguez-Torres, S. A., Roe, N. A., Ross, A. J., Ross, N. P., Rossi, G., Rubiño-Martín, J. A., Saito, S., Salazar-Albornoz, S., Samushia, L., Sánchez, A. G., Satpathy, S., Schlegel, D. J., Schneider, D. P., Scóccola, C. G., Seo, H.-J., Sheldon, E. S., Simmons, A., Slosar, A., Strauss, M. A., Swanson, M. E. C., Thomas, D., Tinker, J. L., Tojeiro, R., Magaña, M. V., Vazquez, J. A., Verde, L., Wake, D. A., Wang, Y., Weinberg, D. H., White, M., Wood-Vasey, W. M., Yèche, C., Zehavi, I., Zhai, Z., and Zhao, G.-B. (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. and Steinhardt, P. J. (1982). Cosmology for grand unified theories with radiatively induced symmetry breaking. *Physical Review Letters*, 48:1220–1223.

Allen, T. J., Grinstein, B., and Wise, M. B. (1987). Non-gaussian density perturbations in inflationary cosmologies. *Physics Letters B*, 197:66–70.

Alsing, J., Charnock, T., Feeney, S., and Wand elt, B. (2019). Fast likelihood-free cosmology with neural density estimators and active learning. *MNRAS*, 488(3):4440–4458.

Alsing, J., Heavens, A., and Jaffe, A. H. (2017). Cosmological parameters, shear maps and power spectra from CFHTLenS using Bayesian hierarchical inference. *Monthly Notices of the Royal Astronomical Society (MNRAS)*, 466:3272–3292.

Alsing, J., Heavens, A., Jaffe, A. H., Kiessling, A., Wandelt, B., and Hoffmann, T. (2016a). Hierarchical cosmic shear power spectrum inference. *MNRAS*, 455:4452–4466.

Alsing, J., Heavens, A., Jaffe, A. H., Kiessling, A., Wandelt, B., and Hoffmann, T. (2016b). Hierarchical cosmic shear power spectrum inference. *Monthly Notices of the Royal Astronomical Society (MNRAS)*, 455:4452–4466.

Alsing, J. and Wandelt, B. (2018). Generalized massive optimal data compression. *MNRAS*, 476(1):L60–L64.

Alsing, J., Wandelt, B., and Feeney, S. (2018). Massive optimal data compression and density estimation for scalable, likelihood-free inference in cosmology. *MNRAS*, 477:2874–2885.

Amendola, L., Appleby, S., Avgoustidis, A., Bacon, D., Baker, T., Baldi, M., Bartolo, N., Blanchard, A., Bonvin, C., Borgani, S., Branchini, E., Burrage, C., Camera, S., Carbone, C., Casarini, L., Cropper, M., de Rham, C., Dietrich, J. P., Di Porto, C., Durrer, R., Ealet, A., Ferreira, P. G., Finelli, F., Garcia-Bellido, J., Giannantonio, T., Guzzo, L., Heavens, A., Heisenberg, L., Heymans, C., Hoekstra, H., Hollenstein, L., Holmes, R., Horst, O., Hwang, Z., Jahnke, K., Kitching, T. D., Koivisto, T., Kunz, M., La Vacca, G., Linder, E., March, M., Marra, V., Martins, C., Majerotto, E., Markovic, D., Marsh, D., Marulli, F., Massey, R., Mellier, Y., Montanari, F., Mota, D. F., Nunes, N. J., Percival, W., Pettorino, V., Porciani, C., Quercellini, C., Read, J., Rinaldi, M., Sapone, D., Sawicki, I., Scaramella, R., Skordis, C., Simpson, F., Taylor, A., Thomas, S., Trotta, R., Verde, L.,

Vernizzi, F., Vollmer, A., Wang, Y., Weller, J., and Zlosnik, T. (2016). Cosmology and Fundamental Physics with the Euclid Satellite. *ArXiv e-prints*.

Amendola, L., Appleby, S., Avgoustidis, A., Bacon, D., Baker, T., Baldi, M., Bartolo, N., Blanchard, A., Bonvin, C., Borgani, S., et al., and (Euclid Theory Working Group) (2018). Cosmology and fundamental physics with the Euclid satellite. *Living Reviews in Relativity*, 21:2.

Anderson, T. (2003). *An Introduction to Multivariate Statistical Analysis*. Wiley Series in Probability and Statistics. Wiley.

Arram, J., Pflanzer, M., Kaplan, T., and Luk, W. (2015). Fpga acceleration of reference-based compression for genomic data. In *Field Programmable Technology (FPT), 2015 International Conference on*, pages 9–16. IEEE.

Audren, B. and Lesgourgues, J. (2011). Non-linear matter power spectrum from Time Renormalisation Group: efficient computation and comparison with one-loop. *Journal of Cosmology and Astro-Particle Physics*, 2011(10):037.

Bacon, D. J., Goldberg, D. M., Rowe, B. T. P., and Taylor, A. N. (2006). Weak gravitational flexion. *MNRAS*, 365:414–428.

Bardeen, J. M. (1980). Gauge-invariant cosmological perturbations. *Phys. Rev. D*, 22:1882–1905.

Bardeen, J. M., Bond, J. R., Kaiser, N., and Szalay, A. S. (1986). The statistics of peaks of Gaussian random fields. *ApJ*, 304:15–61.

Bardeen, J. M., Steinhardt, P. J., and Turner, M. S. (1983). Spontaneous creation of almost scale-free density perturbations in an inflationary universe. *Phys. Rev. D*, 28:679–693.

Bartelmann, M. and Schneider, P. (2001). Weak gravitational lensing. *Phys. Rep.*, 340:291–472.

Baumann, D. (2009). TASI Lectures on Inflation. *arXiv e-prints*, page arXiv:0907.5424.

Baumann, D. (2014). Cosmology, part iii mathematical tripos. http://www.damtp.cam.ac.uk/user/db275/Cosmology/Lectures.pdf - accessed 31/01/17.

Bayes, T. (1763). LII. an essay towards solving a problem in the doctrine of chances. by the late rev. mr. bayes, f. r. s. communicated by mr. price, in a letter to john canton, a. m. f. r. s. *Philosophical Transactions of the Royal Society of London*, 53:370–418.

Becker, M. R. (2013). CALCLENS: weak lensing simulations for large-area sky surveys and second-order effects in cosmic shear power spectra. *MNRAS*, 435:115–132.

Benítez, N. (2000). Bayesian Photometric Redshift Estimation. *ApJ*, 536:571–583.

Bernardeau, F., Colombi, S., Gaztañaga, E., and Scoccimarro, R. (2002). Large-scale structure of the Universe and cosmological perturbation theory. *Physics Reports*, 367(1-3):1–248.

Betoule, M., Kessler, R., Guy, J., Mosher, J., Hardin, D., Biswas, R., Astier, P., El-Hage, P., Konig, M., Kuhlmann, S., Marriner, J., Pain, R., Regnault, N., Balland, C., Bassett, B. A., Brown, P. J., Campbell, H., Carlberg, R. G., Cellier-Holzem, F., Cinabro, D., Conley, A., D'Andrea, C. B., DePoy, D. L., Doi, M., Ellis, R. S., Fabbro, S., Filippenko, A. V., Foley, R. J., Frieman, J. A., Fouchez, D., Galbany, L., Goobar, A., Gupta, R. R., Hill, G. J., Hlozek, R., Hogan, C. J., Hook, I. M., Howell, D. A., Jha, S. W., Le Guillou, L., Leloudas, G., Lidman, C., Marshall, J. L., Möller, A., Mourão, A. M., Neveu, J., Nichol, R., Olmstead, M. D., Palanque-Delabrouille, N., Perlmutter, S., Prieto, J. L., Pritchet, C. J., Richmond, M., Riess, A. G., Ruhlmann-Kleider, V., Sako, M., Schahmaneche, K., Schneider, D. P., Smith, M., Sollerman, J., Sullivan, M., Walton, N. A., and Wheeler, C. J. (2014). Improved cosmological constraints from a joint analysis of the SDSS-II and SNLS supernova samples. *A&A*, 568:A22.

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 Astro-Particle Physics*, 2019(2):050.

Bishop, C. M. (2011). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer.

Blandford, R. and Narayan, R. (1986). Fermat's Principle, Caustics, and the Classification of Gravitational Lens Images. *ApJ*, 310:568.

Böhm, V., Hilbert, S., Greiner, M., and Enßlin, T. A. (2017). Bayesian weak lensing tomography: Reconstructing the 3D large-scale distribution of matter with a lognormal prior. *Phys. Rev. D*, 96(12):123510.

Bonnett, C., Troxel, M. A., Hartley, W., Amara, A., Leistedt, B., Becker, M. R., Bernstein, G. M., Bridle, S. L., Bruderer, C., Busha, M. T., Carrasco Kind, M., Childress, M. J., Castander, F. J., Chang, C., Crocce, M., Davis, T. M., Eifler, T. F., Frieman, J., Gangkofner, C., Gaztanaga, E., Glazebrook, K., Gruen, D., Kacprzak, T., King, A., Kwan, J., Lahav, O., Lewis, G., Lidman, C., Lin, H., MacCrann, N., Miquel, R., O'Neill, C. R., Palmese, A., Peiris, H. V., Refregier, A., Rozo, E., Rykoff, E. S., Sadeh, I., Sánchez, C., Sheldon, E., Uddin, S., Wechsler, R. H., Zuntz, J., Abbott, T., Abdalla, F. B., Allam, S., Armstrong, R., Banerji, M., Bauer, A. H., Benoit-Lévy, A., Bertin, E., Brooks, D., Buckley-Geer, E., Burke, D. L., Capozzi, D., Carnero Rosell, A., Carretero, J., Cunha, C. E., D'Andrea, C. B., da Costa, L. N., DePoy, D. L., Desai, S., Diehl, H. T., Dietrich, J. P., Doel, P., Fausti Neto, A., Fernandez, E., Flaugher, B., Fosalba, P., Gerdes, D. W., Gruendl, R. A., Honscheid, K., Jain, B., James, D. J., Jarvis, M., Kim, A. G., Kuehn, K., Kuropatkin, N., Li, T. S., Lima, M., Maia, M. A. G., March, M., Marshall, J. L., Martini, P., Melchior, P., Miller, C. J., Neilsen, E., Nichol, R. C., Nord, B., Ogando, R., Plazas, A. A., Reil, K., Romer, A. K., Roodman, A., Sako, M., Sanchez, E., Santiago, B., Smith, R. C., Soares-Santos, M., Sobreira, F., Suchyta, E., Swanson, M. E. C., Tarle, G., Thaler, J., Thomas, D., Vikram, V., Walker, A. R., and Dark Energy Survey Collaboration (2016). Redshift distributions of galaxies in the Dark Energy Survey Science Verification shear catalogue and implications for weak lensing. *Phys. Rev. D*, 94(4):042005.

Bookstein, F. L. (1989). Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on pattern analysis and machine intelligence*, 11(6):567–585.

Bouchet, F. R., Colombi, S., Hivon, E., and Juszkiewicz, R. (1995). Perturbative Lagrangian approach to gravitational instability. *Astronomy and Astrophysics*, 296:575.

Bromiley, P. (2003). Products and convolutions of gaussian probability density functions. *Tina-Vision Memo*, 3(4):1.

Buchert, T. and Räsänen, S. (2012). Backreaction in Late-Time Cosmology. *Annual Review of Nuclear and Particle Science*, 62(1):57–79.

Busha, M. T., Wechsler, R. H., Becker, M. R., Erickson, B., and Evrard, A. E. (2013). Catalog Production for the DES Blind Cosmology Challenge. In *American Astronomical Society Meeting Abstracts #221*, volume 221 of *American Astronomical Society Meeting Abstracts*, page 341.07.

Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208.

Cabella, P. and Kamionkowski, M. (2004). Theory of Cosmic Microwave Background Polarization. *arXiv e-prints*, pages astro–ph/0403392.

Cai, X., Pereyra, M., and McEwen, J. D. (2018). Uncertainty quantification for radio interferometric imaging - I. Proximal MCMC methods. *MNRAS*, 480(3):4154–4169.

Carlson, J., White, M., and Padmanabhan, N. (2009). Critical look at cosmological perturbation theory techniques. *Phys. Rev. D*, 80(4):043531.

Carroll, S. (2003). *Spacetime and Geometry: An Introduction to General Relativity*. Pearson.

Carron, J. and Szapudi, I. (2017). What does the N-point function hierarchy of the cosmological matter density field really measure? *MNRAS*, 469(3):2855–2858.

Castro, P. G., Heavens, A. F., and Kitching, T. D. (2005). Weak lensing analysis in three dimensions. *Phys. Rev. D*, 72(2):023516.

Chang, C., Pujol, A., Gaztañaga, E., Amara, A., Réfrégier, A., Bacon, D., Becker, M. R., Bonnett, C., Carretero, J., Castander, F. J., Crocce, M., Fosalba, P., Giannantonio, T., Hartley, W., Jarvis, M., Kacprzak, T., Ross, A. J., Sheldon, E., Troxel, M. A., Vikram, V., Zuntz, J., Abbott, T. M. C., Abdalla, F. B., Allam, S., Annis, J., Benoit-Lévy, A., Bertin, E., Brooks, D., Buckley-Geer, E., Burke, D. L., Capozzi, D., Rosell, A. C., Kind, M. C., Cunha, C. E., D'Andrea, C. B., da Costa, L. N., Desai, S., Diehl, H. T., Dietrich, J. P., Doel, P., Eifler, T. F., Estrada, J., Evrard, A. E., Flaugher, B., Frieman, J., Goldstein, D. A., Gruen, D., Gruendl, R. A., Gutierrez, G., Honscheid, K., Jain, B., James, D. J., Kuehn, K., Kuropatkin, N., Lahav, O., Li, T. S., Lima, M., Marshall, J. L., Martini, P., Melchior, P., Miller, C. J., Miquel, R., Mohr, J. J., Nichol, R. C., Nord, B., Ogando, R., Plazas, A. A., Reil, K., Romer, A. K., Roodman, A., Rykoff, E. S., Sanchez, E.,

Scarpine, V., Schubnell, M., Sevilla-Noarbe, I., Smith, R. C., Soares-Santos, M., Sobreira, F., Suchyta, E., Swanson, M. E. C., Tarle, G., Thomas, D., and Walker, A. R. (2016). Galaxy bias from the Dark Energy Survey Science Verification data: combining galaxy density maps and weak lensing maps. *MNRAS*, 459:3203–3216.

Chang, C., Pujol, A., Mawdsley, B., Bacon, D., Elvin-Poole, J., Melchior, P., Kovács, A., Jain, B., Leistedt, B., Giannantonio, T., Alarcon, A., Baxter, E., Bechtol, K., Becker, M. R., Benoit-Lévy, A., Bernstein, G. M., Bonnett, C., Busha, M. T., Carnero Rosell, A., Castander, F. J., Cawthon, R., da Costa, L. N., Davis, C., De Vicente, J., DeRose, J., Drlica-Wagner, A., Fosalba, P., Gatti, M., Gaztanaga, E., Gruen, D., Gschwend, J., Hartley, W. G., Hoyle, B., Huff, E. M., Jarvis, M., Jeffrey, N., Kacprzak, T., Lin, H., MacCrann, N., Maia, M. A. G., Ogando, R. L. C., Prat, J., Rau, M. M., Rollins, R. P., Roodman, A., Rozo, E., Rykoff, E. S., Samuroff, S., Sánchez, C., Sevilla-Noarbe, I., Sheldon, E., Troxel, M. A., Varga, T. N., Vielzeuf, P., Vikram, V., Wechsler, R. H., Zuntz, J., Abbott, T. M. C., Abdalla, F. B., Allam, S., Annis, J., Bertin, E., Brooks, D., Buckley-Geer, E., Burke, D. L., Carrasco Kind, M., Carretero, J., Crocce, M., Cunha, C. E., D'Andrea, C. B., Desai, S., Diehl, H. T., Dietrich, J. P., Doel, P., Estrada, J., Fausti Neto, A., Fernandez, E., Flaugher, B., Frieman, J., García-Bellido, J., Gruendl, R. A., Gutierrez, G., Honscheid, K., James, D. J., Jeltema, T., Johnson, M. W. G., Johnson, M. D., Kent, S., Kirk, D., Krause, E., Kuehn, K., Kuhlmann, S., Lahav, O., Li, T. S., Lima, M., March, M., Martini, P., Menanteau, F., Miquel, R., Mohr, J. J., Neilsen, E., Nichol, R. C., Petravick, D., Plazas, A. A., Romer, A. K., Sako, M., Sanchez, E., Scarpine, V., Schubnell, M., Smith, M., Smith, R. C., Soares-Santos, M., Sobreira, F., Suchyta, E., Tarle, G., Thomas, D., Tucker, D. L., Walker, A. R., Wester, W., and Zhang, Y. (2017). Dark Energy Survey Year 1 Results: Curved-Sky Weak Lensing Mass Map. *ArXiv e-prints*.

Chang, C., Vikram, V., Jain, B., Bacon, D., Amara, A., Becker, M. R., Bernstein, G., Bonnett, C., Bridle, S., Brout, D., Busha, M., Frieman, J., Gaztanaga, E., Hartley, W., Jarvis, M., Kacprzak, T., Kovács, A., Lahav, O., Lin, H., Melchior, P., Peiris, H., Rozo, E., Rykoff, E., Sánchez, C., Sheldon, E., Troxel, M. A., Wechsler, R., Zuntz, J., Abbott, T., Abdalla, F. B., Allam, S., Annis, J., Bauer, A. H., Benoit-Lévy, A., Brooks, D., Buckley-Geer, E., Burke, D. L., Capozzi, D., Carnero Rosell, A., Carrasco Kind, M., Castander, F. J., Crocce, M., D'Andrea, C. B., Desai, S., Diehl, H. T., Dietrich, J. P., Doel, P., Eifler, T. F., Evrard, A. E., Fausti Neto, A., Flaugher, B., Fosalba, P., Gruen, D.,

Gruendl, R. A., Gutierrez, G., Honscheid, K., James, D., Kent, S., Kuehn, K., Kuropatkin, N., Maia, M. A. G., March, M., Martini, P., Merritt, K. W., Miller, C. J., Miquel, R., Neilsen, E., Nichol, R. C., Ogando, R., Plazas, A. A., Romer, A. K., Roodman, A., Sako, M., Sanchez, E., Sevilla, I., Smith, R. C., Soares-Santos, M., Sobreira, F., Suchyta, E., Tarle, G., Thaler, J., Thomas, D., Tucker, D., and Walker, A. R. (2015). Wide-Field Lensing Mass Maps from Dark Energy Survey Science Verification Data. *Physical Review Letters*, 115(5):051301.

Chapman, E., Abdalla, F. B., Bobin, J., Starck, J.-L., Harker, G., Jelić, V., Labropoulos, P., Zaroubi, S., Brentjens, M. A., de Bruyn, A. G., and Koopmans, L. V. E. (2013). The scale of the problem: recovering images of reionization with Generalized Morphological Component Analysis. *MNRAS*, 429:165–176.

Charnock, T., Lavaux, G., and Wandelt, B. D. (2018). Automatic physical inference with information maximizing neural networks. *Phys. Rev. D*, 97(8):083004.

Cicero (1923). *De Amicitia*. Harvard University Press.

Clerkin, L., Kirk, D., Lahav, O., Abdalla, F. B., and Gaztañaga, E. (2015). A prescription for galaxy biasing evolution as a nuisance parameter. *Monthly Notices of the Royal Astronomical Society*, 448(2):1389–1401.

Clerkin, L., Kirk, D., Manera, M., Lahav, O., Abdalla, F., Amara, A., Bacon, D., Chang, C., Gaztañaga, E., Hawken, A., Jain, B., Joachimi, B., Vikram, V., Abbott, T., Allam, S., Armstrong, R., Benoit-Lévy, A., Bernstein, G. M., Bernstein, R. A., Bertin, E., Brooks, D., Burke, D. L., Rosell, A. C., Carrasco Kind, M., Crocce, M., Cunha, C. E., D'Andrea, C. B., da Costa, L. N., Desai, S., Diehl, H. T., Dietrich, J. P., Eifler, T. F., Evrard, A. E., Flaugher, B., Fosalba, P., Frieman, J., Gerdes, D. W., Gruen, D., Gruendl, R. A., Gutierrez, G., Honscheid, K., James, D. J., Kent, S., Kuehn, K., Kuropatkin, N., Lima, M., Melchior, P., Miquel, R., Nord, B., Plazas, A. A., Romer, A. K., Roodman, A., Sanchez, E., Schubnell, M., Sevilla-Noarbe, I., Smith, R. C., Soares-Santos, M., Sobreira, F., Suchyta, E., Swanson, M. E. C., Tarle, G., and Walker, A. R. (2017). Testing the lognormality of the galaxy and weak lensing convergence distributions from Dark Energy Survey maps. *MNRAS*, 466:1444–1461.

Clifton, T., Ferreira, P. G., Padilla, A., and Skordis, C. (2012). Modified gravity and cosmology. *Phys. Rep.*, 513(1):1–189.

Coe, D., Benítez, N., Sánchez, S. F., Jee, M., Bouwens, R., and Ford, H. (2006). Galaxies in the Hubble Ultra Deep Field. I. Detection, Multiband Photometry, Photometric Redshifts, and Morphology. *AJ*, 132:926–959.

Collister, A. A. and Lahav, O. (2004). ANNz: Estimating Photometric Redshifts Using Artificial Neural Networks. *PASP*, 116(818):345–351.

Connolly, A. J., Szalay, A. S., Bershady, M. A., Kinney, A. L., and Calzetti, D. (1995). Spectral Classification of Galaxies: an Orthogonal Approach. *AJ*, 110:1071.

Cooley, J. W. and Tukey, J. W. (1965). An algorithm for the machine calculation of complex fourier series. *Mathematics of computation*, 19(90):297–301.

Cooper, B., Girdlestone, S., Burovskiy, P., Gaydadjiev, G., Averbukh, V., Knowles, P. J., and Luk, W. (2017). Quantum chemistry in dataflow: Density-fitting mp2. *Journal of Chemical Theory and Computation*, 13(11):5265–5272. PMID: 29019679.

Cooray, A. and Hu, W. (2001). Weak Gravitational Lensing Bispectrum. *ApJ*, 548:7–18.

Cox, R. T. (1946). Probability, Frequency and Reasonable Expectation. *American Journal of Physics*, 14:1–13.

Dark Energy Survey Collaboration, Abbott, T., Abdalla, F. B., Aleksić, J., Allam, S., Amara, A., Bacon, D., Balbinot, E., Banerji, M., Bechtol, K., Benoit-Lévy, A., Bernstein, G. M., Bertin, E., Blazek, J., Bonnett, C., Bridle, S., Brooks, D., Brunner, R. J., Buckley-Geer, E., Burke, D. L., Caminha, G. B., Capozzi, D., Carlsen, J., Carnero-Rosell, A., Carollo, M., Carrasco-Kind, M., Carretero, J., Castander, F. J., Clerkin, L., Collett, T., Conselice, C., Crocce, M., Cunha, C. E., D'Andrea, C. B., da Costa, L. N., Davis, T. M., Desai, S., Diehl, H. T., Dietrich, J. P., Dodelson, S., Doel, P., Drlica-Wagner, A., Estrada, J., Etherington, J., Evrard, A. E., Fabbri, J., Finley, D. A., Flaugher, B., Foley, R. J., Fosalba, P., Frieman, J., García-Bellido, J., Gaztanaga, E., Gerdes, D. W., Giannantonio, T., Goldstein, D. A., Gruen, D., Gruendl, R. A., Guarnieri, P., Gutierrez, G., Hartley, W., Honscheid, K., Jain, B., James, D. J., Jeltema, T., Jouvel, S., Kessler, R., King, A., Kirk, D., Kron, R., Kuehn, K., Kuropatkin, N., Lahav, O., Li, T. S., Lima, M., Lin, H., Maia,

M. A. G., Makler, M., Manera, M., Maraston, C., Marshall, J. L., Martini, P., McMahon, R. G., Melchior, P., Merson, A., Miller, C. J., Miquel, R., Mohr, J. J., Morice-Atkinson, X., Naidoo, K., Neilsen, E., Nichol, R. C., Nord, B., Ogando, R., Ostrovski, F., Palmese, A., Papadopoulos, A., Peiris, H. V., Peoples, J., Percival, W. J., Plazas, A. A., Reed, S. L., Refregier, A., Romer, A. K., Roodman, A., Ross, A., Rozo, E., Rykoff, E. S., Sadeh, I., Sako, M., Sánchez, C., Sanchez, E., Santiago, B., Scarpine, V., Schubnell, M., Sevilla-Noarbe, I., Sheldon, E., Smith, M., Smith, R. C., Soares-Santos, M., Sobreira, F., Soumagnac, M., Suchyta, E., Sullivan, M., Swanson, M., Tarle, G., Thaler, J., Thomas, D., Thomas, R. C., Tucker, D., Vieira, J. D., Vikram, V., Walker, A. R., Wechsler, R. H., Weller, J., Wester, W., Whiteway, L., Wilcox, H., Yanny, B., Zhang, Y., and Zuntz, J. (2016). The Dark Energy Survey: more than dark energy - an overview. *MNRAS*, 460:1270–1299.

Dekel, A. and Lahav, O. (1999). Stochastic Nonlinear Galaxy Biasing. *The Astrophysical Journal*, 520(1):24–34.

DES Collaboration, Abbott, and et al. (2017a). Dark Energy Survey Year 1 Results: Cosmological Constraints from Galaxy Clustering and Weak Lensing. *ArXiv e-prints*.

DES Collaboration, Abbott, T. M. C., Abdalla, F. B., Alarcon, A., Aleksić, J., Allam, S., Allen, S., Amara, A., Annis, J., Asorey, J., Avila, S., Bacon, D., Balbinot, E., Banerji, M., Banik, N., Barkhouse, W., Baumer, M., Baxter, E., Bechtol, K., Becker, M. R., Benoit-Lévy, A., Benson, B. A., Bernstein, G. M., Bertin, E., Blazek, J., Bridle, S. L., Brooks, D., Brout, D., Buckley-Geer, E., Burke, D. L., Busha, M. T., Capozzi, D., Carnero Rosell, A., Carrasco Kind, M., Carretero, J., Castander, F. J., Cawthon, R., Chang, C., Chen, N., Childress, M., Choi, A., Conselice, C., Crittenden, R., Crocce, M., Cunha, C. E., D'Andrea, C. B., da Costa, L. N., Das, R., Davis, T. M., Davis, C., De Vicente, J., DePoy, D. L., DeRose, J., Desai, S., Diehl, H. T., Dietrich, J. P., Dodelson, S., Doel, P., Drlica-Wagner, A., Eifler, T. F., Elliott, A. E., Elsner, F., Elvin-Poole, J., Estrada, J., Evrard, A. E., Fang, Y., Fernandez, E., Ferté, A., Finley, D. A., Flaugher, B., Fosalba, P., Friedrich, O., Frieman, J., García-Bellido, J., Garcia-Fernandez, M., Gatti, M., Gaztanaga, E., Gerdes, D. W., Giannantonio, T., Gill, M. S. S., Glazebrook, K., Goldstein, D. A., Gruen, D., Gruendl, R. A., Gschwend, J., Gutierrez, G., Hamilton, S., Hartley, W. G., Hinton, S. R., Honscheid, K., Hoyle, B., Huterer, D., Jain, B., James, D. J., Jarvis,

M., Jeltema, T., Johnson, M. D., Johnson, M. W. G., Kacprzak, T., Kent, S., Kim, A. G., King, A., Kirk, D., Kokron, N., Kovacs, A., Krause, E., Krawiec, C., Kremin, A., Kuehn, K., Kuhlmann, S., Kuropatkin, N., Lacasa, F., Lahav, O., Li, T. S., Liddle, A. R., Lidman, C., Lima, M., Lin, H., MacCrann, N., Maia, M. A. G., Makler, M., Manera, M., March, M., Marshall, J. L., Martini, P., McMahon, R. G., Melchior, P., Menanteau, F., Miquel, R., Miranda, V., Mudd, D., Muir, J., Möller, A., Neilsen, E., Nichol, R. C., Nord, B., Nugent, P., Ogando, R. L. C., Palmese, A., Peacock, J., Peiris, H. V., Peoples, J., Percival, W. J., Petravick, D., Plazas, A. A., Porredon, A., Prat, J., Pujol, A., Rau, M. M., Refregier, A., Ricker, P. M., Roe, N., Rollins, R. P., Romer, A. K., Roodman, A., Rosenfeld, R., Ross, A. J., Rozo, E., Rykoff, E. S., Sako, M., Salvador, A. I., Samuroff, S., Sánchez, C., Sanchez, E., Santiago, B., Scarpine, V., Schindler, R., Scolnic, D., Secco, L. F., Serrano, S., Sevilla-Noarbe, I., Sheldon, E., Smith, R. C., Smith, M., Smith, J., Soares-Santos, M., Sobreira, F., Suchyta, E., Tarle, G., Thomas, D., Troxel, M. A., Tucker, D. L., Tucker, B. E., Uddin, S. A., Varga, T. N., Vielzeuf, P., Vikram, V., Vivas, A. K., Walker, A. R., Wang, M., Wechsler, R. H., Weller, J., Wester, W., Wolf, R. C., Yanny, B., Yuan, F., Zenteno, A., Zhang, B., Zhang, Y., and Zuntz, J. (2017b). Dark Energy Survey Year 1 Results: Cosmological Constraints from Galaxy Clustering and Weak Lensing. *ArXiv e-prints*.

DESI Collaboration, Aghamousa, A., Aguilar, J., Ahlen, S., Alam, S., Allen, L. E., Allende Prieto, C., Annis, J., Bailey, S., Balland, C., Ballester, O., Baltay, C., Beaufore, L., Bebek, C., Beers, T. C., Bell, E. F., Bernal, J. L., Besuner, R., Beutler, F., Blake, C., Bleuler, H., Blomqvist, M., Blum, R., Bolton, A. S., Briceno, C., Brooks, D., Brownstein, J. R., Buckley-Geer, E., Burden, A., Burtin, E., Busca, N. G., Cahn, R. N., Cai, Y.-C., Cardiel-Sas, L., Carlberg, R. G., Carton, P.-H., Casas, R., Castand er, F. J., Cervantes-Cota, J. L., Claybaugh, T. M., Close, M., Coker, C. T., Cole, S., Comparat, J., Cooper, A. P., Cousinou, M. C., Crocce, M., Cuby, J.-G., Cunningham, D. P., Davis, T. M., Dawson, K. S., de la Macorra, A., De Vicente, J., Delubac, T., Derwent, M., Dey, A., Dhungana, G., Ding, Z., Doel, P., Duan, Y. T., Ealet, A., Edelstein, J., Eftekharzadeh, S., Eisenstein, D. J., Elliott, A., Escoffier, S., Evatt, M., Fagrelius, P., Fan, X., Fanning, K., Farahi, A., Farihi, J., Favole, G., Feng, Y., Fernandez, E., Findlay, J. R., Finkbeiner, D. P., Fitzpatrick, M. J., Flaugher, B., Flender, S., Font-Ribera, A., Forero-Romero, J. E., Fosalba, P., Frenk, C. S., Fumagalli, M., Gaensicke, B. T., Gallo, G., Garcia-Bellido, J., Gaztanaga, E., Pietro

Gentile Fusillo, N., Gerard, T., Gershkovich, I., Giannantonio, T., Gillet, D., Gonzalez-de-Rivera, G., Gonzalez-Perez, V., Gott, S., Graur, O., Gutierrez, G., Guy, J., Habib, S., Heetderks, H., Heetderks, I., Heitmann, K., Hellwing, W. A., Herrera, D. A., Ho, S., Holland, S., Honscheid, K., Huff, E., Hutchinson, T. A., Huterer, D., Hwang, H. S., Illa Laguna, J. M., Ishikawa, Y., Jacobs, D., Jeffrey, N., Jelinsky, P., Jennings, E., Jiang, L., Jimenez, J., Johnson, J., Joyce, R., Jullo, E., Juneau, S., Kama, S., Karcher, A., Karkar, S., Kehoe, R., Kennamer, N., Kent, S., Kilbinger, M., Kim, A. G., Kirkby, D., Kisner, T., Kitanidis, E., Kneib, J.-P., Koposov, S., Kovacs, E., Koyama, K., Kremin, A., Kron, R., Kronig, L., Kueter-Young, A., Lacey, C. G., Lafever, R., Lahav, O., Lambert, A., Lampton, M., Land riau, M., Lang, D., Lauer, T. R., Le Goff, J.-M., Le Guillou, L., Le Van Suu, A., Lee, J. H., Lee, S.-J., Leitner, D., Lesser, M., Levi, M. E., L'Huillier, B., Li, B., Liang, M., Lin, H., Linder, E., Loebman, S. R., Lukić, Z., Ma, J., MacCrann, N., Magneville, C., Makarem, L., Manera, M., Manser, C. J., Marshall, R., Martini, P., Massey, R., Matheson, T., McCauley, J., McDonald, P., McGreer, I. D., Meisner, A., Metcalfe, N., Miller, T. N., Miquel, R., Moustakas, J., Myers, A., Naik, M., Newman, J. A., Nichol, R. C., Nicola, A., Nicolati da Costa, L., Nie, J., Niz, G., Norberg, P., Nord, B., Norman, D., Nugent, P., O'Brien, T., Oh, M., Olsen, K. A. G., Padilla, C., Padmanabhan, H., Padmanabhan, N., Palanque-Delabrouille, N., Palmese, A., Pappalardo, D., Pâris, I., Park, C., Patej, A., Peacock, J. A., Peiris, H. V., Peng, X., Percival, W. J., Perruchot, S., Pieri, M. M., Pogge, R., Pollack, J. E., Poppett, C., Prada, F., Prakash, A., Probst, R. G., Rabinowitz, D., Raichoor, A., Ree, C. H., Refregier, A., Regal, X., Reid, B., Reil, K., Rezaie, M., Rockosi, C. M., Roe, N., Ronayette, S., Roodman, A., Ross, A. J., Ross, N. P., Rossi, G., Rozo, E., Ruhlmann-Kleider, V., Rykoff, E. S., Sabiu, C., Samushia, L., Sanchez, E., Sanchez, J., Schlegel, D. J., Schneider, M., Schubnell, M., Secroun, A., Seljak, U., Seo, H.-J., Serrano, S., Shafieloo, A., Shan, H., Sharples, R., Sholl, M. J., Shourt, W. V., Silber, J. H., Silva, D. R., Sirk, M. M., Slosar, A., Smith, A., Smoot, G. F., Som, D., Song, Y.-S., Sprayberry, D., Staten, R., Stefanik, A., Tarle, G., Sien Tie, S., Tinker, J. L., Tojeiro, R., Valdes, F., Valenzuela, O., Valluri, M., Vargas-Magana, M., Verde, L., Walker, A. R., Wang, J., Wang, Y., Weaver, B. A., Weaverdyck, C., Wechsler, R. H., Weinberg, D. H., White, M., Yang, Q., Yeche, C., Zhang, T., Zhao, G.-B., Zheng, Y., Zhou, X., Zhou, Z., Zhu, Y., Zou, H., and Zu, Y. (2016). The DESI Experiment Part I: Science,Targeting, and Survey Design. *arXiv e-prints*, page arXiv:1611.00036.

Desjacques, V., Jeong, D., and Schmidt, F. (2018). Large-scale galaxy bias. *Physics Reports*, 733:1–193.

Dicke, R. H., Peebles, P. J. E., Roll, P. G., and Wilkinson, D. T. (1965). Cosmic Black-Body Radiation. *ApJ*, 142:414–419.

Dietrich, J. P. and Hartlap, J. (2010). Cosmology with the shear-peak statistics. *MNRAS*, 402:1049–1058.

Dodelson, S. (2003). *Modern cosmology*. Academic press.

Düben, P. D., Russell, F. P., Niu, X., Luk, W., and Palmer, T. (2015). On the use of programmable hardware and reduced numerical precision in earth-system modeling. *Journal of advances in modeling earth systems*, 7(3):1393–1408.

Duchon, J. (1976). Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces. *Revue française d'automatique, informatique, recherche opérationnelle. Analyse numérique*, 10(R3):5–12.

Dyson, F. W., Eddington, A. S., and Davidson, C. (1920). A Determination of the Deflection of Light by the Sun's Gravitational Field, from Observations Made at the Total Eclipse of May 29, 1919.

Eddington, A. S. (1917). The pulsation theory of Cepheid variables. *The Observatory*, 40:290–293.

Efstathiou, G., Davis, M., White, S. D. M., and Frenk, C. S. (1985). Numerical techniques for large cosmological N-body simulations. *ApJS*, 57:241–260.

Einstein, A. (1915a). Die Feldgleichungen der Gravitation. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), Seite 844-847.*

Einstein, A. (1915b). Erklärung der Perihelbewegung des Merkur aus der allgemeinen Relativitätstheorie. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), Seite 831-839.*

Einstein, A. (1916). Näherungsweise Integration der Feldgleichungen der Gravitation. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), Seite 688-696.*

Einstein, A. and de Sitter, W. (1932). On the Relation between the Expansion and the Mean Density of the Universe. *Proceedings of the National Academy of Science*, 18(3):213–214.

Eisenstein, D. J. and Hu, W. (1999). Power Spectra for Cold Dark Matter and Its Variants. *ApJ*, 511(1):5–15.

Ellis, G. F. R. and van Elst, H. (1999). Cosmological Models (Cargèse lectures 1998). In Lachièze-Rey, M., editor, *NATO Advanced Science Institutes (ASI) Series C*, volume 541 of *NATO Advanced Science Institutes (ASI) Series C*, pages 1–116.

Elsner, F. and Wandelt, B. D. (2013). Efficient Wiener filtering without preconditioning. *Astronomy and Astrophysics*, 549:A111.

Firth, A. E., Lahav, O., and Somerville, R. S. (2003). Estimating photometric redshifts with artificial neural networks. *MNRAS*, 339(4):1195–1202.

Fixsen, D. J., Cheng, E. S., Gales, J. M., Mather, J. C., Shafer, R. A., and Wright, E. L. (1996). The Cosmic Microwave Background Spectrum from the Full COBE FIRAS Data Set. *ApJ*, 473:576.

Flaugher, B., Diehl, H. T., Honscheid, K., Abbott, T. M. C., Alvarez, O., Angstadt, R., Annis, J. T., Antonik, M., Ballester, O., Beaufore, L., Bernstein, G. M., Bernstein, R. A., Bigelow, B., Bonati, M., Boprie, D., Brooks, D., Buckley-Geer, E. J., Campa, J., Cardiel-Sas, L., Castander, F. J., Castilla, J., Cease, H., Cela-Ruiz, J. M., Chappa, S., Chi, E., Cooper, C., da Costa, L. N., Dede, E., Derylo, G., DePoy, D. L., de Vicente, J., Doel, P., Drlica-Wagner, A., Eiting, J., Elliott, A. E., Emes, J., Estrada, J., Fausti Neto, A., Finley, D. A., Flores, R., Frieman, J., Gerdes, D., Gladders, M. D., Gregory, B., Gutierrez, G. R., Hao, J., Holland, S. E., Holm, S., Huffman, D., Jackson, C., James, D. J., Jonas, M., Karcher, A., Karliner, I., Kent, S., Kessler, R., Kozlovsky, M., Kron, R. G., Kubik, D., Kuehn, K., Kuhlmann, S., Kuk, K., Lahav, O., Lathrop, A., Lee, J., Levi, M. E., Lewis, P., Li, T. S., Mandrichenko, I., Marshall, J. L., Martinez, G., Merritt, K. W., Miquel, R., Muñoz, F., Neilsen, E. H., Nichol, R. C., Nord, B., Ogando, R., Olsen, J., Palaio, N., Patton, K., Peoples, J., Plazas, A. A., Rauch, J., Reil, K., Rheault, J.-P., Roe, N. A., Rogers, H., Roodman, A., Sanchez, E., Scarpine, V., Schindler, R. H., Schmidt, R., Schmitt, R., Schubnell, M., Schultz, K., Schurter, P., Scott, L., Serrano, S., Shaw, T. M.,

Smith, R. C., Soares-Santos, M., Stefanik, A., Stuermer, W., Suchyta, E., Sypniewski, A., Tarle, G., Thaler, J., Tighe, R., Tran, C., Tucker, D., Walker, A. R., Wang, G., Watson, M., Weaverdyck, C., Wester, W., Woods, R., Yanny, B., and DES Collaboration (2015). The Dark Energy Camera. *AJ*, 150:150.

Fluri, J., Kacprzak, T., Lucchi, A., Refregier, A., Amara, A., Hofmann, T., and Schneider, A. (2019). Cosmological constraints with deep learning from KiDS-450 weak lensing maps. *arXiv e-prints*, page arXiv:1906.03156.

Fosalba, P., Gaztañaga, E., Castander, F. J., and Crocce, M. (2015). The MICE Grand Challenge light-cone simulation - III. Galaxy lensing mocks from all-sky lensing maps. *MNRAS*, 447:1319–1332.

Frigo, M. and Johnson, S. G. (2005). The design and implementation of FFTW3. *Proceedings of the IEEE*, 93(2):216–231. Special issue on "Program Generation, Optimization, and Platform Adaptation".

Fry, J. N. (1996). The Evolution of Bias. *ApJ*, 461:L65.

Gan, L., Fu, H., Mencer, O., Luk, W., and Yang, G. (2017). Data flow computing in geoscience applications. *Advances in Computers Creativity in Computing and DataFlow SuperComputing*, page 125158.

Gaztanaga, E. (1992). N-Point Correlation Functions in the CfA and SSRS Redshift Distribution of Galaxies. *The Astrophysical Journal*, 398:L17.

Gelman, A. et al. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper). *Bayesian analysis*, 1(3):515–534.

Géron, A. (2017). *Hands-on machine learning with Scikit-Learn and TensorFlow*. O'Reilly Media.

Gibbons, G. W. and Hawking, S. W. (1977). Cosmological Event Horizons, Thermodynamics, and Particle Creation. *Phys. Rev.*, D15:2738–2751.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT. `http://www.deeplearningbook.org`.

Górski, K. M., Hivon, E., Banday, A. J., Wandelt, B. D., Hansen, F. K., Reinecke, M., and Bartelmann, M. (2005). HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. *ApJ*, 622:759–771.

Graff, P., Hobson, M. P., and Lasenby, A. (2011). An investigation into the Multiple Optimised Parameter Estimation and Data compression algorithm. *MNRAS*, 413(1):L66–L70.

Guth, A. H. (1981). Inflationary universe: A possible solution to the horizon and flatness problems. *Phys. Rev. D*, 23:347–356.

Guth, A. H. and Pi, S.-Y. (1982). Fluctuations in the new inflationary universe. *Phys. Rev. Lett.*, 49:1110–1113.

Guth, A. H. and Tye, S.-H. H. (1980). Phase transitions and magnetic monopole production in the very early universe. *Physical Review Letters*, 44:631–635.

Habib, S., Heitmann, K., Higdon, D., Nakhleh, C., and Williams, B. (2007). Cosmic calibration: Constraints from the matter power spectrum and the cosmic microwave background. *Phys. Rev. D*, 76(8):083503.

Han, Y. and Ye, J. C. (2018). Framing u-net via deep convolutional framelets: Application to sparse-view ct. *IEEE transactions on medical imaging*, 37(6):1418–1429.

Hand, N., Feng, Y., Beutler, F., Li, Y., Modi, C., Seljak, U., and Slepian, Z. (2018). nbodykit: An Open-source, Massively Parallel Toolkit for Large-scale Structure. *The Astronomical Journal*, 156(4):160.

Harrison, E. R. (1970). Fluctuations at the threshold of classical cosmology. *Phys. Rev. D*, 1:2726–2730.

Hartlap, J., Simon, P., and Schneider, P. (2007). Why your model parameter confidences might be too optimistic. Unbiased estimation of the inverse covariance matrix. *A&A*, 464:399–404.

Hawking, S. (1982). The development of irregularities in a single bubble inflationary universe. *Physics Letters B*, 115(4):295 – 297.

Heavens, A. F., Jimenez, R., and Lahav, O. (2000). Massive lossless data compression and multiple parameter estimation from galaxy spectra. *MNRAS*, 317(4):965–972.

Heitmann, K., Higdon, D., White, M., Habib, S., Williams, B. J., Lawrence, E., and Wagner, C. (2009). The Coyote Universe. II. Cosmological Models and Precision Emulation of the Nonlinear Matter Power Spectrum. *ApJ*, 705(1):156–174.

Hernquist, L., Bouchet, F. R., and Suto, Y. (1991). Application of the Ewald method to cosmological N-body simulations. *ApJS*, 75:231–240.

Hinton, S. R. (2016). ChainConsumer. *The Journal of Open Source Software*, 1:00045.

Hobson, M. P., Efstathiou, G. P., and Lasenby, A. N. (2006). *General relativity: an introduction for physicists*. Cambridge University Press.

Hobson, M. P., Jaffe, A. H., Liddle, A. R., Mukherjee, P., and Parkinson, D. (2010). *Bayesian methods in cosmology*. Cambridge University Press.

Howlett, C., Manera, M., and Percival, W. J. (2015). L-PICOLA: A parallel code for fast dark matter simulation. *A & C*, 12:109–126.

Howlett, C., Ross, A. J., Samushia, L., Percival, W. J., and Manera, M. (2015). The clustering of the sdss main galaxy sample–ii. mock galaxy catalogues and a measurement of the growth of structure from redshift space distortions at z= 0.15. *Monthly Notices of the Royal Astronomical Society*, 449(1):848–866.

Hu, W. and Dodelson, S. (2002). Cosmic Microwave Background Anisotropies. *Annual Review of Astronomy and Astrophysics*, 40:171–216.

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.

Ioffe, S. and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*.

Isserlis, L. (1918). ON A FORMULA FOR THE PRODUCT-MOMENT COEFFICIENT OF ANY ORDER OF A NORMAL FREQUENCY DISTRIBUTION IN ANY NUMBER OF VARIABLES. *Biometrika*, 12(1-2):134–139.

Jackson, J. C. (1972). A critique of Rees's theory of primordial gravitational radiation. *MNRAS*, 156:1P.

Jammalamadaka, S. R. and Sengupta, A. (2001). *Topics in circular statistics*. Series on multivariate analysis ; v. 5. World Scientific, River Edge, N.J.

Jasche, J. and Lavaux, G. (2015). Matrix-free large-scale Bayesian inference in cosmology. *MNRAS*, 447:1204–1212.

Jaynes, E. T. (2003). *Probability theory: the logic of science*. CUP.

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 186(1007):453–461.

Jeffreys, H. (1998). *The theory of probability*. OUP Oxford.

Jenkins, A., Frenk, C. S., White, S. D. M., Colberg, J. M., Cole, S., Evrard, A. E., Couchman, H. M. P., and Yoshida, N. (2001). The mass function of dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 321(2):372–384.

Jennings, W. D., Watkinson, C. A., Abdalla, F. B., and McEwen, J. D. (2019). Evaluating machine learning techniques for predicting power spectra from reionization simulations. *MNRAS*, 483(3):2907–2922.

Jones, E., Oliphant, T., Peterson, P., et al. (2001). SciPy: Open source scientific tools for Python. [Online; accessed ¡today¿].

Kacprzak, T., Kirk, D., Friedrich, O., Amara, A., Refregier, A., Marian, L., Dietrich, J. P., Suchyta, E., Aleksić, J., Bacon, D., Becker, M. R., Bonnett, C., Bridle, S. L., Chang, C., Eifler, T. F., Hartley, W. G., Huff, E. M., Krause, E., MacCrann, N., Melchior, P., Nicola, A., Samuroff, S., Sheldon, E., Troxel, M. A., Weller, J., Zuntz, J., Abbott, T. M. C., Abdalla, F. B., Armstrong, R., Benoit-Lévy, A., Bernstein, G. M., Bernstein, R. A., Bertin, E., Brooks, D., Burke, D. L., Carnero Rosell, A., Carrasco Kind, M., Carretero, J., Castander, F. J., Crocce, M., D'Andrea, C. B., da Costa, L. N., Desai, S., Diehl, H. T., Evrard, A. E., Neto, A. F., Flaugher, B., Fosalba, P., Frieman, J., Gerdes, D. W., Goldstein, D. A., Gruen, D., Gruendl, R. A., Gutierrez, G., Honscheid, K., Jain, B., James, D. J., Jarvis, M., Kuehn, K., Kuropatkin, N., Lahav, O., Lima, M., March, M., Marshall, J. L., Martini, P., Miller, C. J., Miquel, R., Mohr, J. J., Nichol, R. C., Nord, B., Plazas, A. A., Romer, A. K., Roodman, A., Rykoff, E. S., Sanchez, E., Scarpine, V., Schubnell, M., Sevilla-Noarbe,

I., Smith, R. C., Soares-Santos, M., Sobreira, F., Swanson, M. E. C., Tarle, G., Thomas, D., Vikram, V., Walker, A. R., Zhang, Y., and DES Collaboration (2016). Cosmology constraints from shear peak statistics in Dark Energy Survey Science Verification data. *MNRAS*, 463:3653–3673.

Kaiser, N. (1984). On the spatial correlations of Abell clusters. *ApJ*, 284:L9–L12.

Kaiser, N. (1987). Clustering in real space and in redshift space. *MNRAS*, 227:1–21.

Kaiser, N. (2017). Why there is no Newtonian backreaction. *MNRAS*, 469(1):744–748.

Kaiser, N. and Squires, G. (1993). Mapping the dark matter with weak gravitational lensing. *Astrophysical Journal*, 404:441–450.

Kamionkowski, M., Kosowsky, A., and Stebbins, A. (1997). A Probe of Primordial Gravity Waves and Vorticity. *Physical Review Letters*, 78(11):2058–2061.

Kerscher, M., Schmalzing, J., and Buchert, T. (1996). Analyzing Galaxy Catalogues with Minkowski Functionals. In Coles, P., Martinez, V., and Pons-Borderia, M.-J., editors, *Mapping, Measuring, and Modelling the Universe*, volume 94 of *Astronomical Society of the Pacific Conference Series*, page 247.

Kilbinger, M. (2015). Cosmology with cosmic shear observations: a review. *Reports on Progress in Physics*, 78(8):086901.

Kilbinger, M., Benabed, K., Guy, J., Astier, P., Tereno, I., Fu, L., Wraith, D., Coupon, J., Mellier, Y., and Balland, C. (2009). Dark-energy constraints and correlations with systematics from CFHTLS weak lensing, SNLS supernovae Ia and WMAP5. *A&A*, 497(3):677–688.

Kilbinger, M., Fu, L., Heymans, C., Simpson, F., Benjamin, J., Erben, T., Harnois-Déraps, J., Hoekstra, H., Hildebrandt, H., Kitching, T. D., Mellier, Y., Miller, L., Van Waerbeke, L., Benabed, K., Bonnett, C., Coupon, J., Hudson, M. J., Kuijken, K., Rowe, B., Schrabback, T., Semboloni, E., Vafaei, S., and Velander, M. (2013). CFHTLenS: combined probe cosmological model comparison using 2D weak gravitational lensing. *MNRAS*, 430:2200–2220.

Kirk, D., Brown, M. L., Hoekstra, H., Joachimi, B., Kitching, T. D., Mand elbaum, R., Sifón, C., Cacciato, M., Choi, A., and Kiessling, A. (2015). Galaxy Alignments: Observations and Impact on Cosmology. *Space Sci. Rev.*, 193(1-4):139–211.

Kofman, L. A. and Pogosyan, D. Y. (1988). Nonflat perturbations in inflationary cosmology. *Physics Letters B*, 214:508–514.

Koller, D. and Friedman, N. (2009). *Probabilistic graphical models: principles and techniques*. MIT press.

Kos, A., Rankovi, V., and Tomai, S. (2015). Sorting networks on maxeler dataflow supercomputing systems. *Advances in Computers Dataflow Processing*, page 139186.

Kovac, J. M., Leitch, E. M., Pryke, C., Carlstrom, J. E., Halverson, N. W., and Holzapfel, W. L. (2002). Detection of polarization in the cosmic microwave background using DASI. *Nature*, 420:772–787.

Krause, E. and Hirata, C. M. (2010). Weak lensing power spectra for precision cosmology. Multiple-deflection, reduced shear, and lensing bias corrections. *A&A*, 523:A28.

Lahav, O., Fisher, K. B., Hoffman, Y., Scharf, C. A., and Zaroubi, S. (1994). Wiener Reconstruction of All-Sky Galaxy Surveys in Spherical Harmonics. *Astrophysical Journal, Letters*, 423:L93.

Lahav, O. and Liddle, A. R. (2014). The Cosmological Parameters 2014. *arXiv e-prints*, page arXiv:1401.1389.

Lahav, O., Lilje, P. B., Primack, J. R., and Rees, M. J. (1991). Dynamical effects of the cosmological constant. *MNRAS*, 251:128–136.

Lahav, O., Naim, A., Buta, R. J., Corwin, H. G., de Vaucouleurs, G., Dressler, A., Huchra, J. P., van den Bergh, S., Raychaudhury, S., Sodre, L., J., and Storrie-Lombardi, M. C. (1995). Galaxies, Human Eyes, and Artificial Neural Networks. *Science*, 267(5199):859–862.

Lanusse, F., Starck, J.-L., Leonard, A., and Pires, S. (2016). High resolution weak lensing mass mapping combining shear and flexion. *A&A*, 591:A2.

Laplace, P.-S. (1840). *Essai philosophique sur les probabilités*. Bachelier.

Leavitt, H. S. (1908). 1777 variables in the Magellanic Clouds. *Annals of Harvard College Observatory*, 60:87–108.3.

Leavitt, H. S. and Pickering, E. C. (1912). Periods of 25 Variable Stars in the Small Magellanic Cloud. *Harvard College Observatory Circular*, 173:1–3.

Leclercq, F. (2018). Bayesian optimization for likelihood-free cosmological inference. *Phys. Rev. D*, 98(6):063511.

LeCun, Y., Boser, B. E., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. E., and Jackel, L. D. (1990). Handwritten digit recognition with a back-propagation network. In *Advances in neural information processing systems*, pages 396–404.

Leistedt, B., Hogg, D. W., Wechsler, R. H., and DeRose, J. (2018). Hierarchical modeling and statistical calibration for photometric redshifts. *arXiv e-prints*, page arXiv:1807.01391.

Leistedt, B., Mortlock, D. J., and Peiris, H. V. (2016). Hierarchical Bayesian inference of galaxy redshift distributions from photometric surveys. *MNRAS*, 460:4258–4267.

Lemos, P., Lee, E., Efstathiou, G., and Gratton, S. (2019). Model independent H(z) reconstruction using the cosmic inverse distance ladder. *Monthly Notices of the Royal Astronomical Society*, 483(4):4803–4810.

Leonard, A., Lanusse, F., and Starck, J.-L. (2014). GLIMPSE: accurate 3D weak lensing reconstructions using sparsity. *Monthly Notices of the Royal Astronomical Society*, 440:1281–1294.

Lesgourgues, J. (2011). The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview. *arXiv e-prints*, page arXiv:1104.2932.

Lewis, A., Challinor, A., and Lasenby, A. (2000). Efficient computation of CMB anisotropies in closed FRW models. *ApJ*, 538:473–476.

Liang, S., Yin, S., Liu, L., Luk, W., and Wei, S. (2018). Fp-bnn: Binarized neural network on fpga. *Neurocomputing*, 275:1072–1086.

Lightman, A. P. and Schechter, P. L. (1990). The Omega dependence of peculiar velocities induced by spherical density perturbations. *ApJS*, 74:831.

Lin, C.-A. and Kilbinger, M. (2015). A new model to predict weak-lensing peak counts. I. Comparison with N-body simulations. *A&A*, 576:A24.

Linde, A. (1982). A new inflationary universe scenario: A possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems. *Physics Letters B*, 108(6):389 – 393.

Lovelock, D. (1971). The Einstein Tensor and Its Generalizations. *Journal of Mathematical Physics*, 12(3):498–501.

Lovelock, D. (1972). The Four-Dimensionality of Space and the Einstein Tensor. *Journal of Mathematical Physics*, 13(6):874–876.

Loverde, M. and Afshordi, N. (2008). Extended Limber approximation. *Phys. Rev. D*, 78(12):123506.

LSST Dark Energy Science Collaboration (2012). Large Synoptic Survey Telescope: Dark Energy Science Collaboration. *ArXiv e-prints*.

LSST Science Collaboration, Abell, P. A., Allison, J., Anderson, S. F., Andrew, J. R., Angel, J. R. P., Armus, L., Arnett, D., Asztalos, S. J., Axelrod, T. S., and et al. (2009). LSST Science Book, Version 2.0. *ArXiv e-prints*.

Lueckmann, J.-M., Bassetto, G., Karaletsos, T., and Macke, J. H. (2018). Likelihood-free inference with emulator networks. *arXiv e-prints*, page arXiv:1805.09294.

Macaulay, E., Nichol, R. C., Bacon, D., Brout, D., Davis, T. M., Zhang, B., Bassett, B. A., Scolnic, D., Möller, A., D'Andrea, C. B., Hinton, S. R., Kessler, R., Kim, A. G., Lasker, J., Lidman, C., Sako, M., Smith, M., Sullivan, M., Abbott, T. M. C., Allam, S., Annis, J., Asorey, J., Avila, S., Bechtol, K., Brooks, D., Brown, P., Burke, D. L., Calcino, J., Carnero Rosell, A., Carollo, D., Carrasco Kind, M., Carretero, J., Castander, F. J., Collett, T., Crocce, M., Cunha, C. E., da Costa, L. N., Davis, C., De Vicente, J., Diehl, H. T., Doel, P., Drlica-Wagner, A., Eifler, T. F., Estrada, J., Evrard, A. E., Filippenko, A. V., Finley, D. A., Flaugher, B., Foley, R. J., Fosalba, P., Frieman, J., Galbany, L., García-Bellido, J., Gaztanaga, E., Glazebrook, K., González-Gaitán, S., Gruen, D., Gruendl, R. A., Gschwend, J., Gutierrez, G., Hartley, W. G., Hollowood, D. L., Honscheid, K., Hoormann, J. K., Hoyle, B., Huterer, D., Jain, B., James, D. J., Jeltema, T., Kasai, E.,

Krause, E., Kuehn, K., Kuropatkin, N., Lahav, O., Lewis, G. F., Li, T. S., Lima, M., Lin, H., Maia, M. A. G., Marshall, J. L., Martini, P., Miquel, R., Nugent, P., Palmese, A., Pan, Y. C., Plazas, A. A., Romer, A. K., Roodman, A., Sanchez, E., Scarpine, V., Schindler, R., Schubnell, M., Serrano, S., Sevilla-Noarbe, I., Sharp, R., Soares-Santos, M., Sobreira, F., Sommer, N. E., Suchyta, E., Swann, E., Swanson, M. E. C., Tarle, G., Thomas, D., Thomas, R. C., Tucker, B. E., Uddin, S. A., Vikram, V., Walker, A. R., Wiseman, P., and DES Collaboration (2019). First cosmological results using Type Ia supernovae from the Dark Energy Survey: measurement of the Hubble constant. *Monthly Notices of the Royal Astronomical Society*, 486(2):2184–2196.

Mallat, S. (2016). Understanding deep convolutional networks. *Philosophical Transactions of the Royal Society of London Series A*, 374(2065):20150203.

Martinet, N., Schneider, P., Hildebrandt, H., Shan, H., Asgari, M., Dietrich, J. P., Harnois-Déraps, J., Erben, T., Grado, A., Heymans, C., Hoekstra, H., Klaes, D., Kuijken, K., Merten, J., and Nakajima, R. (2017). KiDS-450: Cosmological Constraints from Weak Lensing Peak Statistics - II: Inference from Shear Peaks in N-body Simulations. *ArXiv e-prints*.

Mather, J. C., Fixsen, D. J., Shafer, R. A., Mosier, C., and Wilkinson, D. T. (1999). Calibrator Design for the COBE Far-Infrared Absolute Spectrophotometer (FIRAS). *ApJ*, 512:511–520.

Mawdsley, B., Bacon, D., Chang, C., Melchior, P., Rozo, E., Seitz, S., Jeffrey, N., Gatti, M., Gaztanaga, E., and Gruen, D. (2019). Dark Energy Survey Year 1 Results: Wide field mass maps via forward fitting in harmonic space. *arXiv e-prints*, page arXiv:1905.12682.

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.

Morales, J. L. and Nocedal, J. (2011). Remark on "algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound constrained optimization". *ACM Trans. Math. Softw.*, 38(1):7:1–7:4.

Mukhanov, V. F. and Chibisov, G. V. (1981). Quantum fluctuations and a nonsingular universe. *Soviet Journal of Experimental and Theoretical Physics Letters*, 33:532.

Mukhanov, V. F. and Chibisov, G. V. (1982). The Vacuum energy and large scale structure of the universe. *Sov. Phys. JETP*, 56:258–265. [Zh. Eksp. Teor. Fiz.83,475(1982)].

Navarro, J. F., Frenk, C. S., and White, S. D. M. (1996). The Structure of Cold Dark Matter Halos. *ApJ*, 462:563.

Newman, E. T. and Penrose, R. (1966). Note on the Bondi-Metzner-Sachs Group. *Journal of Mathematical Physics*, 7:863–870.

Noether, E. (1918). Invariante variationsprobleme. *Nachrichten von der Gesellschaft der Wissenschaften zu Gttingen, Mathematisch-Physikalische Klasse*, 1918:235–257.

Nusser, A. and Davis, M. (1994). On the prediction of velocity fields from redshift space galaxy samples. *ApJ*, 421:L1–L4.

Oguri, M., Miyazaki, S., Hikage, C., Mandelbaum, R., Utsumi, Y., Miyatake, H., Takada, M., Armstrong, R., Bosch, J., Komiyama, Y., Leauthaud, A., More, S., Nishizawa, A. J., Okabe, N., and Tanaka, M. (2017). Two- and three-dimensional wide-field weak lensing mass maps from the Hyper Suprime-Cam Subaru Strategic Program S16A data. *ArXiv e-prints*.

Papamakarios, G., Sterratt, D. C., and Murray, I. (2018). Sequential Neural Likelihood: Fast Likelihood-free Inference with Autoregressive Flows. *arXiv e-prints*, page arXiv:1805.07226.

Park, C.-G., Hwang, J.-c., and Noh, H. (2012). Axion as a cold dark matter candidate: Low-mass case. *Phys. Rev. D*, 86(8):083535.

Peacock, J. (1999). *Cosmological Physics*. Cambridge Astrophysics. Cambridge University Press.

Peebles, P. J. E. (1980). *The large-scale structure of the universe*. Princeton University Press.

Peebles, P. J. E. and Yu, J. T. (1970). Primeval Adiabatic Perturbation in an Expanding Universe. *ApJ*, 162:815.

Peel, A., Lin, C.-A., Lanusse, F., Leonard, A., Starck, J.-L., and Kilbinger, M. (2017). Cosmological constraints with weak-lensing peak counts and second-order statistics in a large-field survey. *A&A*, 599:A79.

Pell, O., Mencer, O., Tsoi, K. H., and Luk, W. (2013). *Maximum performance computing with dataflow engines*. Springer.

Penzias, A. A. and Wilson, R. W. (1965). A Measurement of Excess Antenna Temperature at 4080 Mc/s. *ApJ*, 142:419–421.

Perlmutter, S., Aldering, G., Goldhaber, G., Knop, R. A., Nugent, P., Castro, P. G., Deustua, S., Fabbro, S., Goobar, A., Groom, D. E., Hook, I. M., Kim, A. G., Kim, M. Y., Lee, J. C., Nunes, N. J., Pain, R., Pennypacker, C. R., Quimby, R., Lidman, C., Ellis, R. S., Irwin, M., McMahon, R. G., Ruiz-Lapuente, P., Walton, N., Schaefer, B., Boyle, B. J., Filippenko, A. V., Matheson, T., Fruchter, A. S., Panagia, N., Newberg, H. J. M., Couch, W. J., and Project, T. S. C. (1999). Measurements of Ω and Λ from 42 High-Redshift Supernovae. *The Astrophysical Journal*, 517(2):565–586.

Petersen, K. B. and Pedersen, M. S. (2012). The matrix cookbook. Version 20121115.

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. *Phys. Rev. D*, 88(12):123002.

Planck Collaboration, Ade, P. A. R., Aghanim, N., Arnaud, M., Ashdown, M., Aumont, J., Baccigalupi, C., Banday, A. J., Barreiro, R. B., Bartlett, J. G., Bartolo, N., Battaner, E., Battye, R., Benabed, K., Benoît, A., Benoit-Lévy, A., Bernard, J. P., Bersanelli, M., Bielewicz, P., Bock, J. J., Bonaldi, A., Bonavera, L., Bond, J. R., Borrill, J., Bouchet, F. R., Boulanger, F., Bucher, M., Burigana, C., Butler, R. C., Calabrese, E., Cardoso, J. F., Catalano, A., Challinor, A., Chamballu, A., Chary, R. R., Chiang, H. C., Chluba, J., Christensen, P. R., Church, S., Clements, D. L., Colombi, S., Colombo, L. P. L., Combet, C., Coulais, A., Crill, B. P., Curto, A., Cuttaia, F., Danese, L., Davies, R. D., Davis, R. J., de Bernardis, P., de Rosa, A., de Zotti, G., Delabrouille, J., Désert, F. X., Di Valentino, E., Dickinson, C., Diego, J. M., Dolag, K., Dole, H., Donzelli, S., Doré, O., Douspis, M., Ducout, A., Dunkley, J., Dupac, X., Efstathiou, G., Elsner, F., Enßlin, T. A., Eriksen, H. K., Farhang, M., Fergusson, J., Finelli, F., Forni, O., Frailis, M., Fraisse, A. A., Franceschi, E., Frejsel, A., Galeotta, S., Galli, S., Ganga, K., Gauthier, C., Gerbino, M., Ghosh, T., Giard, M., Giraud-Héraud, Y., Giusarma, E., Gjerløw, E., González-Nuevo, J., Górski, K. M., Gratton, S., Gregorio, A., Gruppuso, A., Gudmundsson, J. E.,

Hamann, J., Hansen, F. K., Hanson, D., Harrison, D. L., Helou, G., Henrot-Versillé, S., Hernández-Monteagudo, C., Herranz, D., Hildebrand t, S. R., Hivon, E., Hobson, M., Holmes, W. A., Hornstrup, A., Hovest, W., Huang, Z., Huffenberger, K. M., Hurier, G., Jaffe, A. H., Jaffe, T. R., Jones, W. C., Juvela, M., Keihänen, E., Keskitalo, R., Kisner, T. S., Kneissl, R., Knoche, J., Knox, L., Kunz, M., Kurki-Suonio, H., Lagache, G., Lähteenmäki, A., Lamarre, J. M., Lasenby, A., Lattanzi, M., Lawrence, C. R., Leahy, J. P., Leonardi, R., Lesgourgues, J., Levrier, F., Lewis, A., Liguori, M., Lilje, P. B., Linden-Vørnle, M., López-Caniego, M., Lubin, P. M., Macías-Pérez, J. F., Maggio, G., Maino, D., Mandolesi, N., Mangilli, A., Marchini, A., Maris, M., Martin, P. G., Martinelli, M., Martínez-González, E., Masi, S., Matarrese, S., McGehee, P., Meinhold, P. R., Melchiorri, A., Melin, J. B., Mendes, L., Mennella, A., Migliaccio, M., Millea, M., Mitra, S., Miville-Deschênes, M. A., Moneti, A., Montier, L., Morgante, G., Mortlock, D., Moss, A., Munshi, D., Murphy, J. A., Naselsky, P., Nati, F., Natoli, P., Netterfield, C. B., Nørgaard-Nielsen, H. U., Noviello, F., Novikov, D., Novikov, I., Oxborrow, C. A., Paci, F., Pagano, L., Pajot, F., Paladini, R., Paoletti, D., Partridge, B., Pasian, F., Patanchon, G., Pearson, T. J., Perdereau, O., Perotto, L., Perrotta, F., Pettorino, V., Piacentini, F., Piat, M., Pierpaoli, E., Pietrobon, D., Plaszczynski, S., Pointecouteau, E., Polenta, G., Popa, L., Pratt, G. W., Prézeau, G., Prunet, S., Puget, J. L., Rachen, J. P., Reach, W. T., Rebolo, R., Reinecke, M., Remazeilles, M., Renault, C., Renzi, A., Ristorcelli, I., Rocha, G., Rosset, C., Rossetti, M., Roudier, G., Rouillé d'Orfeuil, B., Rowan-Robinson, M., Rubiño-Martín, J. A., Rusholme, B., Said, N., Salvatelli, V., Salvati, L., Sandri, M., Santos, D., Savelainen, M., Savini, G., Scott, D., Seiffert, M. D., Serra, P., Shellard, E. P. S., Spencer, L. D., Spinelli, M., Stolyarov, V., Stompor, R., Sudiwala, R., Sunyaev, R., Sutton, D., Suur-Uski, A. S., Sygnet, J. F., Tauber, J. A., Terenzi, L., Toffolatti, L., Tomasi, M., Tristram, M., Trombetti, T., Tucci, M., Tuovinen, J., Türler, M., Umana, G., Valenziano, L., Valiviita, J., Van Tent, F., Vielva, P., Villa, F., Wade, L. A., Wandelt, B. D., Wehus, I. K., White, M., White, S. D. M., Wilkinson, A., Yvon, D., Zacchei, A., and Zonca, A. (2016). Planck 2015 results. XIII. Cosmological parameters. *A&A*, 594:A13.

Planck Collaboration, Aghanim, N., Akrami, Y., Ashdown, M., Aumont, J., Baccigalupi, C., Ballardini, M., Banday, A. J., Barreiro, R. B., and Bartolo, N. (2018a). Planck 2018 results. VI. Cosmological parameters. *arXiv e-prints*, page arXiv:1807.06209.

Planck Collaboration, Akrami, Y., Arroja, F., Ashdown, M., Aumont, J., Baccigalupi, C.,

Ballardini, M., Banday, A. J., Barreiro, R. B., Bartolo, N., Basak, S., Battye, R., Benabed, K., Bernard, J. P., Bersanelli, M., Bielewicz, P., Bock, J. J., Bond, J. R., Borrill, J., Bouchet, F. R., Boulanger, F., Bucher, M., Burigana, C., Butler, R. C., Calabrese, E., Cardoso, J. F., Carron, J., Casaponsa, B., Challinor, A., Chiang, H. C., Colombo, L. P. L., Combet, C., Contreras, D., Crill, B. P., Cuttaia, F., de Bernardis, P., de Zotti, G., Delabrouille, J., Delouis, J. M., Désert, F. X., Di Valentino, E., Dickinson, C., Diego, J. M., Donzelli, S., Doré, O., Douspis, M., Ducout, A., Dupac, X., Efstathiou, G., Elsner, F., Enßlin, T. A., Eriksen, H. K., Falgarone, E., Fantaye, Y., Fergusson, J., Fernandez-Cobos, R., Finelli, F., Forastieri, F., Frailis, M., Franceschi, E., Frolov, A., Galeotta, S., Galli, S., Ganga, K., Génova-Santos, R. T., Gerbino, M., Ghosh, T., González-Nuevo, J., Górski, K. M., Gratton, S., Gruppuso, A., Gudmundsson, J. E., Hamann, J., Hand ley, W., Hansen, F. K., Helou, G., Herranz, D., Hivon, E., Huang, Z., Jaffe, A. H., Jones, W. C., Karakci, A., Keihänen, E., Keskitalo, R., Kiiveri, K., Kim, J., Kisner, T. S., Knox, L., Krachmalnicoff, N., Kunz, M., Kurki-Suonio, H., Lagache, G., Lamarre, J. M., Langer, M., Lasenby, A., Lattanzi, M., Lawrence, C. R., Le Jeune, M., Leahy, J. P., Lesgourgues, J., Levrier, F., Lewis, A., Liguori, M., Lilje, P. B., Lilley, M., Lindholm, V., López-Caniego, M., Lubin, P. M., Ma, Y. Z., Macías-Pérez, J. F., Maggio, G., Maino, D., Mand olesi, N., Mangilli, A., Marcos-Caballero, A., Maris, M., Martin, P. G., Martínez-González, E., Matarrese, S., Mauri, N., McEwen, J. D., Meerburg, P. D., Meinhold, P. R., Melchiorri, A., Mennella, A., Migliaccio, M., Millea, M., Mitra, S., Miville-Deschênes, M. A., Molinari, D., Moneti, A., Montier, L., Morgante, G., Moss, A., Mottet, S., Münchmeyer, M., Natoli, P., Nørgaard-Nielsen, H. U., Oxborrow, C. A., Pagano, L., Paoletti, D., Partridge, B., Patanchon, G., Pearson, T. J., Peel, M., Peiris, H. V., Perrotta, F., Pettorino, V., Piacentini, F., Polastri, L., Polenta, G., Puget, J. L., Rachen, J. P., Reinecke, M., Remazeilles, M., Renzi, A., Rocha, G., Rosset, C., Roudier, G., Rubiño-Martín, J. A., Ruiz-Granados, B., Salvati, L., Sandri, M., Savelainen, M., Scott, D., Shellard, E. P. S., Shiraishi, M., Sirignano, C., Sirri, G., Spencer, L. D., Sunyaev, R., Suur-Uski, A. S., Tauber, J. A., Tavagnacco, D., Tenti, M., Terenzi, L., Toffolatti, L., Tomasi, M., Trombetti, T., Valiviita, J., Van Tent, B., Vibert, L., Vielva, P., Villa, F., Vittorio, N., Wandelt, B. D., Wehus, I. K., White, M., White, S. D. M., Zacchei, A., and Zonca, A. (2018b). Planck 2018 results. I. Overview and the cosmological legacy of Planck. *arXiv e-prints*, page arXiv:1807.06205.

Pontzen, A., Roškar, R., Stinson, G. S., Woods, R., Reed, D. M., Coles, J., and Quinn, T. R. (2013). pynbody: Astrophysics Simulation Analysis for Python. Astrophysics Source Code Library, ascl:1305.002.

Press, W. H. and Schechter, P. (1974). Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. *ApJ*, 187:425–438.

Price, M. A., Cai, X., McEwen, J. D., Pereyra, M., and Kitching, T. D. (2018). Sparse Bayesian mass-mapping with uncertainties: local credible intervals. *arXiv e-prints*, page arXiv:1812.04017.

Rees, M. J. (1997). *Before the Beginning: Our Universe and Others*. SIMON & SCHUSTER LTD.

Riess, A. G., Casertano, S., Yuan, W., Macri, L. M., and Scolnic, D. (2019). Large magellanic cloud cepheid standards provide a 1% foundation for the determination of the hubble constant and stronger evidence for physics beyond *lambda*CDM. *The Astrophysical Journal*, 876(1):85.

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., Leibundgut, B., Phillips, M. M., Reiss, D., Schmidt, B. P., Schommer, R. A., Smith, R. C., Spyromilio, J., Stubbs, C., Suntzeff, N. B., and Tonry, J. (1998). Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. *The Astronomical Journal*, 116(3):1009–1038.

Rollins, R. P. (2015). *Chemical and statistical models of the interstellar medium and star-forming regions*. PhD thesis, University College London.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597.

Rykoff, E. S., Rozo, E., Hollowood, D., Bermeo-Hernandez, A., Jeltema, T., Mayers, J., Romer, A. K., Rooney, P., Saro, A., Vergara Cervantes, C., Wechsler, R. H., Wilcox, H., Abbott, T. M. C., Abdalla, F. B., Allam, S., Annis, J., Benoit-Lévy, A., Bernstein, G. M., Bertin, E., Brooks, D., Burke, D. L., Capozzi, D., Carnero Rosell, A., Carrasco Kind, M., Castander, F. J., Childress, M., Collins, C. A., Cunha, C. E., D'Andrea, C. B., da Costa,

L. N., Davis, T. M., Desai, S., Diehl, H. T., Dietrich, J. P., Doel, P., Evrard, A. E., Finley, D. A., Flaugher, B., Fosalba, P., Frieman, J., Glazebrook, K., Goldstein, D. A., Gruen, D., Gruendl, R. A., Gutierrez, G., Hilton, M., Honscheid, K., Hoyle, B., James, D. J., Kay, S. T., Kuehn, K., Kuropatkin, N., Lahav, O., Lewis, G. F., Lidman, C., Lima, M., Maia, M. A. G., Mann, R. G., Marshall, J. L., Martini, P., Melchior, P., Miller, C. J., Miquel, R., Mohr, J. J., Nichol, R. C., Nord, B., Ogando, R., Plazas, A. A., Reil, K., Sahlén, M., Sanchez, E., Santiago, B., Scarpine, V., Schubnell, M., Sevilla-Noarbe, I., Smith, R. C., Soares-Santos, M., Sobreira, F., Stott, J. P., Suchyta, E., Swanson, M. E. C., Tarle, G., Thomas, D., Tucker, D., Uddin, S., Viana, P. T. P., Vikram, V., Walker, A. R., Zhang, Y., and DES Collaboration (2016). The RedMaPPer Galaxy Cluster Catalog From DES Science Verification Data. *ApJS*, 224:1.

Sachs, R. K. and Wolfe, A. M. (1967). Perturbations of a Cosmological Model and Angular Variations of the Microwave Background. *ApJ*, 147:73.

Salopek, D. S., Bond, J. R., and Bardeen, J. M. (1989). Designing density fluctuation spectra in inflation. *Phys. Rev. D*, 40:1753–1788.

Schneider, M. D., Ng, K. Y., Dawson, W. A., Marshall, P. J., Meyers, J. E., and Bard, D. J. (2017). Probabilistic Cosmological Mass Mapping from Weak Lensing Shear. *ApJ*, 839:25.

Schneider, P. (1985). A new formulation of gravitational lens theory, time-delay, and Fermat's principle. *A&A*, 143:413–420.

Schneider, P. and Seitz, C. (1995a). Steps towards nonlinear cluster inversion through gravitational distortions. 1: Basic considerations and circular clusters. *A&A*, 294:411–431.

Schneider, P. and Seitz, C. (1995b). Steps towards nonlinear cluster inversion through gravitational distortions. I. Basic considerations and circular clusters. *A&A*, 294:411–431.

Scoccimarro, R. (1998). Transients from initial conditions: a perturbative analysis. *Monthly Notices of the Royal Astronomical Society*, 299(4):1097–1118.

Scolnic, D. M., Jones, D. O., Rest, A., Pan, Y. C., Chornock, R., Foley, R. J., Huber, M. E., Kessler, R., Narayan, G., Riess, A. G., Rodney, S., Berger, E., Brout, D. J., Challis,

P. J., Drout, M., Finkbeiner, D., Lunnan, R., Kirshner, R. P., Sanders, N. E., Schlafly, E., Smartt, S., Stubbs, C. W., Tonry, J., Wood-Vasey, W. M., Foley, M., Hand, J., Johnson, E., Burgett, W. S., Chambers, K. C., Draper, P. W., Hodapp, K. W., Kaiser, N., Kudritzki, R. P., Magnier, E. A., Metcalfe, N., Bresolin, F., Gall, E., Kotak, R., McCrum, M., and Smith, K. W. (2018). The Complete Light-curve Sample of Spectroscopically Confirmed SNe Ia from Pan-STARRS1 and Cosmological Constraints from the Combined Pantheon Sample. *ApJ*, 859:101.

Seitz, C. and Schneider, P. (1995). Steps towards nonlinear cluster inversion through gravitational distortions II. Generalization of the Kaiser and Squires method. *A&A*, 297:287.

Seitz, C. and Schneider, P. (1997). Steps towards nonlinear cluster inversion through gravitational distortions. III. Including a redshift distribution of the sources. *A&A*, 318:687–699.

Seitz, S. and Schneider, P. (2001). A new finite-field mass reconstruction algorithm. *A&A*, 374:740–745.

Seljak, U. and Zaldarriaga, M. (1996). A line of sight approach to cosmic microwave background anisotropies. *ApJ*, 469:437–444.

Seljak, U. and Zaldarriaga, M. (1997). Signature of gravity waves in polarization of the microwave background. *Phys. Rev. Lett.*, 78:2054–2057.

Sellentin, E. and Heavens, A. F. (2016). Parameter inference with estimated covariance matrices. *MNRAS*, 456:L132–L136.

Shan, H., Liu, X., Hildebrandt, H., Pan, C., Martinet, N., Fan, Z., Schneider, P., Asgari, M., Harnois-Déraps, J., Hoekstra, H., Wright, A., Dietrich, J. P., Erben, T., Getman, F., Grado, A., Heymans, C., Klaes, D., Kuijken, K., Merten, J., Puddu, E., Radovich, M., and Wang, Q. (2017). KiDS-450: Cosmological Constraints from Weak Lensing Peak Statistics-I: Inference from Analytical Prediction of high Signal-to-Noise Ratio Convergence Peaks. *ArXiv e-prints*.

Shandarin, S. and Zeldovich, Y. B. (1984). Topological mapping properties of collisionless potential and vortex motion. *Physical review letters*, 52(17):1488.

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.

Shirasaki, M., Yoshida, N., and Ikeda, S. (2018). Denoising Weak Lensing Mass Maps with Deep Learning. *arXiv e-prints*, page arXiv:1812.05781.

Simon, P., Taylor, A. N., and Hartlap, J. (2009). Unfolding the matter distribution using three-dimensional weak gravitational lensing. *MNRAS*, 399:48–68.

Sivia, D. and Skilling, J. (2006). *Data Analysis: A Bayesian Tutorial*. Oxford University Press.

Skilling, J. (2004). Nested Sampling. In Fischer, R., Preuss, R., and Toussaint, U. V., editors, *American Institute of Physics Conference Series*, volume 735 of *American Institute of Physics Conference Series*, pages 395–405.

Springel, V. (2005). The cosmological simulation code GADGET-2. *MNRAS*, 364(4):1105–1134.

Starck, J.-L., Fadili, J., and Murtagh, F. (2007). The undecimated wavelet decomposition and its reconstruction. *IEEE Transactions on Image Processing*, 16(2):297–309.

Starck, J.-L., Murtagh, F., and Fadili, J. (2015). *Sparse Image and Signal Processing: Wavelets and Related Geometric Multiscale Analysis*. Cambridge University Press, New York, NY, USA, 2nd edition.

Starobinsky, A. (1982). Dynamics of phase transition in the new inflationary universe scenario and generation of perturbations. *Physics Letters B*, 117(3):175 – 178.

Sunyaev, R. A. and Zeldovich, I. B. (1980). Microwave background radiation as a probe of the contemporary structure and history of the universe. *ARA&A*, 18:537–560.

Tanabashi, M., Hagiwara, K., Hikasa, K., Nakamura, K., Sumino, Y., Takahashi, F., Tanaka, J., Agashe, K., Aielli, G., Amsler, C., Antonelli, M., Asner, D. M., Baer, H., Banerjee, S., Barnett, R. M., Basaglia, T., Bauer, C. W., Beatty, J. J., Belousov, V. I., Beringer, J., Bethke, S., Bettini, A., Bichsel, H., Biebel, O., Black, K. M., Blucher, E., Buchmuller, O., Burkert, V., Bychkov, M. A., Cahn, R. N., Carena, M., Ceccucci, A., Cerri, A., Chakraborty, D., Chen, M. C., Chivukula, R. S., Cowan, G., Dahl, O., D'Ambrosio,

G., Damour, T., de Florian, D., de Gouvêa, A., DeGrand, T., de Jong, P., Dissertori, G., Dobrescu, B. A., D'Onofrio, M., Doser, M., Drees, M., Dreiner, H. K., Dwyer, D. A., Eerola, P., Eidelman, S., Ellis, J., Erler, J., Ezhela, V. V., Fetscher, W., Fields, B. D., Firestone, R., Foster, B., Freitas, A., Gallagher, H., Garren, L., Gerber, H. J., Gerbier, G., Gershon, T., Gershtein, Y., Gherghetta, T., Godizov, A. A., Goodman, M., Grab, C., Gritsan, A. V., Grojean, C., Groom, D. E., Grünewald, M., Gurtu, A., Gutsche, T., Haber, H. E., Hanhart, C., Hashimoto, S., Hayato, Y., Hayes, K. G., Hebecker, A., Heinemeyer, S., Heltsley, B., Hernández-Rey, J. J., Hisano, J., Höcker, A., Holder, J., Holtkamp, A., Hyodo, T., Irwin, K. D., Johnson, K. F., Kado, M., Karliner, M., Katz, U. F., Klein, S. R., Klempt, E., Kowalewski, R. V., Krauss, F., Kreps, M., Krusche, B., Kuyanov, Y. V., Kwon, Y., Lahav, O., Laiho, J., Lesgourgues, J., Liddle, A., Ligeti, Z., Lin, C. J., Lippmann, C., Liss, T. M., Littenberg, L., Lugovsky, K. S., Lugovsky, S. B., Lusiani, A., Makida, Y., Maltoni, F., Mannel, T., Manohar, A. V., Marciano, W. J., Martin, A. D., Masoni, A., Matthews, J., Meißner, U. G., Milstead, D., Mitchell, R. E., Mönig, K., Molaro, P., Moortgat, F., Moskovic, M., Murayama, H., Narain, M., Nason, P., Navas, S., Neubert, M., Nevski, P., Nir, Y., Olive, K. A., Pagan Griso, S., Parsons, J., Patrignani, C., Peacock, J. A., Pennington, M., Petcov, S. T., Petrov, V. A., Pianori, E., Piepke, A., Pomarol, A., Quadt, A., Rademacker, J., Raffelt, G., Ratcliff, B. N., Richardson, P., Ringwald, A., Roesler, S., Rolli, S., Romaniouk, A., Rosenberg, L. J., Rosner, J. L., Rybka, G., Ryutin, R. A., Sachrajda, C. T., Sakai, Y., Salam, G. P., Sarkar, S., Sauli, F., Schneider, O., Scholberg, K., Schwartz, A. J., Scott, D., Sharma, V., Sharpe, S. R., Shutt, T., Silari, M., Sjöstrand , T., Skands, P., Skwarnicki, T., Smith, J. G., Smoot, G. F., Spanier, S., Spieler, H., Spiering, C., Stahl, A., Stone, S. L., Sumiyoshi, T., Syphers, M. J., Terashi, K., Terning, J., Thoma, U., Thorne, R. S., Tiator, L., Titov, M., Tkachenko, N. P., Törnqvist, N. A., Tovey, D. R., Valencia, G., Van de Water, R., Varelas, N., Venanzoni, G., Verde, L., Vincter, M. G., Vogel, P., Vogt, A., Wakely, S. P., Walkowiak, W., Walter, C. W., Wands, D., Ward, D. R., Wascko, M. O., Weiglein, G., Weinberg, D. H., Weinberg, E. J., White, M., Wiencke, L. R., Willocq, S., Wohl, C. G., Womersley, J., Woody, C. L., Workman, R. L., Yao, W. M., Zeller, G. P., Zenin, O. V., Zhu, R. Y., Zhu, S. L., Zimmermann, F., Zyla, P. A., Anderson, J., Fuller, L., Lugovsky, V. S., Schaffner, P., and Particle Data Group (2018). Review of Particle Physics*. *Phys. Rev. D*, 98(3):030001.

Tassev, S., Zaldarriaga, M., and Eisenstein, D. J. (2013). Solving large scale structure in ten

easy steps with COLA. *J. Cosmology Astropart. Phys.*, 2013(6):036.

Tegmark, M. and Peebles, P. J. E. (1998). The Time Evolution of Bias. *The Astrophysical Journal*, 500(2):L79–L82.

Tegmark, M., Taylor, A. N., and Heavens, A. F. (1997). Karhunen-Loève Eigenvalue Problems in Cosmology: How Should We Tackle Large Data Sets? *ApJ*, 480(1):22–35.

Tibshirani, R. (1994). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, 58:267–288.

Trotta, R. (2008). Bayes in the sky: Bayesian inference and model selection in cosmology. *Contemporary Physics*, 49(2):71–104.

Unruh, W. G. (1976). Notes on black-hole evaporation. *Phys. Rev. D*, 14:870–892.

van Uitert, E., Joachimi, B., Joudaki, S., Heymans, C., Köhlinger, F., Asgari, M., Blake, C., Choi, A., Erben, T., Farrow, D. J., Harnois-Déraps, J., Hildebrandt, H., Hoekstra, H., Kitching, T. D., Klaes, D., Kuijken, K., Merten, J., Miller, L., Nakajima, R., Schneider, P., Valentijn, E., and Viola, M. (2017). KiDS+GAMA: Cosmology constraints from a joint analysis of cosmic shear, galaxy-galaxy lensing and angular clustering. *ArXiv e-prints*.

Van Waerbeke, L., Benjamin, J., Erben, T., Heymans, C., Hildebrandt, H., Hoekstra, H., Kitching, T. D., Mellier, Y., Miller, L., Coupon, J., Harnois-Déraps, J., Fu, L., Hudson, M., Kilbinger, M., Kuijken, K., Rowe, B., Schrabback, T., Semboloni, E., Vafaei, S., van Uitert, E., and Velander, M. (2013). CFHTLenS: mapping the large-scale structure with gravitational lensing. *MNRAS*, 433:3373–3388.

Verde, L. and Heavens, A. F. (2001). On the trispectrum as a gaussian test for cosmology. *The Astrophysical Journal*, 553(1):14–24.

Vikram, V., Chang, C., Jain, B., Bacon, D., Amara, A., Becker, M. R., Bernstein, G., Bonnett, C., Bridle, S., Brout, D., Busha, M., Frieman, J., Gaztanaga, E., Hartley, W., Jarvis, M., Kacprzak, T., Kovács, A., Lahav, O., Leistedt, B., Lin, H., Melchior, P., Peiris, H., Rozo, E., Rykoff, E., Sánchez, C., Sheldon, E., Troxel, M. A., Wechsler, R., Zuntz, J., Abbott, T., Abdalla, F. B., Armstrong, R., Banerji, M., Bauer, A. H., Benoit-Lévy, A., Bertin, E., Brooks, D., Buckley-Geer, E., Burke, D. L., Capozzi, D., Carnero

Rosell, A., Carrasco Kind, M., Castander, F. J., Crocce, M., Cunha, C. E., D'Andrea, C. B., da Costa, L. N., DePoy, D. L., Desai, S., Diehl, H. T., Dietrich, J. P., Estrada, J., Evrard, A. E., Fausti Neto, A., Fernandez, E., Flaugher, B., Fosalba, P., Gerdes, D., Gruen, D., Gruendl, R. A., Honscheid, K., James, D., Kent, S., Kuehn, K., Kuropatkin, N., Li, T. S., Maia, M. A. G., Makler, M., March, M., Marshall, J., Martini, P., Merritt, K. W., Miller, C. J., Miquel, R., Neilsen, E., Nichol, R. C., Nord, B., Ogando, R., Plazas, A. A., Romer, A. K., Roodman, A., Sanchez, E., Scarpine, V., Sevilla, I., Smith, R. C., Soares-Santos, M., Sobreira, F., Suchyta, E., Swanson, M. E. C., Tarle, G., Thaler, J., Thomas, D., Walker, A. R., and Weller, J. (2015). Wide-field lensing mass maps from Dark Energy Survey science verification data: Methodology and detailed analysis. *Phys. Rev. D*, 92(2):022006.

Visser, M. (2004). Jerk, snap and the cosmological equation of state. *Classical and Quantum Gravity*, 21:2603–2615.

Wallis, C. G. R., McEwen, J. D., Kitching, T. D., Leistedt, B., and Plouviez, A. (2017). Mapping dark matter on the celestial sphere with weak gravitational lensing. *ArXiv e-prints*.

Weinberg, S. (1989). The cosmological constant problem. *Reviews of Modern Physics*, 61:1–23.

Weinberg, S. (2008). *Cosmology*. Oxford University Press.

Weir, N., Fayyad, U. M., Djorgovski, S. G., and Roden, J. (1995). The SKICAT System for Processing and Analyzing Digital Imaging Sky Surveys. *PASP*, 107:1243.

Whelan, J. and Iben, Jr., I. (1973). Binaries and Supernovae of Type I. *ApJ*, 186:1007–1014.

Wick, G. C. (1950). The evaluation of the collision matrix. *Phys. Rev.*, 80:268–272.

Wiener, N. (1949). *Extrapolation, interpolation, and smoothing of stationary time series*, volume 7. MIT press Cambridge, MA.

Ye, J. C., Han, Y., and Cha, E. (2018). Deep convolutional framelets: A general deep learning framework for inverse problems. *SIAM Journal on Imaging Sciences*, 11(2):991–1048.

Zaldarriaga, M. (2001). Nature of the E-B decomposition of CMB polarization. *Physical Review D*, 64(10):103001.

Zaroubi, S., Hoffman, Y., Fisher, K. B., and Lahav, O. (1995). Wiener Reconstruction of the Large-Scale Structure. *ApJ*, 449:446.

Zeiler, M. D. and Fergus, R. (2014). Visualizing and understanding convolutional networks. In *European conference on computer vision*, pages 818–833. Springer.

Zel'dovich, Y. B. (1970). Gravitational instability: An approximate theory for large density perturbations. *A&A*, 5:84–89.

Zeldovich, Y. B. (1972). A hypothesis, unifying the structure and the entropy of the Universe. *MNRAS*, 160:1P.

Zhu, C., Byrd, R. H., Lu, P., and Nocedal, J. (1997). Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization. *ACM Trans. Math. Softw.*, 23(4):550–560.

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.