

Title

A Thesis Presented

by

Evan Saraivanov

to

The Graduate School

in Partial Fulfillment of the Requirements

for the Degree of

Master of Arts

in

Physics

Stony Brook University

Month 2023

Stony Brook University

The Graduate School

Evan Saraivanov

We, the thesis committee for the above candidate for the Master of Arts degree,
hereby recommend acceptance of this thesis.

Vivian Miranda – Thesis Advisor

Assistant Professor, Department of Physics and Astronomy, Yang Institute for
Theoretical Physics

Person – Chairperson of Defense
title,affil

Navid Vafaei-Najafabadi

Assistant Professor, Department of Physics and Astronomy

Name of Outside Member
Position
Institution

This thesis is accepted by the Graduate School.

The Dean
Dean of the Graduate School

Abstract of the Thesis

Title

by

Evan Saraivanov

Master of Arts

in

Physics

Stony Brook University

2023

Abstract.

Contents

List of Figures	vii
Acknowledgements	ix
1 General Relativity and Cosmology	1
1.1 Motivation	1
1.2 Einstein's Field Equation	3
1.3 Cosmology	5
2 Measurement of Weak Lensing and the Cosmic Microwave Background	10
2.1 Weak Lensing	10
2.1.1 The Linear Matter Power Spectrum	11
2.1.2 Galaxy Density	12
2.1.3 Galaxy Shear	13
2.1.4 Weak Lensing Correlation Functions	16
2.1.5 Systematics	17
2.2 CMB	20
2.2.1 TT Power Spectrum	21
2.2.2 TE and EE Power Spectra	22
2.3 Tensions Between Weak Lensing and CMB	22
3 Computing Techniques	27
3.1 Markov Chain Monte Carlo	27
3.1.1 The Metropolis-Hastings Algorithm	28
3.2 Tension Metrics	29
3.2.1 Parameter-based Metrics	29
3.2.2 Evidence-Based Metrics	31
3.2.3 DES v Planck Results	32

3.3	Normalizing Flows	33
4	Growth-Geometry Split in DES-Y3	37
4.1	Split Matter Power Spectrum	37
4.2	Data and Analysis Pipeline	39
4.2.1	DES Parameter Priors	39
4.2.2	Software Pipeline	41
4.2.3	Pipeline Validation on Synthetic Data	42
4.3	Results	43
4.3.1	LCDM	43
4.3.2	wCDM	45
4.3.3	Tension Analysis	45
4.4	Conclusion	46
5	Data Emulators	50
5.1	Neural Networks	50
5.2	Architecture Choices	53
5.2.1	Multi-Layer Perceptron	54
5.2.2	Residual Network	54
5.2.3	Bottlenecked Residual Network	56
5.2.4	Attention	56
5.3	Results	57
5.3.1	Training and Testing	58
5.3.2	Application to Tension Calibration	59
	Bibliography	65
	Bibliography	65

List of Figures

1.1	Observer freely falling with particles experience tidal forces within the observers reference frame.	2
1.2	Definition of proper time (bold line) vs. coordinate time (dashed line).	3
1.3	Example of parallel transport along a connection on a sphere.	4
1.4	An image of galaxies from the James Webb Space Telescope, showing regions of high matter density and regions of low matter density [1].	6
1.5	CMB TT power spectrum, emphasizing the scale of anisotropies [2]. .	7
2.1	Image of weak lensing taken by the Hubble Space Telescope [3].	10
2.2	Planck TT power spectrum	21
2.3	Planck TE power spectrum	23
2.4	Planck EE power spectrum	23
2.5	Measurements of H_0 from different experiments.	25
2.6	Measurements of S_8 from different experiments.	26
3.1	Result of tension metrics between Planck and DES (figure 6 in [4]). A-priori tension is the gaussian approximation described in equation 3.20.	33
3.2	An example non-Gaussian distribution (non-zero skewness). .	35
3.3	An example of normalizing flow as the transformation is ‘turned on’, represented by the parameter λ . The non-Gaussian distribution at $\lambda = 0$ is transformed into the Gaussian distribution at $\lambda = 1$	36
4.1	Non-linear power spectrum from Euclid emulator. When either the geometry or growth parameters are varied, the other is kept fixed at $\Omega^X = 0.3$	38
4.2	Effect of changing growth and geometry parameters on $\sigma_8(z)$.	39

4.3	Relative difference between Euclid emulator and Halofit.	42
4.4	Cosmic shear, 2x2pt combinations, and 3x2pt posteriors for DES-Y3 with the ALL prior.	43
4.5	Comparison of DES-Y1 and DES-Y3 posteriors.	44
4.6	Split posteriors for DES-Y3 3x2pt and ALL prior.	45
4.7	DES-Y1 and DES-Y3 posteriors for 2-point correlation functions.	46
4.8	Full 1d results for Λ CDM split	47
4.9	w CDM posteriors.	48
4.10	Full 1d results for Λ CDM split	48
4.11	49
5.1	Arrangement of neurons.	51
5.2	Neurons with synapses.	51
5.3	Neurons, synapses, and IO.	52
5.4	A multi-layer perceptron.	55
5.5	A residual block.	55
5.6	A bottlenecked residual block.	56
5.7	A transformer block	60
5.8	Loss of a ResNet model with 3 ResNet blocks and 256 neurons in each layer.	61
5.9	Testing loss on a $T = 8$ chain.	62
5.11	Average χ^2 as a function of the number of parameters for each architecture.	64

Acknowledgements

Acknowledgements

Chapter 1

General Relativity and Cosmology

1.1 Motivation

With Einstein's formulation of special relativity came a beautiful expansion of physics into the high energy and large scale regime. A postulate of special relativity was that light is a universal speed limit, and with this postulate a plethora of established physical theories became problematic. Fortunately for theories such as electromagnetism no incompatibilities were introduced. However, for gravity this was not the case. The gravitational field had no wave equation to establish a speed of propagation, meaning changes to the gravitational field were felt instantaneously globally. Attempts to reconcile this, such as using a retarded propagator or promoting the newtonian potential to a scalar field theory, proved incapable of correcting for special relativistic effects. Never-the-less, Newtonian gravity has stood the test of time to be accurate in the non-relativistic limit. Thus any theory of gravity should reduce to the Newtonian theory for speeds $v \ll c$.

The famous thought experiment of an Einstein elevator demonstrates a striking inconsistency within Newtonian gravity. Consider an observer freely falling in a spherical gravitational field. With nothing to look at, this person would not be able to determine they were in a gravitational field rather than just linearly accelerating (this is the equivalence principle). Now suppose this observer has point particles nearby, one to each side, one above, and one below (fig 1.1). What the observer will see is the particles below and above them will move

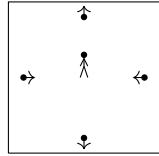


Figure 1.1: Observer freely falling with particles experience tidal forces within the observers reference frame.

farther away because the gravitational force scales as $1/r^2$. They will see the particles to the side of them moving towards them due to the spherical symmetry of the gravitational field. These fictitious forces are referred to as the tidal forces, as they also give rise to the seen twice per day on earth. This interesting thought experiment shows how freely falling reference frames are not inertial frames, the first inconsistency recognized by Einstein.

In the early 1900s, Einstein published his first paper on special relativity. In special relativity, physical space is enhanced to a four dimensional space, typically denoted \mathbb{M}^4 or $\mathbb{R}^{3,1}$. Setting $c = 1$ (which is done in the rest of this thesis), the metric is given by

$$g = \text{diag}(-1, 1, 1, 1) \quad (1.1)$$

This space is typically called *Minkowski space* (thus the notation \mathbb{M}^4). Distances under this metric are called the *proper time* interval, denoted τ .

$$-d\tau^2 = -dt^2 + dx^2 + dy^2 + dz^2 \quad (1.2)$$

Proper time represents the amount of time it that passes for an observer travelling on a path with 0 spatial displacement. Naturally, this quantity should be invariant even for an observer in a different reference frame. For such an observer, the time that elapses will be different from τ , and will be accounted for by the spatial displacement as seen by equation 1.2. The set of transformations that leave τ invariant are the Lorentz transformations.

Lets suppose we have two observers, one is in an inertial frame with coordinates x and the other is in a non-inertial frame with coordinates x' . These coordinates are related $x = x(x')$. In the inertial frame, the equations of motion are given by

$$\frac{d^2x^\mu}{d\tau^2} = 0 \quad (1.3)$$

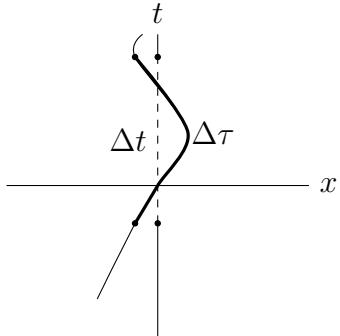


Figure 1.2: Definition of proper time (bold line) vs. coordinate time (dashed line).

We can look at how this changes for the observer in a non-inertial frame.

$$\frac{d^2x'^\lambda}{d\tau^2} + \underbrace{\frac{\partial x'^\lambda}{\partial x^\rho} \frac{\partial^2 x^\rho}{\partial x'^\mu \partial x'^\nu}}_{\equiv \Gamma_{\mu\nu}^\lambda} \frac{dx'^\mu}{d\tau} \frac{dx'^\nu}{d\tau} = 0 \quad (1.4)$$

Equation 1.4 looks suspiciously like the geodesic equation from differential geometry. The exception is we have defined the ‘Christoffel symbol’ here, which is not the most general form of the Christoffel symbols. This does give motivation for how to search for a compatible theory of gravity by providing a clear link between gravity and geometry.

1.2 Einstein’s Field Equation

Christoffel symbols take a very general form, where they signify the failure of derivatives to commute. This failure results from curvature of the space itself. As said in the previous section, special relativity promotes the classical \mathbb{R}^3 space to \mathbb{M}^4 . However, we now need to allow curved space, not simply minkowski space. So lets take an arbitrary semi-Riemannian 4-manifold M . In very loose terms, a manifold is a locally euclidean space, which is a sufficient definition for this application.

Naturally, computing derivatives on curved surfaces is non-trivial as there is no way to compare tangent vectors at two different points of a manifold. One can, however, use the fact that derivatives give tangent vectors and look at a single point $p \in M$ to construct the tangent space, denoted $T_p M$, which is the set of all tangent vectors at p . The generalization of the derivative is called a

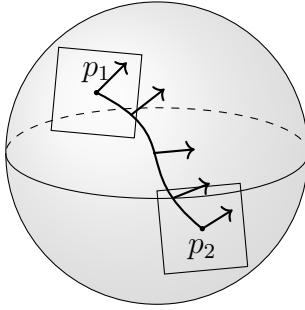


Figure 1.3: Example of parallel transport along a connection on a sphere.

connection [5], which tells one how to connect the tangent spaces at different points infinitessimally (figure 1.3). Although beyond the scope of this thesis, it is worth noting that, due to the holonomy of curved surfaces, this will generally depend on the path taken. In general relativity, a very special connection is used called the *Levi-Civita connection*, denoted ∇ , which is torsionless and metric preserving. That is, for all vector fields u, v, w on M

$$\begin{aligned} ug(v, w) &= g(\nabla_u v, w) + g(v, \nabla_u w), \quad (\text{metric preserving}) \\ [u, v] &= \nabla_u v - \nabla_v u, \quad (\text{torsion free}) \end{aligned} \tag{1.5}$$

The only remaining geometric property for the Levi-Civita connection is curvature, which can be neatly described using the christoffel symbols defined by

$$\nabla_\mu \partial_\nu = \Gamma_{\mu\nu}^\rho \partial_\rho. \tag{1.6}$$

In this case, the partials here are referring to the basis of vector fields on M . Additionally this curvature has a special name as well, the *Riemann curvature*, a type (3,1) tensor defined by

$$R(u, v)w = (\nabla_u \nabla_v - \nabla_v \nabla_u - \nabla_{[u,v]})w \tag{1.7}$$

or in coordinate form by

$$R_{\mu\nu\rho}^\sigma = \nabla_\mu \Gamma_{\nu\rho}^\sigma - \nabla_\nu \Gamma_{\mu\rho}^\sigma \tag{1.8}$$

To emphasize the elegance of Einstein's equation, it is convenient to write the Riemann curvature as a matrix valued differential 2-form \mathcal{R} and the christoffel symbols as a matrix valued differential 1-form \mathcal{G} [5] (differential forms are antisymmetric covariant tensors). With this notion, its easy to see the Bianchi

identity for \mathcal{R} with respect to the Levi-Civita connection is

$$d_{\nabla} \mathcal{R} = \nabla_{[\tau} R^{\sigma}_{\mu\nu]\rho} = 0 \quad (1.9)$$

Where the brackets denote antisymmetric permutations of the indices. Working through the algebra gives

$$\nabla^{\mu} G_{\mu\nu} = 0 \quad G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R \quad (1.10)$$

Where R and $R_{\mu\nu}$ are the Ricci scalar and Ricci tensor found by contracting indices in the Riemann tensor. This looks familiar to the conservation of energy equation

$$\nabla^{\mu} T_{\mu\nu} = 0 \quad (1.11)$$

Additionally, the metric itself is divergenceless. Putting these all together gives us Einstein's Field Equation

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi\kappa T_{\mu\nu} \quad (1.12)$$

1.3 Cosmology

We can see in Einstein's equation the constant Λ , called the cosmological constant. A great question in gravitational physics and cosmology is ‘what value does Λ take, and what is its source’. The theory of Λ CDM, the standard model of cosmology, will be discussed, where dark matter is ‘cold’ (so it clumps to form halos) and dark energy acts as the cosmological constant [6].

When discussing Λ CDM (or cosmology in general), there are two main assumptions that are made on a global scale:

- The universe is homogeneous. That is, there is no preferred location in the universe.
- The universe is isotropic. That is, there is no preferred direction of the universe.

Both of these assumptions should sound strange; Our universe is clearly does not follow these assumptions. If we look into the night sky, we can see regions of densely packed stars and galaxies and other regions with no stars or galaxies, violating the assumption of homogeneity (figure 1.4). On the other hand, if one observes the temperature of the cosmic microwave background

(CMB), one finds there is an angle dependence of the observed temperature (figure 1.5), violating the assumption of isotropy. To reconcile this, the metric will be modified by perturbations. In this thesis, only scalar perturbations are considered.



Figure 1.4: An image of galaxies from the James Webb Space Telescope, showing regions of high matter density and regions of low matter density [1].

With Edwin Hubble's discovery of the expanding universe, any metric assigned to the universe must include homogeneous spatial expansion. Together with the above assumptions results in the FLRW metric.

$$g = \text{diag}(-1, a^2, a^2, a^2) \quad (1.13)$$

To perturb this metric around a homogenous universe, it requires two fields. The first, denoted Ψ , is the newtonian gravitational potential. Because newtonian gravity works in the non-relativistic limit, it enters the time component of the metric. The second field, denoted Φ , represents spatial curvature perturbations. The perturbed metric is given by

$$g = \text{diag}(-1 - 2\Psi, a^2(1 - 2\Phi), a^2(1 - 2\Phi), a^2(1 - 2\Phi)) \quad (1.14)$$

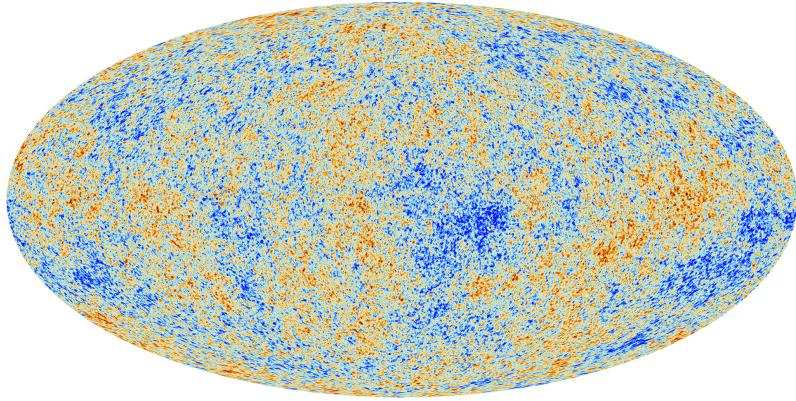


Figure 1.5: CMB TT power spectrum, emphasizing the scale of anisotropies [2].

By computing the Christoffel symbols, one finds

$$\begin{aligned}\Gamma_{00}^i &= \frac{1}{a^2} \partial_i \Psi \\ \Gamma_{0j}^i &= \delta_{ij} (H + \dot{\Phi}) \\ \Gamma_{jk}^i &= (\delta_{ij} \partial_k + \delta_{ik} \partial_j - \delta_{jk} \partial_i) \Phi\end{aligned}\tag{1.15}$$

Given Einstein's field equations and the perturbed FLRW metric, one can find the Friedman equations. The first comes from the 00 component of Einstein's field equations.

$$H^2(t) = \frac{8\pi G}{3} \left(\rho + \frac{\Lambda}{8\pi G} \right) - \frac{k}{a^2}\tag{1.16}$$

What can be done with this is to convert the curvature k and the cosmological constant Λ into densities as well, denoted ρ_k and ρ_Λ respectively. If those terms are absorbed into ρ , one finds

$$H^2(t) = \frac{8\pi G}{3} \rho\tag{1.17}$$

and the critical density is defined by the density associated with H_0 ,

$$H_0^2 \equiv \frac{8\pi G}{3} \rho_{\text{crit}}\tag{1.18}$$

The desirable property of this formulation is that the value of the density today can tell us about the geometry of the universe. If $\rho > \rho_{\text{crit}}$ the curvature is positive (de Sitter geometry). If $\rho < \rho_{\text{crit}}$ the curvature is negative (anti-de Sitter geometry). If $\rho = \rho_{\text{crit}}$, the universe is flat. If we define the relative

densities by

$$\Omega_i = \frac{\rho_i}{\rho_{\text{crit}}} \quad (1.19)$$

The relation becomes

$$\Omega + \Omega_k = 1 \quad (1.20)$$

Meaning if the non-curvature relative densities are less than 1, then the universe is curved with sign equal to the sign of Ω_k . This relates to the way we determine distances in the universe ???. The corresponding distance is called the *angular distance*, which relates the angle subtended by two rays to the distance between them. By figure ??, this depends on the geometry of the universe.

$$D_A = \begin{cases} R & K = 0 \\ R \sin(D/R) & K > 0 \\ R \sinh(D/R) & K < 0 \end{cases} . \quad (1.21)$$

The second Friedman equation comes from the trace of Einstein's equation.

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P) \quad (1.22)$$

In cosmology, there are other distances which can be more useful than the distance given by the FLRW metric. In the FLRW metric the distance between two points grows in time. We can avoid this by defining the *comoving distance* in which distances remain fixed through time. If we look at a coordinate function x^μ at $t = t_0$, at a later time the coordinate function can be written as $x^\mu \rightarrow a(t)x^\mu$, thus by dividing by the scale factor $a(t)$ we can define the comoving coordinates as

$$\chi = \int_{t_0}^t \frac{1}{a(t')} dt' \quad (1.23)$$

with the standard Minkowski metric. This can be taken a step further by determining how far light has travelled since $t = 0$

$$\eta = \int_0^t \frac{1}{a(t')} dt' \quad (1.24)$$

Since we can't see anything beyond this distance, it is often called the *comoving horizon*.

This chapter provides a brief overview of the relevant pieces of cosmology that will be used throughout the rest of this thesis. In the next two chapters,

we will use these pieces to see how we can measure cosmological parameters from cosmological surveys. The fourth chapter will be a study on a parameter splitting technique, which provides a model independent consistency check in Λ CDM. The fifth chapter provides a study of applying neural networks to cosmological analysis, and a scientific application to the evaluation of cosmological tensions.

Chapter 2

Measurement of Weak Lensing and the Cosmic Microwave Background

2.1 Weak Lensing

In general, contemporary weak lensing surveys measure the light emitted from galaxies in two components: galaxy position and galaxy shear. In the following sections the theoretical base for cosmological weak lensing surveys are described. However, before discussing the two relevant observables, one needs to formalize the concept of weak lensing.



Figure 2.1: Image of weak lensing taken by the Hubble Space Telescope [3].

2.1.1 The Linear Matter Power Spectrum

As we have seen before, we consider deviations from the FLRW metric using the scalar fields Φ and Ψ . In the absence of anisotropic stress, $\Phi = -\Psi$, which we will assume here for demonstration (one can use codes such as CMB or CLASS to do computations including anisotropy if desired).

Throughout time in the universe, we consider there to be three distinct epochs/regions [6]:

- The super-horizon region, or the region before inflation where radiation dominates
- The horizon-crossing region, or the region during inflation where the universe transfers from radiation to matter domination
- The sub-horizon region, or the region of after rapid inflation where matter dominates

Additionally, the formation of matter inhomogeneity comes from gravitational instability, the idea that primordial curvature perturbations in the super-horizon grew to form small matter density contrasts which grew in size after crossing the horizon. After the rate of inflation slowed, gravitational forces caused clusters of matter to become the galaxies and clusters we see today.

This gives us a blueprint for writing the matter power spectrum. We have the primordial curvature perturbation \mathcal{R} which source the initial matter inhomogeneity that formed late-time large scale structure. Thus, we can write it as

$$\mathcal{R} = \frac{5}{3}\Phi_{\text{large-scale}}(k, a_{\text{late}}) \quad (2.1)$$

Additionally, we have a transfer function which describes the transition from radiation to matter domination, defined as

$$T(k) = \frac{\Phi(k, a_{\text{late}})}{\Phi_{\text{large-scale}}(k, a_{\text{late}})} \quad (2.2)$$

And we have a growth factor, defined as the linear growth of matter density contrast during matter domination, defined as

$$D_+(a) = a \frac{\Phi(k, a)}{\Phi(k, a_{\text{late}})} \quad (2.3)$$

Thus we can write the gravitational potential at all times as

$$\Phi(k, a) = \frac{3}{5a} \mathcal{R}(k) T(k, a) D_+(a) \quad (2.4)$$

This together with the Poisson equation for the matter density contrast field gives the equation

$$\delta_m(k) = \frac{2k^2 a}{3\Omega_m H_0^2} \Phi(k, a) \quad (2.5)$$

$$\delta_m(k) = \frac{2}{5} \frac{k^2}{\Omega_m H_0^2} \mathcal{R}(k) T(k, a) D_+(a) \quad (2.6)$$

The power spectrum of an observable is defined as the two-point correlation function of the observable in Fourier space.

$$P_{\mathcal{O}}(k) = \langle \mathcal{O}(k) \mathcal{O}(k) \rangle \quad (2.7)$$

Thus the linear matter power spectrum (because we restricted the matter density contrast field to linear order terms in the perturbation) is given by

$$P_m^L(k, a) = \frac{8\pi^2}{25\Omega_m^2 H_0^4} A_s D_+^2(a) T^2(k) \frac{k^{n_s}}{k_p^{(n_s-1)}} \quad (2.8)$$

2.1.2 Galaxy Density

In practice, we cannot measure the matter power spectrum. Dark matter prohibits direct measurement, thus one must determine observables which trace the matter distribution in the universe. The most natural option is to observe the galaxy distribution. However, large scale structures and galaxies provide many sources for non-linearities in the power spectrum and result in a non-trivial relation between the two. By taking certain data selection criteria, one can cut out the non-linear contributions. The relation between the galaxy and matter power spectra are then related by the linear bias factor.

$$\begin{aligned} \delta_g(k) &= b_1 \delta_m(k) \\ P_g(k, a) &= (b_1)^2 P_m^L(k, a) \end{aligned} \quad (2.9)$$

Realistically, in galaxy surveys we do not have access to exact redshifts for each observed galaxy. Rather, an image is taken of the sky in multiple color bands from which the redshift is determined. This procedure is called photometric redshift. To account for this, one can attempt to redefine the galaxy power

spectrum. First, we define the distribution of distances as

$$W(\chi) = \frac{1}{N_g} \frac{dN_g}{d\chi} \quad (2.10)$$

The procedure for determining this distribution will be discussed in section (section for dz and IA). This allows us to simply write the projected density contrast as

$$\Delta_g = \int_0^\infty W(\chi) \delta_g(x, \tau) d\chi \quad (2.11)$$

To complete this discussion, we will make some approximations. We will only consider small scales for which $\sin(\theta) \sim \theta \sim 1/l$ where l is the Fourier conjugate of θ . Second is the limber approximation, which states that, as long as the galaxy power spectrum P_g slowly varies over the region $\Delta k \sim 1/l\chi$, we approximate it as a constant. These approximations greatly simplify the integrals required when computing power spectrum of Fourier transformed Δ_g . This power spectrum is called the angular power spectrum. Taking into account the approximations above and computing the fourier transform, one finds an angular power spectrum of

$$C_g(l) = \int \frac{1}{\chi^2} W^2(\chi) P_g(k = (l + 1/2)/\chi, \eta) d\chi \quad (2.12)$$

2.1.3 Galaxy Shear

Photons move on null geodesics, thus $d\chi = -d\eta$, where η is the conformal time and χ is the conformal distance. This tells us that $dx^i/d\chi = -dx^i/d\eta = -adx^i/dt = -adx^i/d\lambda d\lambda/dt = -a/P^0 P^\mu = -a/E * p/a\hat{p}^i = -\hat{p}^i$ (for photons and other massless particles). Thus, under lensing, the difference between the observed and true positions within the source plane is given by

$$\chi\theta^i = x_\perp^i = - \int_0^\chi \hat{p}_\perp^i(\chi'') d\chi'' \quad (2.13)$$

Our goal is to write this as a function of the potential only. Thus we look to compute $d\hat{p}^i/dt$.

$$\frac{d\hat{p}^i}{dt} = \frac{1}{p^2} (\dot{p}^i p - \dot{p} p^i) \quad (2.14)$$

By computing \dot{p}^i , we can find \dot{p} as well.

$$\frac{dp^i}{d\lambda} = P^i \frac{d}{d\lambda}(a(1 + \Phi) + (1 + \Phi)a \frac{d}{d\lambda}P^i) \quad (2.15)$$

We have $d/d\lambda = P^\mu \partial_\mu$, so the first term becomes

$$P^i \frac{d}{d\lambda}(a(1 + \Phi)) = aP^i(P^0(H + \dot{\Phi} + H\Phi) + P^j \partial_j \Phi) \quad (2.16)$$

The second term is evaluated by using the geodesic equation up to first order in the perturbations.

$$\begin{aligned} \frac{d}{d\lambda}P^i &= -\Gamma_{\mu\nu}^i P^\mu P^\nu \\ &= -E \left(\frac{E}{a^2} \partial^i \Psi + \frac{1}{a} 2p^i (1 - \Phi - \Psi)(H + \dot{\Phi}) - \frac{2}{a^2 E} p^i p^j \partial_j \Phi + \frac{p^2}{a^2 E} \partial^i \Phi \right) \end{aligned} \quad (2.17)$$

Plugging everything back in, and only keeping terms to linear order, we get

$$\frac{dp^i}{dt} = -p^i(H + \dot{\Phi}) - \frac{1}{aE} (p^i p^j \partial_j \Phi - p^2 \partial^i \Phi) - \frac{E}{a} \partial^i \Psi \quad (2.18)$$

The time derivative of the modulus of the momentum, p , is then given by

$$\begin{aligned} \frac{dp}{dt} &= \frac{d}{dt} \sqrt{\delta_{ij} p^i p^j} = \frac{1}{p} \delta_{ij} \dot{p}^i p^j \\ &= -p(H + \dot{\Phi}) - \hat{p}_i \frac{E}{a} \partial^i \Psi \end{aligned} \quad (2.19)$$

Now, putting everything together, plugging these results into eq () gives

$$\frac{d\hat{p}^i}{dt} = \frac{E}{ap} \left(\frac{p^2}{E^2} \partial_j \Phi - \partial_j \Psi \right) (\delta^{ij} - \hat{p}^i \hat{p}^j) \quad (2.20)$$

Going back to the photon scenario, where $E = p$, this simplifies to

$$\frac{d\hat{p}^i}{dt} = \frac{1}{a} (\partial_j \Phi - \partial_j \Psi) (\delta^{ij} - \hat{p}^i \hat{p}^j) \quad (2.21)$$

Interestingly, $\delta^{ij} - \hat{p}^i \hat{p}^j$, is the projection operator of the j -th direction onto the i -th. So, if we orient our axes so that the photon approximately travels along

the j -th direction, we can sum of i and k so that the orthogonal components of the momentum are evaluated by

$$\frac{d\hat{p}_\perp^i}{d\chi} = -a \frac{d\hat{p}_\perp^i}{dt} = -\partial_i(\Phi - \Psi) \quad (2.22)$$

In the absence of anisotropic stress (like in the late universe where weak lensing is measured), $\Phi = -\Psi$ and

$$\frac{d\hat{p}_\perp^i}{d\chi} = 2\partial_i\Phi \quad (2.23)$$

Integrating this equation gives us

$$\hat{p}_\perp(\chi'') = -2 \int_0^{\chi''} \partial_i\Phi(\chi') d\chi' + C_i \quad (2.24)$$

where the constant C_i comes from integrating a derivative, where it is only defined up to a constant. Now, plugging back in we have

$$\theta^i = \frac{2}{\chi} \int_0^\chi \int_0^{\chi''} \partial_i\Phi(\chi') d\chi' d\chi'' - C_i \quad (2.25)$$

In the absence of lensing, this should reduce to $\theta^i = \theta_0^i$, the true source position. Hence $C^i = \theta_0^i$ and the total deflection angle is given by

$$\begin{aligned} \Delta\theta^i &= \frac{2}{\chi} \int_0^\chi \int_0^{\chi''} \partial_i\Phi(\chi') d\chi' d\chi'' \\ &= \frac{2}{\chi} \int_0^\chi \partial_i\Phi(\chi') (\chi - \chi') d\chi' \end{aligned} \quad (2.26)$$

Finally, since $x^i = \chi\theta^i$, we have $\partial_i = \partial_{\theta^i}/\chi$. Now we introduce the *lensing potential* defined by

$$\begin{aligned} \Delta\theta^i &= \frac{\partial}{\partial\theta^i} \psi_L(\theta) \\ \psi_L(\theta) &= 2 \int_0^\chi \frac{1}{\chi'} \Phi(\chi') \left(1 - \frac{\chi'}{\chi}\right) d\chi' \end{aligned} \quad (2.27)$$

This result is great, but it's not very experimentally illuminating. Similar to the galaxy vs matter power spectrum, the absence of the observation of dark matter means we cannot reliably determine the potential Φ . Again, an attempt to relate the lensing potential to some property of galaxies can give us a measureable quantity. Lensing distorts the shape of the galaxies (see figure 2.1), where we assume all galaxies are elliptical. Define the galaxy shape tensor by

$$q_{ij} \equiv \frac{1}{F} \int I \theta^i \theta^j d\theta^i d\theta^j \quad (2.28)$$

This integral is symmetric under $i \leftrightarrow j$. Writing this as a matrix gives

$$q_{ij} = \frac{1}{2} \text{tr}(q) \begin{pmatrix} 1 + \epsilon_1 & \epsilon_2 \\ \epsilon_2 & 1 - \epsilon_1 \end{pmatrix} \quad (2.29)$$

Shape distortion specifically occurs when the deflection angle is not constant across a galaxy. Thus we can look at the transformation tensor as

$$A_{ij} \equiv \frac{\partial \theta_S^i}{\partial \theta^j} = \begin{pmatrix} 1 + \kappa - \gamma_+ & -\gamma_\times \\ -\gamma_\times & 1 + \kappa + \gamma_+ \end{pmatrix} = \delta_{ij} + \psi_{ij} \quad (2.30)$$

Within this tensor, the γ_+ and γ_- represent shear and κ represents magnification and changes in galaxy light flux (F). The E -mode is given by

$$E(l) = \left(\frac{l^i l^j}{l^2} - \frac{1}{2} \delta^{ij} \right) (-\psi^{ij}(l)) \quad (2.31)$$

Which relates to the shear components by

$$\begin{pmatrix} \gamma_+ \\ \gamma_\times \end{pmatrix} = \begin{pmatrix} \cos(2\alpha_l) \\ \sin(2\alpha_l) \end{pmatrix} E(l) \quad (2.32)$$

This connects the observable shear to the theoretical lensing potential.

2.1.4 Weak Lensing Correlation Functions

In the previous two sections, we have derived two observables directly connected to cosmological theories. By observing the galaxies we can determine Ω_b , A_s , n_s and H_0 . By observing shear we can determine Ω_m . The details for how to determine these parameters will be discuss in chapter (chapter). Together these two can fully determine the values of the 5 Λ CDM parameters. We are still not ready, since these two quantities themselves are not observables. Using these two quantities, however, one can construct three different

two-point correlation functions which are observable.

The first correlation function is the autocorrelation of galaxy density, called galaxy clustering. Given a density in a particular angular position, the correlation function can be interpreted as the likelihood of observing a similar density at another close position. The galaxy clustering correlation function is given by

$$w_g(\theta) = \mathcal{F}(C_g(l)) = \frac{1}{2\pi} \int_0^\infty l C_g(l) J_0(l\theta) dl \quad (2.33)$$

Where J_0 is the 0-th Bessel function and \mathcal{F} denotes the fourier transform.

Next is the shear autocorrelation. In this case, there are two components of shear: the tangential shear γ_+ and the cross shear γ_\times . The cross correlation $\langle \gamma_+ \gamma_\times \rangle$ vanishes, so we are left with combinations

$$\langle \gamma_+ \gamma_+ \rangle \pm \langle \gamma_\times \gamma_\times \rangle = \xi_\pm \quad (2.34)$$

$$\xi_{+-} = \frac{1}{2\pi} \int_0^\infty l C_{EE}(l) J_{0,4}(l\theta) C_{EE}(l) dl \quad (2.35)$$

Finally, we have the cross correlation between galaxy density and tangential shear. First, the angular power spectrum $C_{gE}(l)$ is given by

$$C_{gE}(l) = \frac{3}{2} \Omega_m H_0^2 \int_0^\infty \frac{1}{\chi a(\chi)} W_g(\chi) P_g(k, \eta) d\chi \quad (2.36)$$

$$\gamma = \langle \Delta_g \gamma_+ \rangle = -\frac{1}{2\pi} \int_0^\infty l J_2(l\theta) C_{gE}(l) dl \quad (2.37)$$

The correlation function $\langle \Delta_g \gamma_\times \rangle$ vanishes in an isotropic universe.

2.1.5 Systematics

Up to now, we have neglected any systematics of the weak lensing observables. There are two distinct types of systematics that will be discussed in this section: The ones which affect the calculation of the correlation functions, and ones which affect the correlation functions after they have been computed. The latter type are usually referred to as ‘fast’ parameters, and we will see why

in the next chapter. In the meantime here is an explanation of the relevant systematics.

Photometric Redshift Uncertainty

As we discussed in the galaxy density section, to perform the computation one must determine the redshift distribution $W(\chi)$. This is a highly non-trivial task. To model the uncertainty in this distribution, one applies a constant shift to the redshift of observed galaxies.

$$n_g(z) \mapsto n_g(z + \Delta z) \quad (2.38)$$

The shift Δz must be determined before one can compute the galaxy clustering and galaxy lensing correlation functions.

Intrinsic Alignment of Galaxies

As can be seen in figure (fig), galaxy shapes are not completely random. In the absence of lensing, there is still some non-zero correlation between the shapes. This is referred to as Intrinsic Alignment. There are two models of intrinsic alignment that will be considered in this thesis: the Tidal Alignment Tidal Torquing (TATT) model [7, 8] and the Non-linear Alignment (NLA) model. The TATT depends on the tidal tensor s_{ij} which contains the information about tidal forces due to non-uniform gravitational fields. Tidal forces are a source of intrinsic alignment and shear (figure). The model is given by

$$\gamma_{ij}^I = \underbrace{C_1 s_{ij}}_{\text{tidal alignment}} + \underbrace{b_{TA} C_1 (\delta_m \times s_{ij})}_{\text{density weighting}} + \underbrace{C_2 \left(s_i^k s_{kj} - \frac{1}{3} \delta_{ij} s^2 \right)}_{\text{tidal torquing}} \quad (2.39)$$

Where the coefficients C_a are given by

$$\begin{aligned} C_1 &= -\frac{A_1 \bar{C} \Omega_m}{a G(z)} \left(\frac{1+z}{1+z_0} \right)^{\eta_1} \\ C_2 &= 5 \frac{A_2 \bar{C} \Omega_m}{(a G(z))^2} \left(\frac{1+z}{1+z_0} \right)^{\eta_2} \end{aligned} \quad (2.40)$$

The TATT model reduces to the NLA model when $A_2 = 0 = b_{TA}$

Galaxy Bias

As discussed in the galaxy clustering section, we relate the galaxy and matter density by a bias parameter b . However we can pick a fiducial value to relate the two

$$\delta_{g,\text{fid}} = b_{\text{fid}} \delta_m \quad (2.41)$$

then consider a multiplicative deviation

$$\delta_g = (b \cdot b_{\text{fid}}) \delta_m = b \delta_{g,\text{fid}} \quad (2.42)$$

Thus, the parameter b can be found to account for different galaxy biases.

Galaxy bias is the first of many ‘fast’ parameters. Since the bias is constant, it can be pulled out of the integral for the correlation functions, meaning it can be applied after the correlation function has been computed. The affected correlation functions are the galaxy clustering and galaxy-galaxy lensing.

$$\begin{aligned} w^i(\theta) &= b_{(i)}^2 w^i(\theta) \\ \gamma_t^i(\theta) &= b_{(i)} \gamma_t^i(\theta) \end{aligned} \quad (2.43)$$

where i here denotes the redshift bin

Shear Calibration

Observed shear is not a completely astrophysical phenomenon. For example, a major source of additional shear comes from the observation lens itself. The exact methods for how this is handled (such as the point spread function) are beyond the scope of this thesis. This is usually modelled as a multiplicative correction to the shear components.

$$\begin{aligned} \xi_{\pm}^{ij} &= (1 + m^i)(1 + m^j) \xi_{\pm}^{ij} \\ \gamma_t^{ij} &= (1 + m^j) \gamma_t^{ij} \end{aligned} \quad (2.44)$$

where i and j denote the source and lens bin respectively. The multiplicative shear calibration is also a fast parameter.

Point Mass

Currently, we are unable to model the contribution of small scales to the tangential shear [9]. The tangential shear can be written in terms of the

surface mass density Σ in the tangential direction

$$\gamma_t(R/D_A) = \frac{\bar{\Sigma}(0, R) - \Sigma(R)}{\Sigma_{\text{crit}}} = \frac{\Delta\Sigma(R)}{\Sigma_{\text{crit}}} \quad (2.45)$$

We can only model the tangential shear to r_{\min} , which allows us to write the point mass contribution (named for the R^{-2} dependence) as

$$\Delta\Sigma(R) = \Delta\Sigma^{\text{model}}(R) + \frac{B}{R^2} \quad (2.46)$$

Lensing Magnification

As discussed in (section), the intrinsic shape of galaxies also gets magnified. This effect is introduced as a parameter C_l^i and is fixed experimentally.

2.2 CMB

Although CMB data is not used very much in this thesis, it is worth discussing the measurements qualitatively. They consensus among physicists and astronomers that the universe beyond the horizon was a hot, dense plasma. As such, the mean free path of photons were much shorter than they are today, thus compton scattering dominated photon dynamics in the early universe. As the universe cooled, atoms began to form. in a process called recombination. During this time, the universe was still opaque as photons scattered on the newly formed atoms. As the universe began to expand, the mean free path of photons increased and inter-atomic interactions decreased, meaning the atoms could be ionized by absorbing free photons in a process called reionization. The absorption of the photons caused the universe to become transparent. One can compute the optical depth of this transition from opaque to transparent as

$$T = T_0(1 - e^{-\tau_{\text{rei}}}) \quad (2.47)$$

τ_{rei} concludes the introduction of cosmological parameters. τ_{rei} can only be determined from CMB and is usually fixed in weak lensing analysis.

The CMB is a black body, meaning the intensity of the light can be used to determine the temperature according to the black body equation where $I \propto T^4$. The temperature, along with the E -mode polarization, give three possible two point correlation functions: TT , TE , and EE .

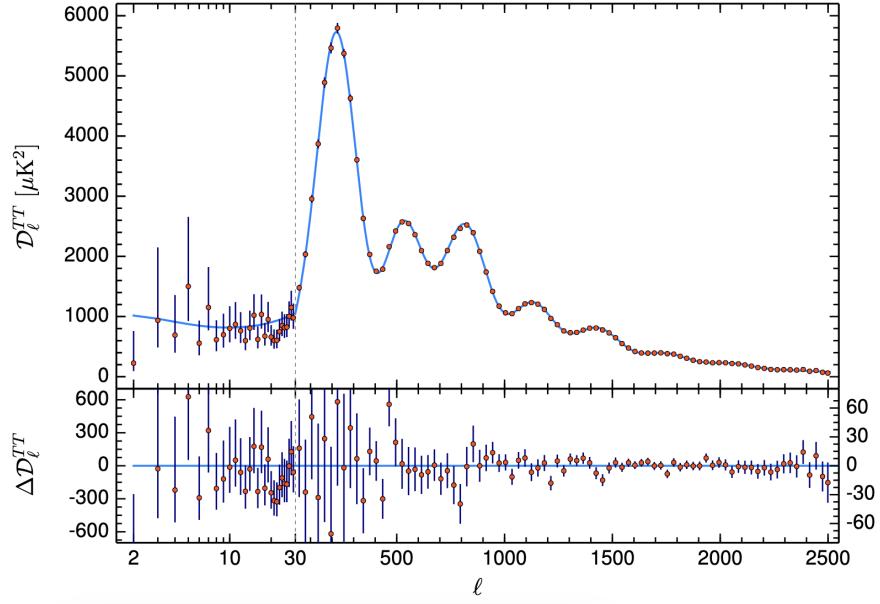


Figure 2.2: Planck TT power spectrum

2.2.1 TT Power Spectrum

The TT power spectrum measures the anisotropies of the temperature field. Thus, it's better to think of it as the deviation from the mean temperature. There are some prominent features of this spectrum worth discussing. The first is the oscillations, known as baryon acoustic oscillations. The second is the overall downward trend from large to small scales. This is generally due to diffusion of light through the early universe, and is given the name diffusion damping.

Now one can observe qualitatively how the cosmological parameters will affect the TT power spectrum. First, let's consider Ω_b , the baryon density. If the density is increased, there should be an increase in Compton scattering and decrease in the mean free path of photons, meaning the large scale anisotropy will be larger. Inversely, a reduction in the baryon density would decrease the large scale anisotropy. This is typically observed as the height of the first peak in the TT power spectrum. There will also be a small change in the peak position due to the change in the sound horizon. The height of the second peak, however, decreases. This is due to the increased pressure, which drives down the overall temperature fluctuations, resulting every other peak changing

in opposite directions.

Next is to consider a change in Ω_c . The dominant effect from changing Ω_c comes from the integrated Sachs-Wolfe (ISW) effect, which is an additional gravitational redshift that occurs between the time of reionization and now. Since the evolution of dark matter structure is slow, increasing Ω_c reduces the ISW effect, thus reducing the anisotropy and causing an overall downward shift in the TT power spectrum.

The remaining parameters are considered in unison: \mathcal{A}_s , n_s , and τ_{rei} . Naturally, τ_{rei} will have an overall suppression effect, The higher the anisotropy, the higher the suppression, thus resulting in strong suppression at large scales and weak suppression at small scales. The amplitude and spectral index enter the power spectrum in a similar way they did for weak lensing. Thus the effect of \mathcal{A}_s and n_s can look identical to changes in τ_{rei} .

Finally is a consideration of H_0 , arguably the most simple case. H_0 affects the expansion rate of the universe, thus a larger expansion rate will shift the position of the peaks in the power spectrum. Increases in H_0 will shift the peaks to smaller scales (found by backpropagating H_0 through time), and inversely, lowering H_0 will shift the peaks to larger scales.

2.2.2 TE and EE Power Spectra

The other two power spectra are shown in figures (figs). The main consideration with these two power spectra is to observe the way it breaks the degeneracy between τ_{rei} and \mathcal{A}_s, n_s . In the TT case, τ_{rei} suppresses the power spectrum as it suppresses the fraction of light that reaches an observer. However, the additional scattering actually causes anisotropies in the polarization power spectrum to increase. Thus the effects of \mathcal{A}_s and τ_{rei} are similar in the TT power spectrum but opposite in the *EE* power spectrum. Thus, one needs all 3 power spectra to reliably infer the cosmological parameters.

2.3 Tensions Between Weak Lensing and CMB

As seen, there are a total of five parameters needed to describe Λ CDM in weak lensing: Ω_m , Ω_b , H_0 , \mathcal{A}_s , and n_s . This parameterization is not unique.

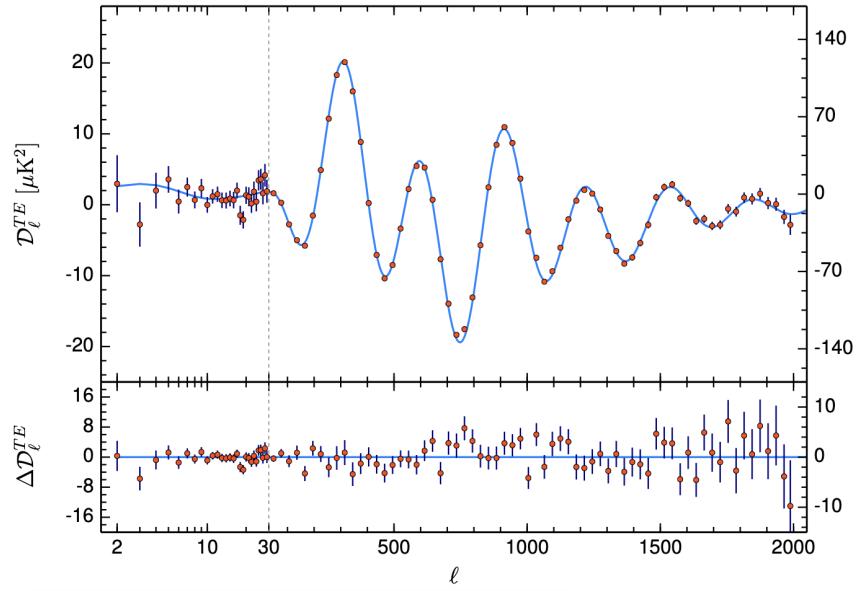


Figure 2.3: Planck TE power spectrum

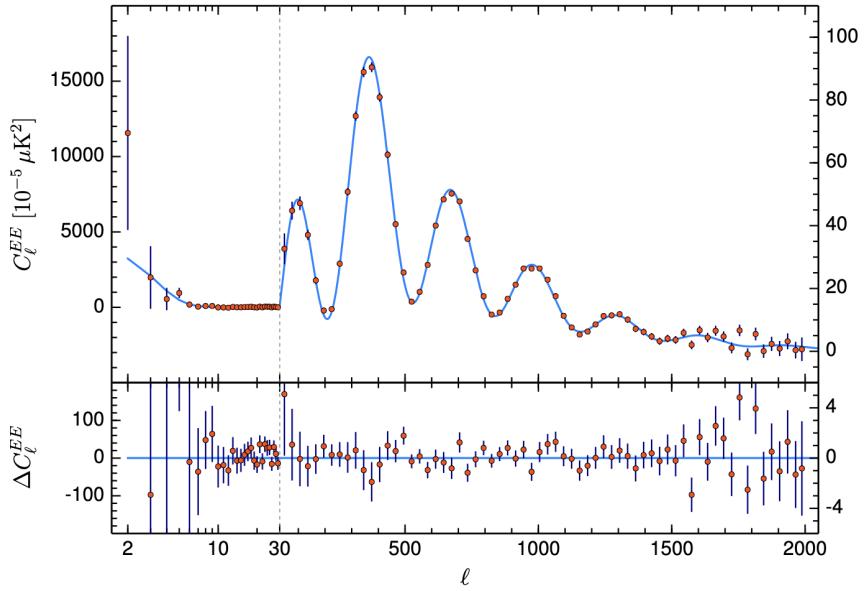


Figure 2.4: Planck EE power spectrum

Another common parameter, which tracks \mathcal{A}_f , is σ_8

$$\sigma_8^2 = \int_0^\infty P(k, r) \left(\frac{3j_1(kr)}{kr} \right)^2 \frac{k^2}{2\pi^2} dk \Bigg|_{r=8} \quad (2.48)$$

which can be described as the amplitude of matter fluctuations at the scale $8h^{-1}\text{Mpc}$. With this parameter instead of \mathcal{A}_s , weak lensing gives string constraints on $S_8 = \sigma_8 \sqrt{\Omega_m/0.3}$.

The reality is, however, that CMB measurements of some of these parameters differ between the parameters measured from weak lensing. The most famous discrepancy is the H_0 tension (figure 2.5). Weak lensing measurements are significantly higher than CMB measurements; a nearly 5σ difference. The other famous tension is the S_8 tension at about a 3σ difference (figure 2.6). There have been extensive efforts to propose a model that can resolve these tensions (modified gravity, early dark energy, quintessence, etc.), but as of now no model is able to simultaneously resolve both tensions. For example, early dark energy resolves the H_0 tension, but the S_8 tension tends to either remain the same or increase. The remaining chapters will explore how to compute these tensions and demonstrate a proposed consistency test.

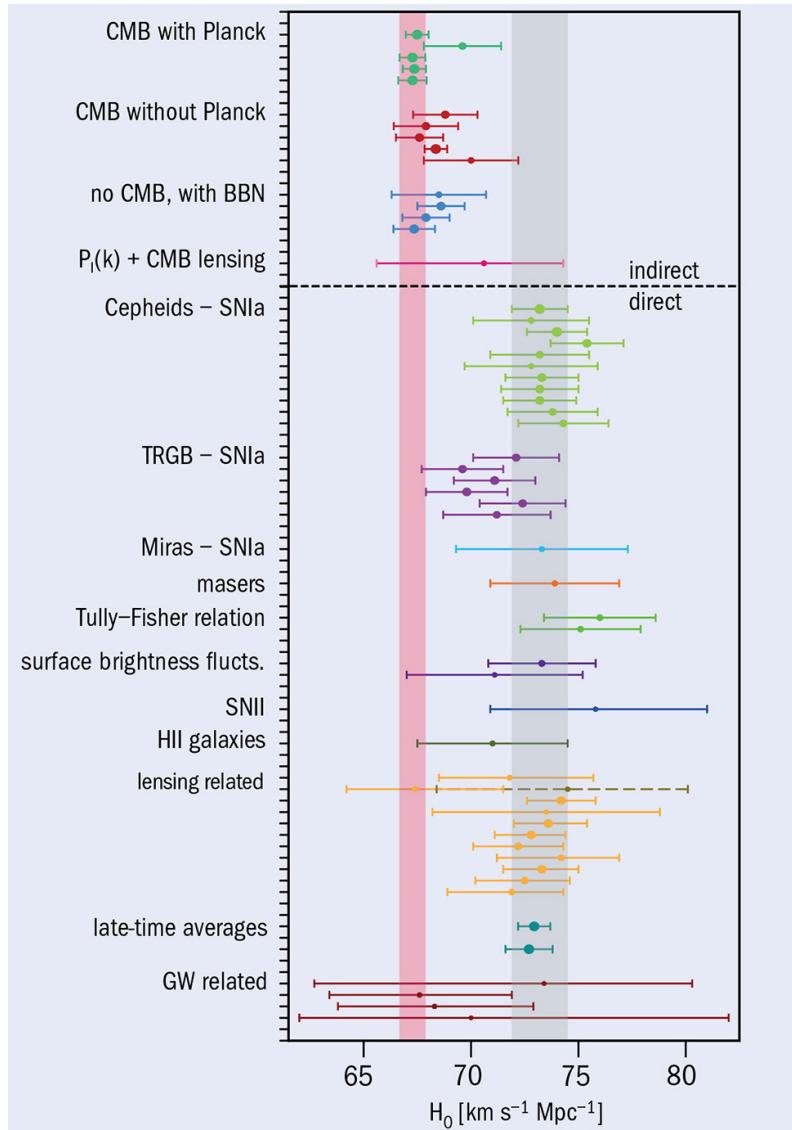


Figure 2.5: Measurements of H_0 from different experiments.

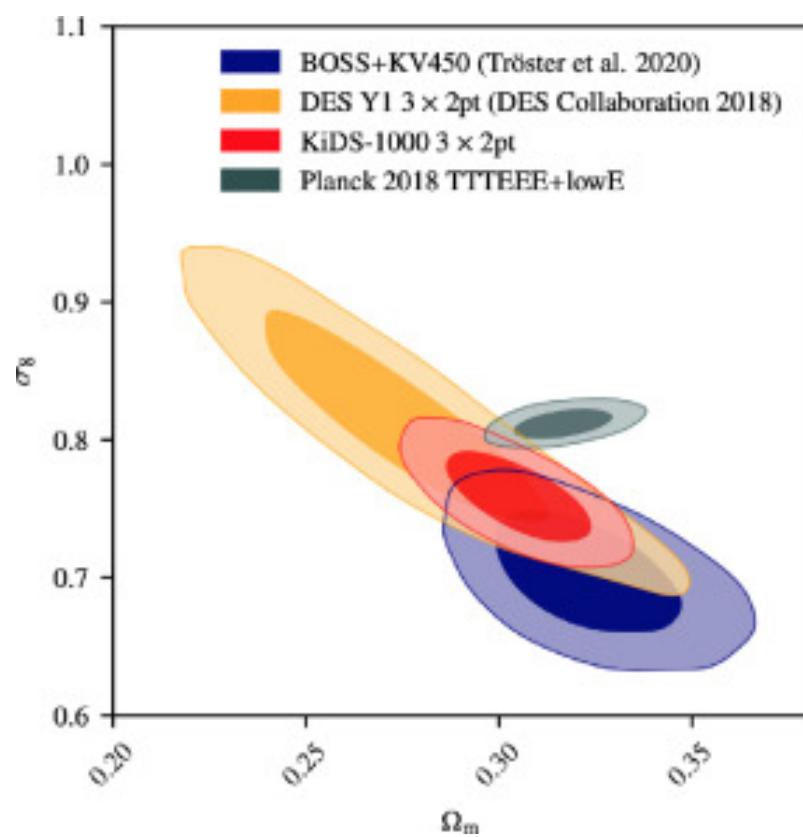


Figure 2.6: Measurements of S_8 from different experiments.

Chapter 3

Computing Techniques

The previous chapter demonstrated how the cosmological parameters go into computing the data vectors. But given some data, how does one find the cosmological parameters? Its clear from the complexity of the calculations that one cannot simply invert the functions to find the associated cosmological parameters. Instead, one would generate a set of cosmological parameters, calculate the data associated with those parameters, then compute the probability of observing the data given the set of parameters. Using this language, its clear that this computation involves conditional probability densities, thus we can use Bayes' theorem to help us determine the parameters.

3.1 Markov Chain Monte Carlo

Markov chains are a step process in which the next step only depends on the current step. Formally, a sequence X_1, \dots, X_n of random elements is a *Markov Chain* if the conditional distribution X_{n+1} depends only on X_n . The set in which X_i take values is called the *state space* of the chain. This is a very abstract definition, however Markov chains play a regular role in daily life. For example, weather patterns form Markov chains, or the order of words in a sentence. The key here is that, if it rains one day, how likely is it to rain the next day? Or given a word, what is the most probable next word? For a cosmological application, the key is to start with some set of parameters, and slowly move towards the most probable set of parameters.

Bayes' theorem states

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)} \quad (3.1)$$

Lets set $x = \theta$, the cosmological parameters, and $y = d$, the data vector representing the 3×2 correlation functions. Then

$$P(\theta|d) = \frac{P(d|\theta)P(\theta)}{P(d)} \quad (3.2)$$

The left hand side is typically called the *posterior*, while $P(d|\theta)$ is called the likelihood, $P(\theta)$ the prior, and $P(d)$ the evidence. In cosmological notation [10], this is usually written

$$\mathcal{P}(\theta) = \frac{\mathcal{L}(\theta)\Pi(\theta)}{\mathcal{E}} \quad (3.3)$$

The evidence is typically very difficult to compute. One must be able to compute an integral of a complicated function over the entire parameter space (it is possible via techniques such as nested sampling, however these usually take a very long time). Thus, with a given likelihood and prior, its not feasible to compute the posterior exactly. Markov chain Monte Carlo (MCMC) describes a probabalistic technique that can generate samples from the posterior only using the likelihood and prior and without computing the evidence.

3.1.1 The Metropolis-Hastings Algorithm

The Metropolis-Hastings algorithm [11–13] is the most common algorithm for doing MCMC. Suppose we want to sample from a distribution $p(x)$. We start with a proposal distribution $g(x_n)$. Sample from the proposal distribution to find the next state x_{n+1} with probability $g(x_{n+1}|x_n)$. This transition from state n to state $n + 1$ must follow the *detailed balance condition*

$$p(x_n)g(x_{n+1}|x_n)A(x_n \rightarrow x_{n+1}) = p(x_{n+1})g(x_n|x_{n+1})A(x_{n+1} \rightarrow x_n) \quad (3.4)$$

where A is an *acceptance probability* which I will define more precisely later. Using Bayes' Theorem on $p(x)$, the evidence cancels out on each side, and thus the detailed balance condition can be simplified to only rely on the likelihood and prior of $p(x)$, which I will denote π and \mathcal{L} .

$$\pi(x_n)\mathcal{L}(x_n)g(x_{n+1}|x_n)A(x_n \rightarrow x_{n+1}) = \pi(x_{n+1})\mathcal{L}(x_{n+1})g(x_n|x_{n+1})A(x_{n+1} \rightarrow x_n) \quad (3.5)$$

$$\Rightarrow \frac{A(x_n \rightarrow x_{n+1})}{A(x_{n+1} \rightarrow x_n)} = \frac{\pi(x_{n+1})\mathcal{L}(x_{n+1})g(x_n|x_{n+1})}{\pi(x_n)\mathcal{L}(x_n)g(x_{n+1}|x_n)} \equiv R_{n,n+1} \quad (3.6)$$

This allows us to define the acceptance probability as

$$A(x_n \rightarrow x_{n+1}) = \min(1, R_{n,n+1}) \quad (3.7)$$

This probability is used to determine whether the chain moves to x_{n+1} or stays at x_n . The chain converges when it reaches a stationary state.

There are a few properties that can be observed for this algorithm:

- Having an asymmetrical proposal $g(x)$ can allow for faster convergence of the chain.
- The initial sampling may not accurately reflect samples for $p(x)$. This is regarded as the ‘burn-in’ and is generally discarded from the samples.
- MCMC Sampling loses sampling power for multi-modal distributions.

3.2 Tension Metrics

Given the 5-dimensional parameter space, there is great difficulty in computing the tension between two measurements. If the data is projected onto one parameter, this can hide the true separation in parameter space (see figure for a gaussian example). Thus, one needs a method to compute the tension in the full parameter space. In addition, posterior distributions are not generally Gaussian, further increasing the complexity. There are many metrics used to measure tension [4], each with different assumptions and consequences.

3.2.1 Parameter-based Metrics

Metric 1: Parameter Difference

The main idea of this metric is that if two data sets largely agree, their difference posterior will be centered at 0, so mass of the posterior above $\mathcal{P}(0)$ will be close to zero. To formalize this notion, start by sampling from the posteriors $\mathcal{P}(\theta_A)$ and $\mathcal{P}(\theta_B)$. Define $\Delta\theta = \theta_A - \theta_B$. Under this reparameterization the two posteriors are $\mathcal{P}(\theta_A)$ and $\mathcal{P}(\theta_A - \Delta\theta)$. If the posteriors are independent one can marginalize over θ_A to get the parameter difference distribution.

$$\mathcal{P}(\Delta\theta) = \int \mathcal{P}(\theta_A)\mathcal{P}(\theta_A - \Delta\theta)d\theta_A \quad (3.8)$$

Thus we can approximate the tension as the volume of the posterior contours with $\mathcal{P}(\Delta\theta) > \mathcal{P}(0)$.

$$\text{PTE} = \int_{\mathcal{P}(\Delta\theta) > \mathcal{P}(0)} \mathcal{P}(\Delta\theta) d\Delta\theta \quad (3.9)$$

This has the benefit of only requiring independent parameter sampling. It makes no assumption of the gaussianity of the posterior or the likelihood. The difficulty here, however, is actually determining $\mathcal{P}(\theta)$. There are multiple methods used, the most novel of which is discussed later.

Metric 2: Parameter Difference in Update Form

As discussed above, we can compute the parameter Q_{UDM} by

$$Q_{\text{UDM}} = (\mu_A - \mu_{A+B})^T (C_A - C_{A+B})^{-1} (\mu_A - \mu_{A+B}) \quad (3.10)$$

The difference of means is precisely the mean of the parameter difference distribution, and we are using the covariance $C_A + C_{A+B}$. Thus it is clear Q_{UDM} is χ^2 distributed with degrees of freedom given by $\text{rank}(C_A - C_{A+B})$. It is clear, however, that this metric relies on the parameter difference to be gaussian distributed because of its reliance on Q_{UDM} being χ^2 distributed. Despite this, we proceed anyway. With proper calibration this metric can be useful even for non-gaussian posteriors.

Metric 3: Goodness of Fit Degradation

For this metric, first find the *maximum a posteriori*, the point $\hat{\theta}$ that maximizes the posterior. Then compute the likelihood $\mathcal{L}(\hat{\theta})$. Lastly compute the difference in likelihoods between the sum of chain A and B and the joint chain $A + B$

$$Q_{\text{DMAP}} = 2\mathcal{L}_A(\hat{\theta}_A) + 2\mathcal{L}_B(\hat{\theta}_B) - 2\mathcal{L}_{A+B}(\hat{\theta}_{A+B}) \quad (3.11)$$

Q_{DMAP} is again χ^2 -distributed. Interesting part here is how to determine the number of degrees of freedom. The likelihood function, which is on data-space, has degrees of freedom of approximately 1500 for LSST, but we give it an input of cosmological parameters which belong to a 5 dimensional parameter space. Thus the degrees of freedom is related to the number of parameters that are constrained by the likelihood. This can be found by examining the ratio of the variance in the posterior to the variance in the prior; a large ratio of $\mathcal{O}(1)$ means a parameter is not constrained by the likelihood and a ratio $\ll 1$ means a parameter is well constrained by the likelihood. Thus we define the number

of degrees of freedom as

$$d = N - \text{tr}(C_{\Pi}^{-1} C_{\mathcal{P}}) \quad (3.12)$$

Metric 4: Eigentension

This metric differs from the others by restricting our parameters to ones that are well-measured. We do this as follows:

1. Diagonalize the covariance matrix on one of our data sets.
2. Take the ratio of the variance in the prior to the variance in the posterior and apply an *ad hoc* cut to determine which eigenmodes are well-measured (generally 100 is a good choice).
3. Project the other data set onto the eigenmodes.
4. Compute the tension using one of the previous methods.

The idea is that one should not include poorly measured eigenmodes in our tension analysis because the difference is dominated by the prior rather than the likelihood.

3.2.2 Evidence-Based Metrics

- Bayesian evidence ratio given by

$$R = \frac{\mathcal{E}_{AB}}{\mathcal{E}_A \mathcal{E}_B} \quad (3.13)$$

Where A and B are data sets. This method can be written many ways using bayes theorem, which will make it apparent that this metric depends heavily on the prior volume.

- Bayesian suspiciousness. This metric attempts to remove the dependence on the prior volume by defining the suspiciousness as

$$\log S = \log R - \log I \quad (3.14)$$

Of particular interest here is the new value I which is the information ratio, which is defined in terms of the KL divergence.

$$\log I = \mathcal{D}_A + \mathcal{D}_B - \mathcal{D}_{AB} \quad (3.15)$$

$$\mathcal{D} = \int \mathcal{P} \log \left(\frac{\mathcal{P}}{\Pi} \right) \quad (3.16)$$

Interpreting the Results

Given some probability P of a parameter shift, the following formula can give you the number of standard deviations if the probability shift comes from a gaussian distribution

$$n_\sigma = \sqrt{2} \operatorname{Erf}^{-1}(P) \quad (3.17)$$

However we don't have any concrete reason to believe P comes from a Gaussian distribution in general. However, it does come from a gaussian distribution when the posteriors are Gaussian, thus we can do a test of the metrics by taking two unit Gaussian posteriors and fixing the separation between them. In fact, this example can be computed analytically.

$$\begin{aligned} \mathcal{P}(\Delta\theta) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\theta^2/2} e^{-(\theta-\Delta\theta)^2/2} d\theta \\ &= \frac{1}{2\pi} \cdot \sqrt{\pi} e^{-(\Delta\theta)^2/4} \\ &= \frac{1}{\sqrt{4\pi}} e^{-(\Delta\theta)^2/4} \end{aligned} \quad (3.18)$$

The parameter difference posterior is a gaussian with standard deviation $\sqrt{2}$. The separation is fixed by a , hence the shift is $\mathcal{P}(a)$. Hence the shift probability is

$$\Delta = \int_{-a}^a e^{-(\Delta\theta)^2/4} d\Delta\theta \quad (3.19)$$

Thus we can compare the analytic result to the computational result for each of the metrics to see which ones are most accurate. In practice, this integral is computed with Monte-Carlo integration, where the error is evaluated using the Clopper-Pearson interval on the binomial distribution.

3.2.3 DES v Planck Results

A subset of these excercises have been performed previously by the Dark Energy Survey (DES) against the planck results. In [4], Lemos et. al. generates an a-priori tension on the cosmological parameters by shifting the DES fiducial cosmology along σ_8 and Ω_m . They appoximate the a-priori tension using the

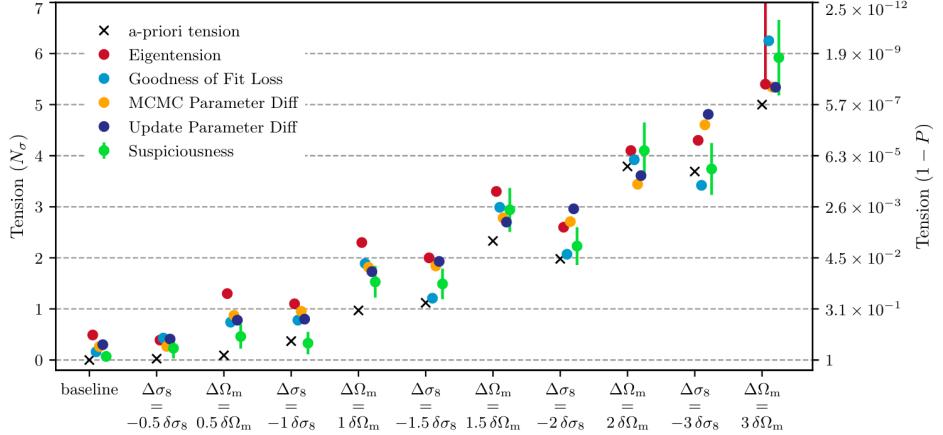


Figure 3.1: Result of tension metrics between Planck and DES (figure 6 in [4]). A-priori tension is the gaussian approximation described in equation 3.20.

gaussian approximation,

$$\chi^2 = \delta\theta(C_{\text{DES}} + C_{\text{Planck}})\delta\theta \quad (3.20)$$

where $\delta\theta$ is (σ_8, Ω_m) and C the covariance of DES and Planck respectively.

They then compute the tension metrics and compare to the a-priori gaussian tension. The main result demonstrates what the resulting tension is from a fixed shift along a single parameter. Their result is presented in this thesis in figure 3.1. An additional note is that gaussian tension tends to underestimate the tension presented in other metrics, highlighting the importance of capturing non-gaussianities.

3.3 Normalizing Flows

The method of normalizing flows (MAF) implemented here uses Masked Autoencoders (MADE) to construct the flow [10, 14, 15]. Suppose we have an input to the flow x_i . The output of the map is $y_i = \mu(x_{1:i-1}) + \sigma(x_{1:i-1})x_i$. The μ and σ are found using neural networks which receive masked inputs $x_{1:i-1} = (x_1, \dots, x_{i-1}, 0, \dots, 0)$. Since the input only depends on the first $i-1$ inputs, the normalizing flow is *autoregressive* and the Jacobian is triangular.

The implementation in tensorflow uses *bijectors* which implements a local diffeomorphism between a manifold M and a target manifold N (which are our

parameter spaces), i.e. $\phi : M \rightarrow N$ such that ϕ is differentiable and injective. In tensorflow it has three operations, Forward, Inverse, and log_deg_jacobian, which are exactly the three we want. By constructing a bijector for each masked input, the full normalizing map can be constructed.

An intuitive way to think of normalizing flow is as a reparameterization. Consider our posterior as a manifold X . The manifold is constructed via the parameterization with charts $\phi_i : X \rightarrow U_i \subset \mathbb{R}^n$ for i in indexing set I . The charts determine the atlas $A = \bigsqcup_{i \in I} (U_i, \phi_i)$. The goal is to determine a new parameterization for X given by the charts $\psi_j : X \rightarrow V_j \subset \mathbb{R}^n$ in which samples in X are gaussian distributed. Let $Y \subset X$ be given by $Y = \phi_i^{-1}(U_i) \cap \psi_j^{-1}(V_j)$. Then the transition function is given by $\tau : \phi_i(Y) \rightarrow \psi_j(Y)$. To ensure our parameterizations are diffeomorphic, τ is required to be differentiable and its inverse differentiable (in fact, we are working with smooth reparameterizations so the transition function and its inverse should be smooth).

The next question one may have is ‘what do flows have to do with this?’ Above, the concept of normalizing flows is described as a reparameterization, but there is another good description involving flows. Again, let X be our parameter space, the graph $(\theta, \mathcal{P}(\theta))$, and TX the tangent bundle of X . Each MAF can be thought of as a finite sequence of vector fields $v_i : X \rightarrow TX$. Each point is moved along its integral curve determined by each v_i . What is great about this definition is that it immediately implies an MAF is a diffeomorphism.

Proposition 1. *Suppose $v : X \rightarrow TX$ is a vector field on X and $x \in X$. Then for any $\epsilon > 0$ there exists a unique function $f : B[-\epsilon, \epsilon] \rightarrow X$ such that f is an integral curve of with infinitesimal generator v and with $f(0) = x$.*

Proof. Given a vector $v_x : U \rightarrow T_x X$ with $x \in U$, for it to be in the tangent space $T_p X$ it must have a curve with $f(0) = x$ and $f'(0) = v_x$. This follows from the definition of $T_x X$.

Suppose there are two such curves, α and β both mapping from $I = (-\epsilon, \epsilon)$ to X and with $\alpha(0) = x = \beta(0)$ and both with $\alpha'(0) = v_x = \beta'(0)$. Then they both solve the system of autonomous ODE’s for $i = 1, \dots, n$

$$\begin{aligned} \dot{\gamma}^i(t) &= v^i(\gamma(t)) \\ \gamma^i(0) &= c^i \end{aligned} \tag{3.21}$$

This can be solved by $\gamma^i(t) = \int_{U \subset X} v^i(\gamma(t)) \text{vol.}$. Since v is a tangent vector

field, there exists a curve ϕ with $\phi' = v$, so the solution is $\gamma^i(t) = \phi^i(t) + C^i$. The integration constants are fixed by the initial condition $C^i = -c^i$, hence any two functions satisfying this system of ODE's, including α and β , are equivalent for all $t \in I$. \square

Autoregressive models have several pitfalls that we would like to avoid in this project. Namely, autoregressive models are sensitive to the order of the input, however our result does not have any dependence.

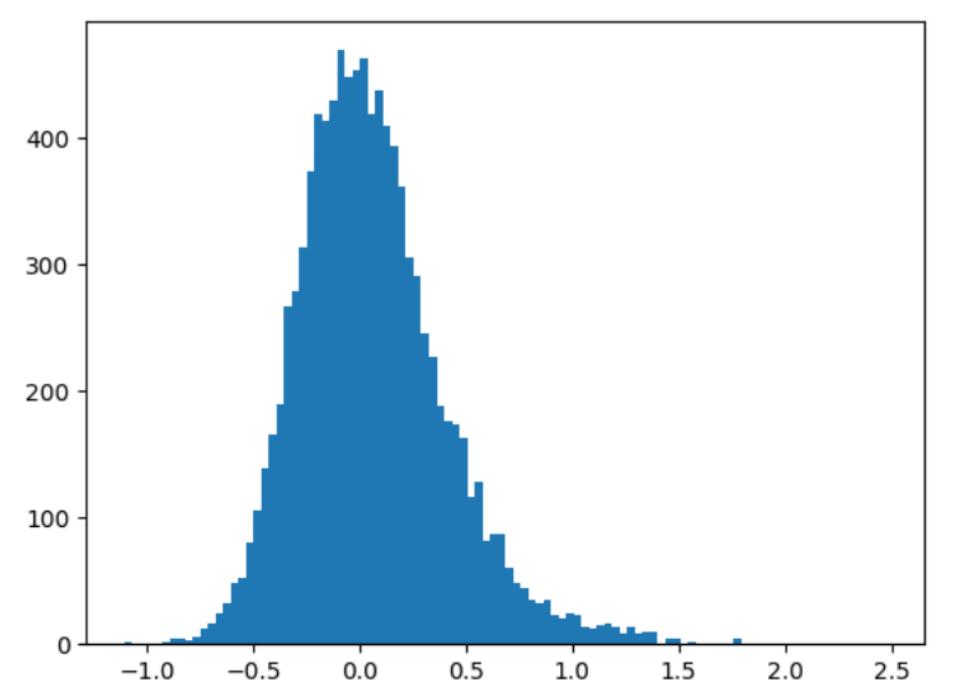


Figure 3.2: An example non-Gaussian distribution (non-zero skewness).

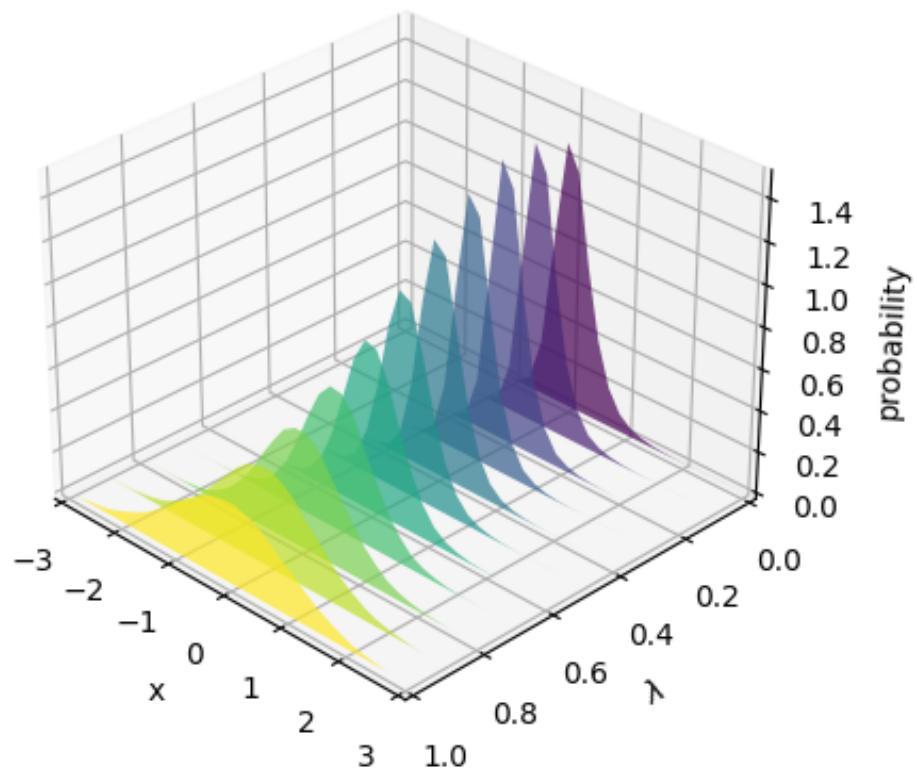


Figure 3.3: An example of normalizing flow as the transformation is ‘turned on’, represented by the parameter λ . The non-Gaussian distribution at $\lambda = 0$ is transformed into the Gaussian distribution at $\lambda = 1$.

Chapter 4

Growth-Geometry Split in DES-Y3

Parameter splitting is a common method used to attempt to resolve the tensions discussed in the previous chapter. The idea is to split a parameter and choose analysis settings that restrict each split parameter to describe specific cosmological effects. One of these splits is a growth-geometry split, where we split the relative matter density Ω_m and the dark energy equation of state w into two parameters. The growth parameter is designed to describe the late-time growth of structure and the geometry parameter designed to describe the background evolution. After splitting the parameters, one can do an analysis to determine how the tension changes and if the split parameters have distinct values from each other. This chapter describes the growth geometry split in Ω_m and w in DES-Y1 [16] and DES-Y3. These results are also available in a paper written by Kunhao Zhong, another Masters student, and myself, on arXiv [17] and are pending publication to Physical Review.

4.1 Split Matter Power Spectrum

Since the matter density Ω_m is split into two parameters, Ω_m^{growth} and Ω_m^{geo} , single growth factor and a single matter power spectrum; these are split as well. From equation (eq), the linear matter power spectrum is proportional to the square of the linear growth factor $D_+(z(a)) = G(z)/(1+z)$, so the growth

and geometry power spectra are related by

$$P_{\text{split}}^L(k, z) = P_{\text{geo}}^L(k, z) \left(\frac{G_{\text{growth}}(z)}{G_{\text{geo}}(z)} \right)^2 \quad (4.1)$$

which means σ_8 is also split according to

$$\sigma_{8,\text{split}}^2 = \sigma_{8,\text{geo}}^2 \left(\frac{G_{\text{growth}}(z)}{G_{\text{geo}}(z)} \right)^2 \quad (4.2)$$

In this analysis, however, we want to go beyond linear scales. To do this, we use the Euclid Emulator to emulate the non-linear matter power spectrum by computing a boost factor

$$P(k, z) = P^L(k, z) \times B(k, z) \quad (4.3)$$

While the growth parameters are scale independent, there is a crossing for

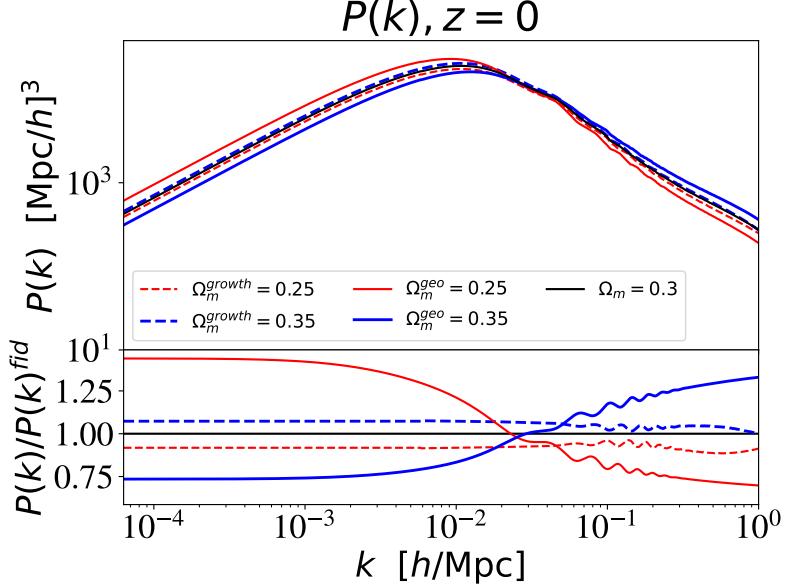


Figure 4.1: Non-linear power spectrum from Euclid emulator. When either the geometry or growth parameters are varied, the other is kept fixed at $\Omega^X = 0.3$.

the geometry parameters. This is due to the enhanced gravitational effects for higher Ω_m^{geo} , or suppressed gravitational effects for lower Ω_m^{geo} . From this, one can predict that the value of σ_8 will change, the change being larger for Ω_m^{geo} than Ω_m^{growth} , and increases in either Ω_m will increase σ_8 . We see this is true in figure 4.2a One surprising feature of figure 4.2b is that these effects are the

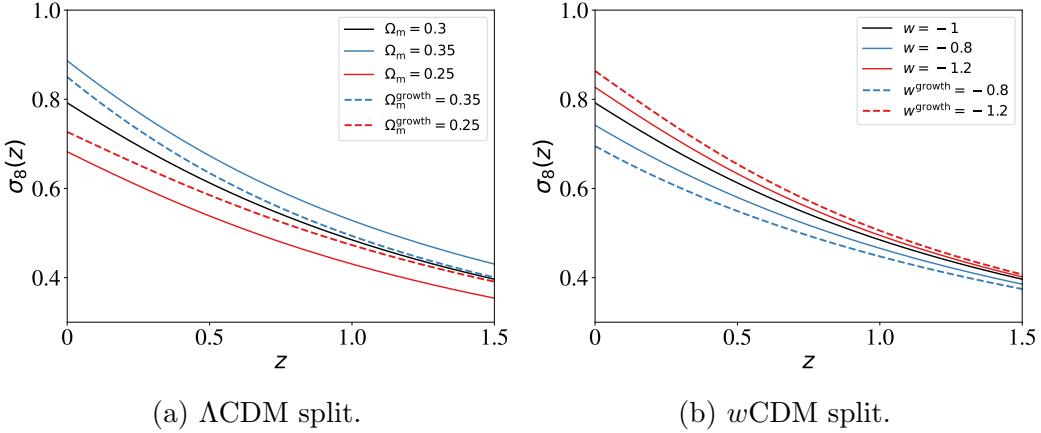


Figure 4.2: Effect of changing growth and geometry parameters on $\sigma_8(z)$.

parameter	prior
Ω_m^{geo}	$\mathcal{U}(0.1, 0.9)$
w^{geo}	$\mathcal{U}(-3, -0.01)$
$\mathcal{A}_s \times 10^9$	$\mathcal{U}(1.7, 2.5)$
n_s	$\mathcal{U}(0.92, 1.0)$
H_0	$\mathcal{U}(61, 73)$
τ	$\mathcal{U}(0.01, 0.8)$
Ω_m^{growth}	$\mathcal{U}(0.24, 0.4)$
w^{growth}	$\mathcal{U}(-1.7, -0.7)$

Table 4.1: Summary of cosmological parameters and their priors.

exact opposite. However, this result is not surprising. Increasing w , where $w < 0$, coincides with decreasing the pressure or increasing density.

4.2 Data and Analysis Pipeline

4.2.1 DES Parameter Priors

This analysis is done on data from the Dark Energy Survey (DES), both year 1 (Y1) and year 3 (Y3), combined with CMB data. The cosmological parameters of interest and their priors are given in table 4.1. The priors on $(\Omega_m^{\text{geo}}, w^{\text{geo}}, \mathcal{A}_s \times 10^9, n_s, H_0)$ are taken from the Euclid Emulator prior. τ is only included in chains that also contain CMB data. The external data sets used are

DES Systematics Parameter	Y1 Prior	Y3 Prior
Linear Galaxy bias		
$b_g^i (i \in [1, 5])$	Flat(0.8, 3.0)	Flat(0.8, 3.0)
Intrinsic Alignment (NLA)		
A_1	Flat(-5, 5)	
A_2	Flat(-5, 5)	
Intrinsic Alignment (TATT)		
A_1		Flat(-5, 5)
A_2		Flat(-5, 5)
η_1		Flat(-5, 5)
η_2		Flat(-5, 5)
b_{TA}		Flat(0, 2)
Source photo-z		
$\Delta z_s^1 \times 10^2$	Gauss(-0.1, 1.6)	Gauss(0, 1.8)
$\Delta z_s^2 \times 10^2$	Gauss(-0.19, 1.3)	Gauss(0, 1.5)
$\Delta z_s^3 \times 10^2$	Gauss(0.9, 1.1)	Gauss(0, 1.1)
$\Delta z_s^4 \times 10^2$	Gauss(-1.8, 2.2)	Gauss(0, 1.7)
Lens photo-z		
$\Delta z_l^1 \times 10^2$	Gauss(0.8, 0.7)	Gauss(0.6, 0.4)
$\Delta z_l^2 \times 10^2$	Gauss(-0.5, 0.7)	Gauss(0.1, 0.3)
$\Delta z_l^3 \times 10^2$	Gauss(0.6, 0.6)	Gauss(0.4, 0.3)
$\Delta z_l^4 \times 10^2$	Gauss(0, 0.01)	Gauss(-0.2, 0.5)
$\Delta z_l^5 \times 10^2$	Gauss(0, 0.01)	Gauss(-0.7, 0.1)
Multiplicative shear calibration		
$m_1 \times 10^2$	Gauss(1.2, 2.3)	Gauss(-0.6, 0.9)
$m_2 \times 10^2$	Gauss(1.2, 2.3)	Gauss(-2.0, 0.8)
$m_3 \times 10^2$	Gauss(1.2, 2.3)	Gauss(-2.4, 0.8)
$m_4 \times 10^2$	Gauss(1.2, 2.3)	Gauss(-3.7, 0.8)
Lens magnification		
$C_1^1 \times 10^2$		Fixed (0.63)
$C_1^2 \times 10^2$		Fixed (-3.04)
$C_1^3 \times 10^2$		Fixed (-1.33)
$C_1^4 \times 10^2$		Fixed (2.50)
$C_1^5 \times 10^2$		Fixed (1.93)
Point mass marginalization		
$B_i (i \in [1, 5])$		Flat(-5, 5)

Table 4.2: Summary of Priors on DES-Y1 and DES-Y3 systematics parameters.

- CMBP: The Planck CMB TTTEEE power spectra together with the non-linear low- ℓ EE power spectrum. The spectra are truncated after the first peak ($35 < \ell < 396$). This mitigates CMB lensing effects, which affect small-scale anisotropies. It also minimizes the late time Integrated Sachs Wolfe effect, which also has a larger effect on small scales.
- SNIa: Pantheon Type Ia supernova. As a distance probe, this prior constrains geometry parameters. There are growth effects that we choose not to model at this point, namely the peculiar velocity distributions of supernovae.
- BBN: We use a derived constraint on $100\Omega_bh^2$ from Big Bang Nucleosynthesis as
- BAO: BAO data is taken from the SDSS DR7 main galaxy sample, the 6dF galaxy survey, and the combined SDSS BOSS DR12 low- z and CMASS galaxy samples. Again, as a distance probe, this is taken as geometry information.

To understand these choices of priors, we consider the following combinations:

- Prior 1 (P1): Emulator prior + CMBP
- Prior 2 (P2): Emulator prior + SNIa + BBN + BAO
- Prior 3 (ALL): Emulator prior + CMBP + SNIa + BBN + BAO

4.2.2 Software Pipeline

To compute the linear matter power spectrum we use CAMB Boltzmann code, which is capable of solving the linear Boltzmann equation and computing transfer functions. For going beyond linear scales, we use the Euclid emulator, a fast neural network based emulator for Halofit. Previous DES studies used Halofit itself, a full N-body simulation to determine the non-linear evolution of dark matter. A baseline comparison between the Euclid emulator and Halofit was done (figure 4.3) We use the Cocoa (COBE-COsmolike joint Architecture). This interfaces Cobaya, the MCMC sampler we use, with Cosmolike, the data vector calculation framework. DES-Y3 has already been implemented in Cobaya, Cosmolike, and Cocoa. To assess the convergence of the MCMC, we use the $R - 1 < 0.02$.

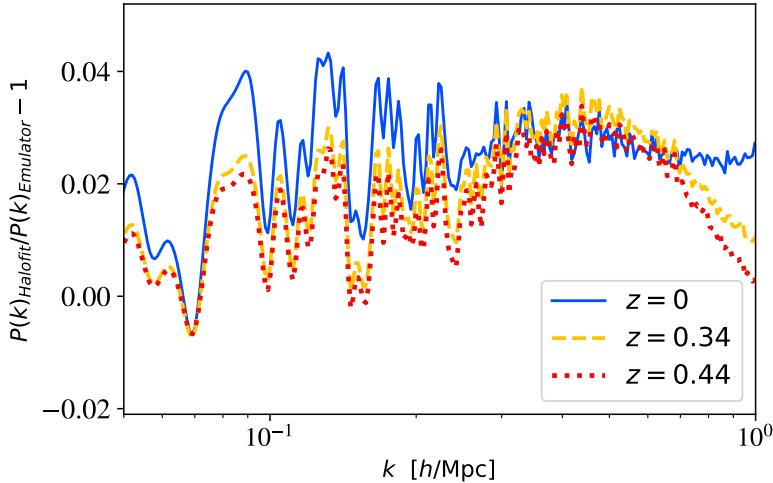


Figure 4.3: Relative difference between Euclid emulator and Halofit.

One benefit of this analysis is that growth parameters are semi-fast; they do not require CAMB to recompute the matter power spectrum, they only require Cosmolike to update the data vectors. Rerunning Cosmolike with only changing the growth parameters results in a 2-fold speedup.

4.2.3 Pipeline Validation on Synthetic Data

To validate this method, we generate a synthetic data vector at the Planck best-fit parameters without lensing, $(\mathcal{A}_s \times 10^{-9}, n_s, H_0, \Omega_m, \Omega_b) = (2.101, 0.965, 67.32, 0.317, 0.049)$, and we hope to find no difference between growth and geometry parameters and hope each parameter is equal to the fiducial value.

Using the ALL prior and external data, there is marginal improvement between the cosmic shear constraints on Ω_m^{growth} compared to the prior. However there is significant improvement in the 3x2pt case compared to cosmic shear.

The constraining power of 3x2pt on Ω_m^{growth} is identical between P1 and ALL. This is because ...

In w CDM, there is poor constraining power on Ω_m^{growth} and w^{growth} , even with the most informative 3x2pt case. For this case, we give constraining power along the first principle component

$$\text{PC}_1 \equiv 0.7071\Delta\Omega_m - 0.7071\Delta w. \quad (4.4)$$

In all 3 sets of priors and external data, and despite the inclusion of additional nuisance parameters from the TATT model, we find similar constraining power on Ω_m^{growth} between DES-Y1 and DES-Y3. In the 3x2pt case with the ALL prior gives a smaller ($\sim 17\%$) error bar. Constraints on Ω_m^{geo} are prior dominated and nearly identical for the w CDM split in DES-Y1 and DES-Y3.

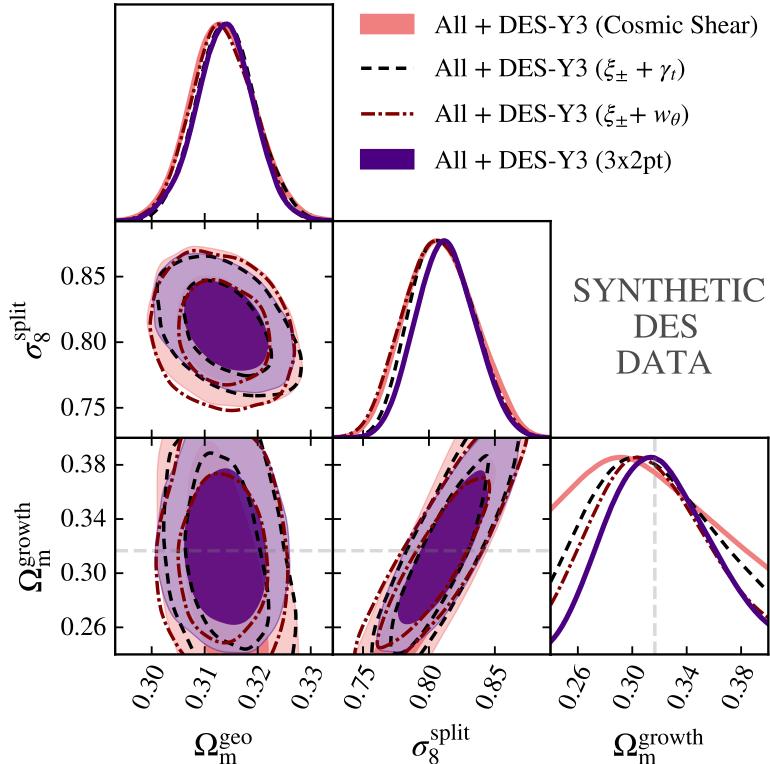


Figure 4.4: Cosmic shear, 2x2pt combinations, and 3x2pt posteriors for DES-Y3 with the ALL prior.

4.3 Results

4.3.1 LCDM

For the most constraining case, 3x2pt with all prior, there is no evidence of a difference between Ω_m growth and geometry parameters (figure 4.6). Exploring this further, however, we can look at various 2-point correlation functions in DES-Y1 and DES-Y3 (figure 4.7). We see a significant change in the DES-Y3 cosmic shear-galaxy clustering posterior and the galaxy lensing-galaxy clus-

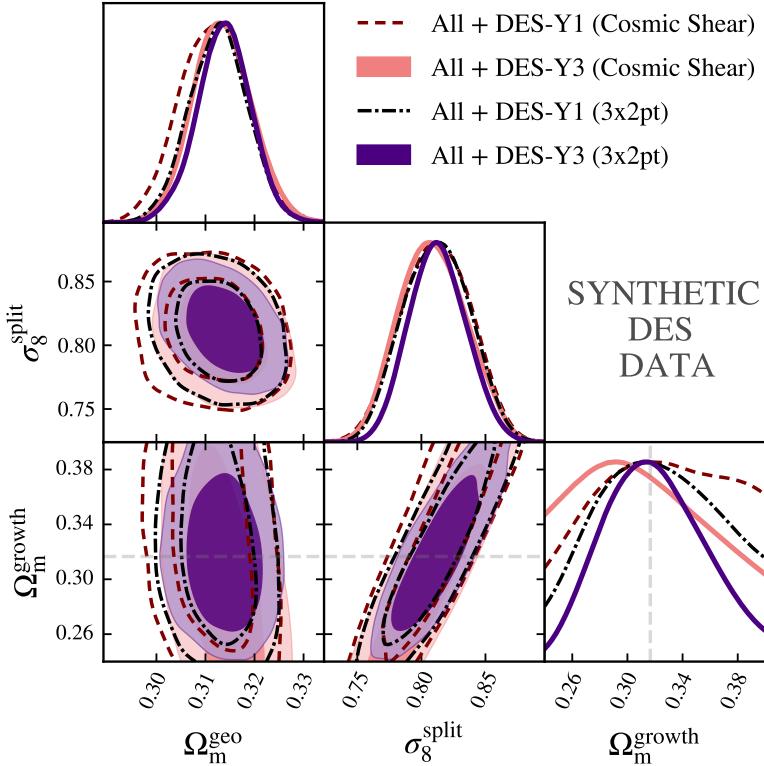


Figure 4.5: Comparison of DES-Y1 and DES-Y3 posteriors.

tering posterior. However, these shifts from $\Delta\Omega_m = 0$ are prior dominated, not data dominated, thus do not constitute evidence for a shift. The full results of the 1d marginalizations with each choice of prior is summarized in figure 4.8. Aside from the previously mention 2-point correlation functions, every $\Delta\Omega_m$ result is compatible with 0 to within 1-sigma. Similarly to the synthetic chains, the 3x2pt data chains with ALL prior have improved constraining power on $\Delta\Omega_m$ in DES-Y3 over DES-Y1. However, the improvement is around 10%, slightly less than the 17% observed in the synthetic data.

The mentioned 2-point correlation functions that do have significant shifts in Ω_m are determined to be at 1.75σ for the cosmic shear-galaxy clustering and 2.60σ for galaxy clustering-galaxy-galaxy lensing. Because of known inconsistencies in the RedMaGiC galaxy samples, these are not attributed to discovery.

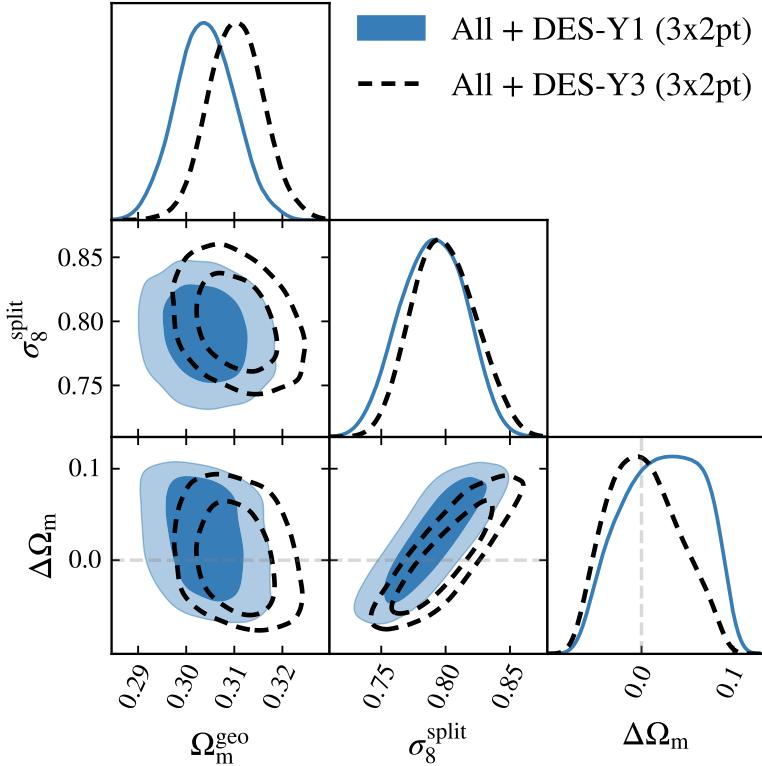


Figure 4.6: Split posteriors for DES-Y3 3x2pt and ALL prior.

4.3.2 wCDM

The results for w CDM are summarized in figures 4.9 and 4.10. Again, there is no evidence of a shift in PC_1 from 0. The exception is, once again, the DES-Y3 2x2pt with ALL prior. The significance of the shift is 4.48σ . Again, this shift may be attributed to the RedMaGiC.

4.3.3 Tension Analysis

We conclude by testing the internal consistency between the Y1 and Y3 data sets independently. This is done using the parameter difference method, where we use normalizing flows to learn the posterior distribution. The results are summarized in figure 4.11. The tension is computed with the parameters $(\mathcal{A}_s, n_s, H_0, \Omega_m^{\text{geo}}, \sigma_8^{\text{split}})$ (with w^{geo} added in the w CDM case). Interestingly, the tension appears to be generated predominantly by \mathcal{A}_s . Additionally, the 2x2pt correlation functions have a significant degradation in the goodness of fit when CMB priors are added. This reflects the RedMaGiC inconsistency

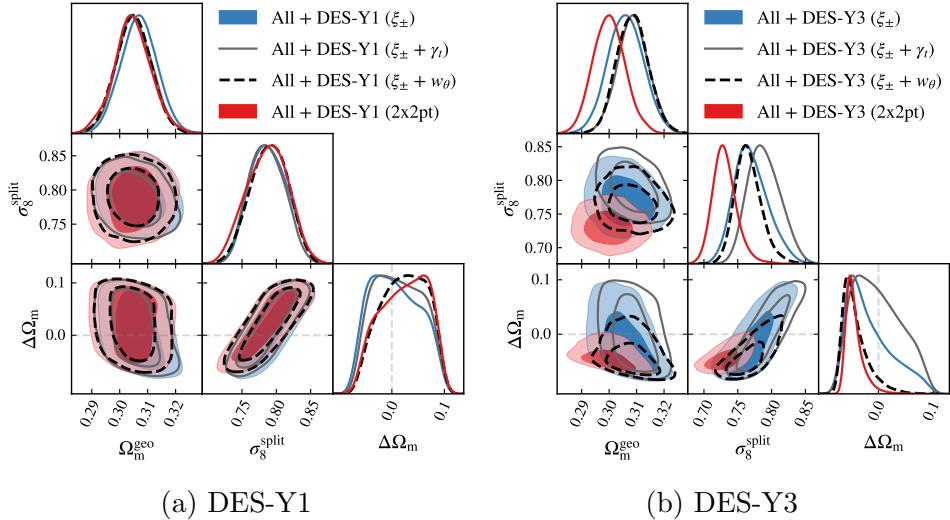


Figure 4.7: DES-Y1 and DES-Y3 posteriors for 2-point correlation functions.

discussed above. A study allowing X_{lens} to vary is required to fully analyze this shift.

In the w CDM case, we once again see the 2x2pt has a significant shift. Contrary to the split Λ CDM case, the A_s tension is much smaller, making the detection of a shift in PC₁ more meaningful. Once again, an analysis involving X_{lens} is required to assess the source of this detection.

4.4 Conclusion

We split Ω_m and w into two parameters, one governing the background evolution and the other governing the late-time scale-independent growth of structure. We consider DES data together with various external data sets and priors from the CMB, supernovae, BBN, and BAO. We then run MCMC on DES-Y1 and DES-Y3 data and evaluate the difference between the growth and geometry parameters.

As demonstrated, there is little evidence of a difference between growth and geometry parameters in DES-Y1 and DES-Y3. The exception is the 2x2pt chains, however there is an internal inconsistency with the RedMaGiC samples that may source this result. Because we fixed $X_{\text{lens}} = 1$, an additional analysis can be done investigating if X_{lens} sources significant shift. Additionally, one can repeat the analysis with the MagLim lensing samples. In the future,

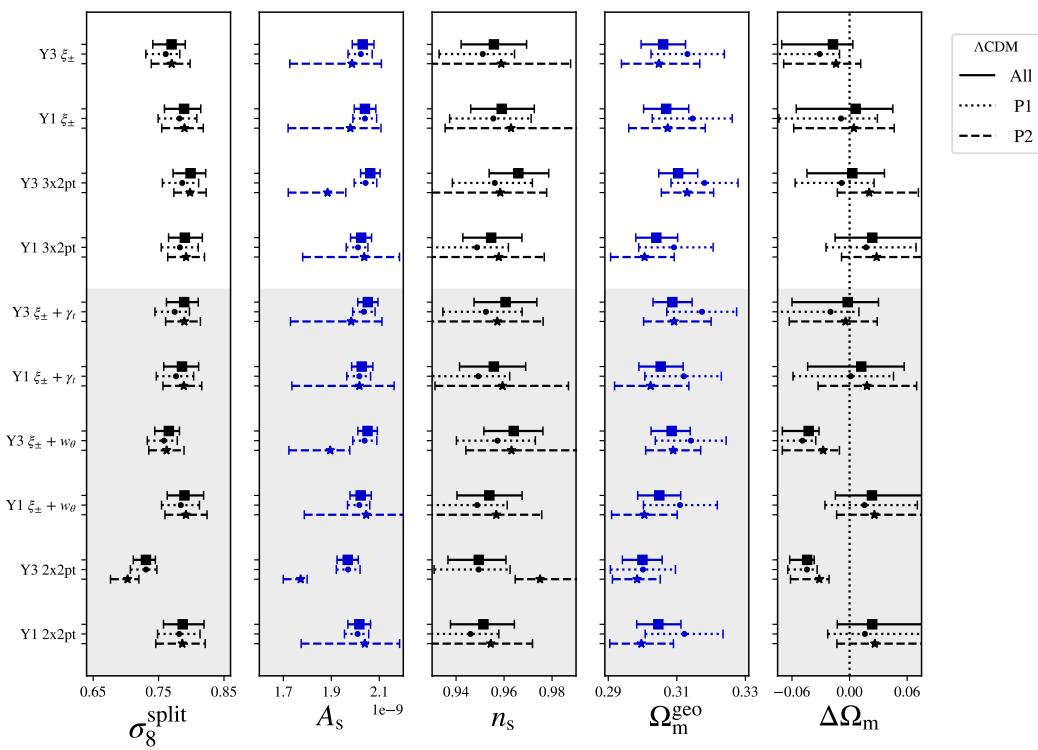
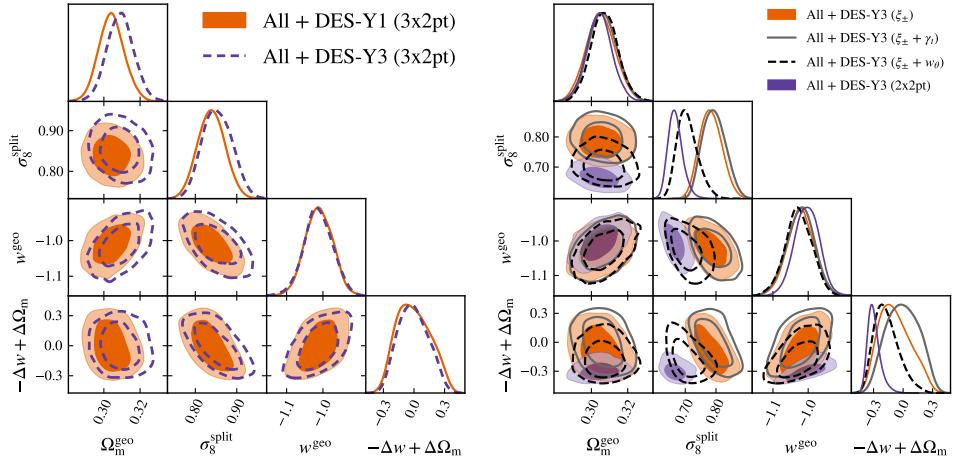


Figure 4.8: Full 1d results for ΛCDM split

additional datasets from Rubin Observatory's LSST, the Euclid mission, Dark Energy Spectroscopic Instrument, Simons Observatory, and CMB-S4 can be used to improve the constraints derived in this analysis



(a) DES-Y1 vs DES-Y3 3x2pt w CDM (b) DES-Y3 2-point correlation function posteriors.

Figure 4.9: w CDM posteriors.

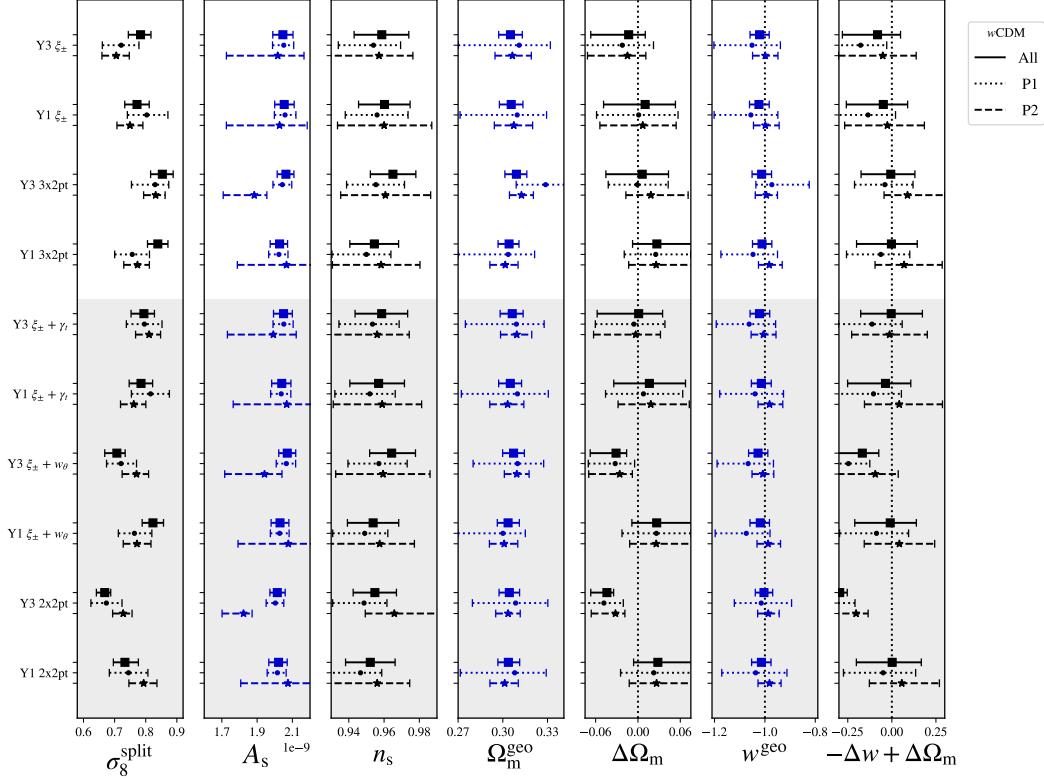
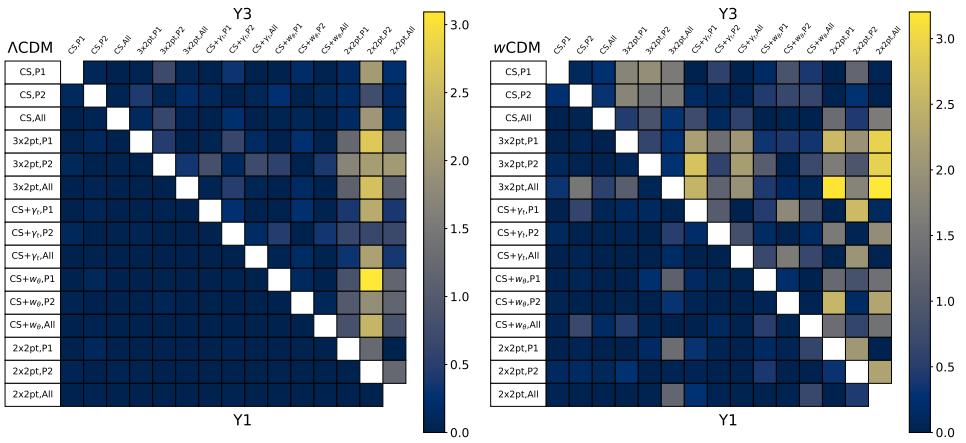


Figure 4.10: Full 1d results for Λ CDM split



(a) Internal tension in the Λ CDM (b) Internal tension in the w CDM split

Figure 4.11

Chapter 5

Data Emulators

As can be seen from the growth-geometry split analysis, a great amount of time and computing resources are needed to do a comprehensive analysis of novel cosmological models and extensions to Λ CDM. There are a growing number of proposed extension theories, and more data sets that can test them. One of the ways to improve the efficiency of data and theory analysis is to create a faster way to compute data vectors when doing MCMC sampling. In this chapter, I will describe how one can use neural networks (NN) to learn this mapping, bypassing the need for expensive programs such as Comsolike.

5.1 Neural Networks

Although the term ‘neural network’ is almost synonymous with magic in the modern world, the concept is very natural from elementary mathematics. We start with the most basic example: linear regression.

In linear regression, one considers a model of the form $y = mx + b$. The goal is to find the values of m and b that minimizes the square of the residual

$$L(y_{i,\text{truth}}, y_{i,\text{model}}) = (y_{i,\text{truth}} - y_{i,\text{model}})^2. \quad (5.1)$$

In this very special case, there is an analytic way to minimize L , but in general that is highly non-trivial. Additionally, there are much more complicated data sets one would want to model, such as data vectors in cosmological surveys. The remaining question is: how does one intelligently generalize this type of model to more complicated ones?

We have a tool that can do this already: the brain! The brain is a large system of cells called neurons. Each neuron is connected to others by a synapse, and after accumulating a large enough electric charge, the neuron will activate, sending the electric signal to a connected neuron.

Using this, lets try to construct a general model. We start with the neurons. Now we add in the synapses, the connections between neurons. In general,

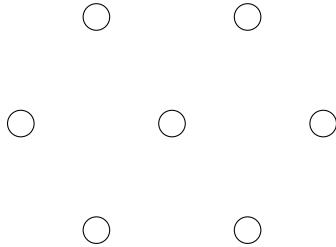


Figure 5.1: Arrangement of neurons.

not every neuron is connected to each other. It should be noted, along each synapse is an activation function. Rather than having a threshold potential like a real synapse, NNs use functions that represent the activation of the neuron. Also, each synapse has a weight that represents the strength of the connection. Next, we define which neurons receive an external signal, and

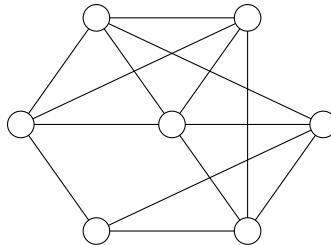


Figure 5.2: Neurons with synapses.

which give a signal to the outside. These act as our input and output. Finally, we have a choice to make in the activation function. This is the chance to greatly increase the modelling capability of NNs. As we will see in the future, the only requirement is that the function is continuous, thus we can use non-linear activation functions, allowing modelling of highly non-linear outputs. The goal is to define a loss function function that we want to minimize (in the linear regression, the loss is the square of the residual). The loss function is a function of the output of the NN, and the output is parameterized by the parameters of the NN. Thus, if $y = \phi(\alpha_i, x)$ represents the output of the

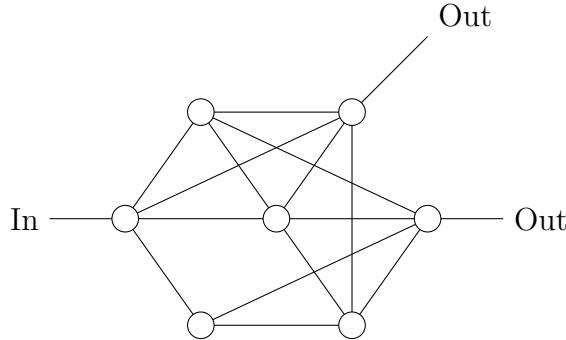


Figure 5.3: Neurons, synapses, and IO.

neural network with parameters α_i and input x , the loss function is, in fact, parameterized by the neural network as well.

$$L(y) = L(\phi(\alpha_i, x)) = \mathcal{L}(\alpha_i, x) \quad (5.2)$$

Thus, a neural network is, in simple terms, an optimization problem. The issue, however, is that neural networks generally have a large number of parameters (thousands, millions, even billions!), so this optimization is difficult in practice. Also, what may appear as a simple loss function in terms of the output can become an extremely complicated (and importantly, non-monotonic) function of α_i .

The general method of optimization is called *gradient descent*. The strategy is the following:

1. Define an initial set of parameters $\alpha_{i,\text{init}}$.
2. Split the training data into n batches
3. Generate a new set of parameters for each batch, centered around $\alpha_{i,\text{init}}$, denoted α_i^n .
4. Compute the loss for each batch.
5. Compute the gradient of the neural network.
6. Update the parameters according to the gradient.
7. Repeat until gradient is 0.

This simple algorithm has one glaring flaw, it is prone to getting stuck in local minima. Ideally, we want to search for the global minimum for the loss function. The way around this is to add some stochasticity to the gradient descent so that the network does not simply search for zero gradient. Instead, the parameters are updated probabilistically from the gradient.

The gradient is usually computed using *backpropagation*, in which the gradient is computed at each layer and the total gradient is computed according to the chain rule. This presents an issue: if the gradient is small at each layer, the gradient will be small on the order of ϵ^n , where n is the number of layers. This can lead to a misleading gradient that apparently vanishes, or a gradient that explodes to large values. This is the problem of *vanishing gradients*, and we will discuss ways to avoid this in the next section.

The last issue to mention is that, given a large number of parameters, its possible to fit the training data very well, but the model is not able to fit external data sets. This is called *overfitting*. There are many strategies to resolve this. The two we will use is dropout and L2 regularization. Dropout randomly sets a fraction of the weights to zero, meaning the NN will not be able to rely on individual neurons to get a reliable output. This ensures the full network is being used. L2 regularization penalizes the model for learning large weights, which again means the model relies on particular neurons. This is accomplished by modifying the loss function by adding the L2 norm of the network.

$$L(y) \mapsto L(y) + \sum_i \alpha_i^2 \quad (5.3)$$

In summary, a NN is a directed graph, where each edge has a weight and an activation function, and each node has a value that it passes along to the edge. The inputs are propagated to the output, and the goal is to minimize the loss of the output. Its important to pick a smart loss function depending on ones needs from the NN and a smart activation function depending on the values of each node. The loss function is minimized using stochastic gradient descent. It is important to also implement strategies to avoid overfitting on the training data by penalizing the model for relying to heavily on certain regions of the graph.

5.2 Architecture Choices

As discussed before, the weights, activation functions, and connections are choices we can make for a neural network. Together, they comprise the

architecture. In this section we will discuss the architecture choices we have made.

Before discussing the graph architecture, we want to highlight the choice of activation function. Before giving any data to the neural network, we do some preprocessing to the data. The first thing we do is normalize the input and output. The input is the cosmological parameters, which we process each parameter to follow a normal distribution.

$$x^i = \frac{\theta^i - \bar{\theta}^i}{\sigma_i} \quad (5.4)$$

The output is the cosmic shear data vector. We preprocess this by diagonalizing and normalizing each component.

$$y^i = \frac{(P^{-1}d)^i}{\sqrt{(PCP^{-1})^{ii}}} \quad (5.5)$$

Where C is the covariance of the data and P the change of basis matrix to the eigenbasis of C . This generally means we will have a lot of negative parameters, and the input and output are both symmetric about 0, thus we choose to use an antisymmetric activation function so that the symmetry around 0 can be preserved. In our case, we use $\tanh(x)$. Additionally, all of these NNs are feedforward sequential models. Sequential means we separate the neurons into layers, and feedforward means the data moves only from layer n to layer $n + 1$ and never backwards.

5.2.1 Multi-Layer Perceptron

The first graph architecture we study is the multi-layer perceptron (the name is adopted from convolutional neural networks ‘perceiving’ image features). In this architecture a node in a given layer is connected to every node in the following layer, so this network is called ‘simply connected’ (figure 5.4).

5.2.2 Residual Network

As mentioned above, there is an issue of vanishing gradients with deep neural networks. One way we can avoid this is by changing the quantity we want to model. Note that, if the model weights approach 0, gradients can vanish, but we can avoid this by adding the input to the output (figure 5.5). Suppose we have a block of layers parameterized by α_i , and its input is from another layer parameterized by β_i so that $w = \chi(\beta_i, x)$ for an input x . Then the block of

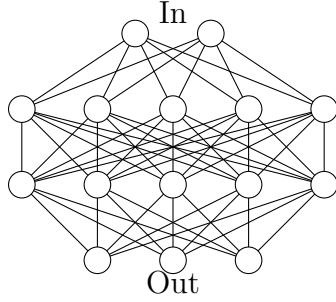


Figure 5.4: A multi-layer perceptron.

layers, whose map is denoted ϕ , will be

$$\phi(\alpha_i, \chi(\beta_i, x)) = \psi(\alpha_i, \beta_i, x) + \chi(\beta_i, x) \quad (5.6)$$

This section of the model is going to attempt to learn

$$\psi(\alpha_i, \beta_i, x) - \chi(\beta_i, x) \quad (5.7)$$

This allows gradients with respect to the parameters β_i to propagate to the output without the multiplication by the gradients of α_i , which usually solves the vanishing gradient problem. This is called a *residual neural network* (ResNet).

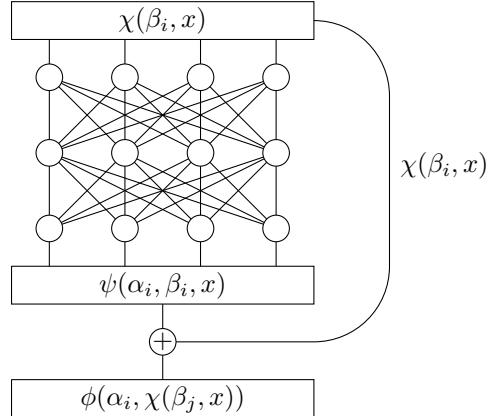


Figure 5.5: A residual block.

5.2.3 Bottlenecked Residual Network

The ResNet allows us to create much deeper networks, but the number of parameters stays the same as the linear model, making ResNets difficult to train (and especially the minimum of two linear layers). We can make a small modification to the ResNet that can resolve this, where there is an additional layer and the number of dimensions reduces in the middle (figure 5.6). This forces the neural network to prioritize certain features and ignore ones that are not as necessary. Due to the shape this architecture makes, its referred to as a *bottlenecked residual network* (ResBottle) (figure 5.6). In the examples

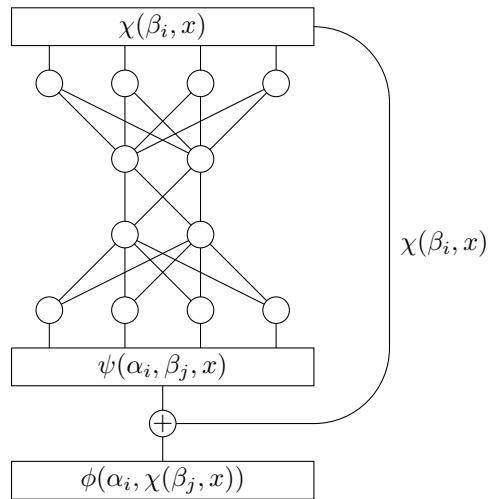


Figure 5.6: A bottlenecked residual block.

drawn (figures 5.5 and 5.6), there are 40 parameters in the ResNet, but only 28 parameters in the ResBottle. This allows us to either simplify the model, add more layers, or make the layers much wider without blowing up the number of parameters that need to be trained.

5.2.4 Attention

This will be the most complicated model to test. If we look at the structure of the data vector, there is a certain order to the data and correlation between the components. The data vectors is separated into redshift bins, and within each redshift bin are bins representing different scales. Thus, a NN that can learn sequences may be useful. Fortunately, there are NNs we encounter everyday that use sequences: chatbots! The basis for this is usually attention networks [?]. These work by examining dot products between vectors so that the output is directly related to the angle between the two vectors.

suppose we are given a sequence of input vectors, $x = (x_1, x_2, \dots, x_n)$, then we can construct dot products of the x_i and use them as coefficients for the output.

$$y_k = (x_i \cdot x_j)x_k \quad (5.8)$$

We can introduce generalize the dot product using linear transformations. That is, we can introduce the transformation matrices W_Q , W_K , W_V . Let $q_i = W_Q x_i$, $k_i = W_K x_i$, and $v_i = W_V x_i$. Then we can also write a dot product as

$$y_k = (q_i^T k_j)v_k \quad (5.9)$$

Finally, we can introduce a non-linearity, which also keeps the outputs sequence y normalized. By introducing the *softmax* function.

$$y_k = \text{softmax}(q_i^T k_j / \sqrt{d})v_k \quad (5.10)$$

where the softmax function is an exponential normalization given by

$$\text{softmax}(W)_i = \frac{e^{w_i}}{e^{\sum_i w_i}} \quad (5.11)$$

and d is the embedding dimension (i.e $W_K \in \mathbb{R}^{d_x \times d}$). We allow the linear matrices W_Q , W_K , and W_V to contain learnable components. We then pass each output y_i to its own mlp. We aslo keep the skip connections describe in the residual network section. The attention together with the MLPs and skip connections is called a transformer block, drawn in figure 5.7.

5.3 Results

We present the results in two sections: the first will be about the training and testing results, and the second comparing between a cocoa chain and an emulated chain. For training and testing, we use the "train, validate, test" approach. We generate three completely independent chains, one is only for training the model, and is a high temperature chain with data well beyond the expected range. The second is used to validate that the model is not overfitting as the training proceeds. The last set is another high temperature chain (but not as high as the training) which is used to test the model as if it were being used as normal. The testing involves reintroducing specific analysis choices such as the scale cut.

When deciding on which model to use, we want to look at several aspects:

- The number of parameters. The more parameters we have, the longer it takes to train, and the longer it takes to evaluate. This will make it difficult for groups without top-of-the-line equipment.
- The $\Delta\chi^2$ between the emulator and accepted computational methods. Many experiments have a χ^2 requirement for their emulators; for example DES requires $\Delta\chi^2 < 0.3$ between any approximations and Cosmolike.
- Generalizability. As we increase the volume covered in the parameter space, which models are able to keep up with the growing modelling volume? Which models can handle more input parameters (which many models need, such as early dark energy with ~ 10 extra parameters).

These get analyzed on the test data set. Because of the second point above, we define the loss function as the difference in χ^2 between the emulator and cosmolike. This allows us to capture the covariances in the data.

5.3.1 Training and Testing

We train the emulator by using a rough Gaussian approximation on the cosmological parameters. The benefit is that running MCMC chains on a simple gaussian without the need for cosmolike is extremely fast. Given the samples, the data vector calculation is trivially parallelizable, so the entire process is significantly faster than running an MCMC in cosmolike.

First, we examine the training and validation loss at each epoch. The training loss is the average loss of all batches in an epoch. Figure 5.8 displays the training loss of a ResNet model with 3 ResNet blocks and 256 neurons in each hidden layer. In this case, we can see a consistent decline in both the training and validation loss. Additionally, we can see the points where the learning rate decreases (see epoch 50 and epoch 85 in figure 5.8) which occur when validation loss plateaus, following the expected behaviour. This demonstrates that we are likely not overfitting the training data. Using a trained model, we evaluate $\Delta\chi^2$ on the test data, a $T = 8$ gaussian approximated chain (figure 5.9). As we can see, our model is able to generalize well to a $T = 8$ chain.

Additionally, we can test how high of a temperature we can train on and still get satisfactory performance.

5.4 Application to Tension Calibration

As mentioned in section ??, tension metrics require calibration between experiments. In this section, I will apply the use of emulators to do a first study of tension metric error. By shifting the parameters and generating noise realizations on the data, we can get a sense of how the error in the data propagates to error in the tension. This will provide additional information as to which tension metrics perform best, as well as adding context to how tension may vary given new measurements. Additionally, given that $n_\sigma = 0$ means no tension, the error are tension metrics will allow us to gauge how reasonable it may be for the tension to be resolved by additional identical measurements.

I will be shifting the LSST fiducial cosmology to $\pm 5\sigma$ in both the σ_8 and Ω_m directions (figure ??). At each of these shifts, I will compute 1000 noise realizations on LSST cosmic shear using the data covariance. Finally, the tension will be computed using the methods described in section ??.

I chose to use the ResNet model with 3 ResBlocks and 256 neurons in each hidden layer. This model provides good stability in the predictions, and is sufficiently small that running chains on the noise realizations is tractable.

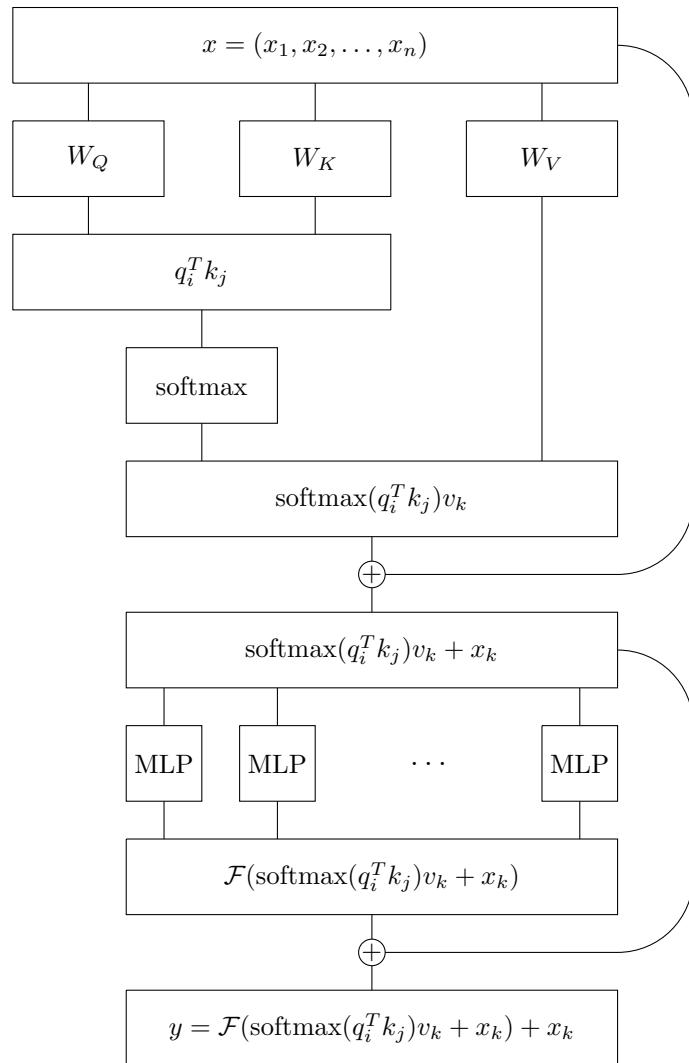


Figure 5.7: A transformer block

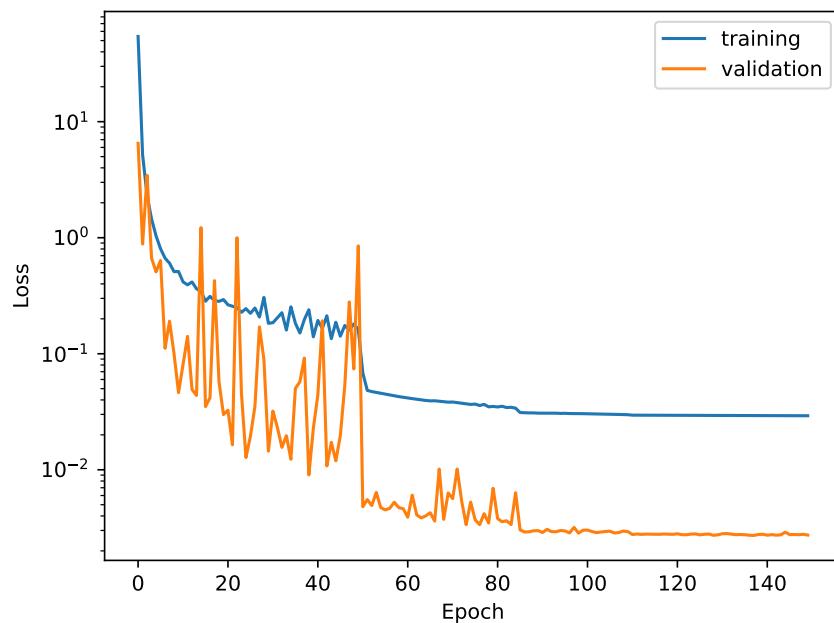


Figure 5.8: Loss of a ResNet model with 3 ResNet blocks and 256 neurons in each layer.

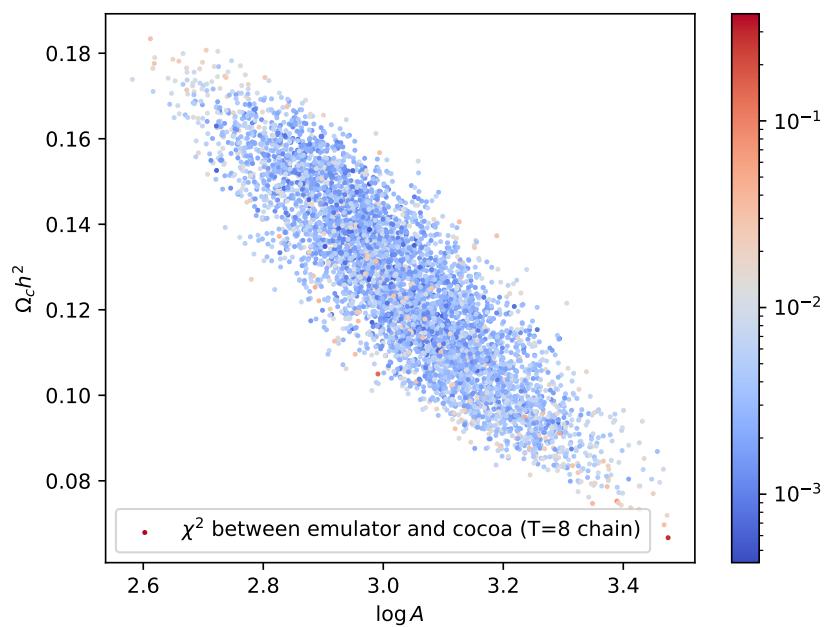
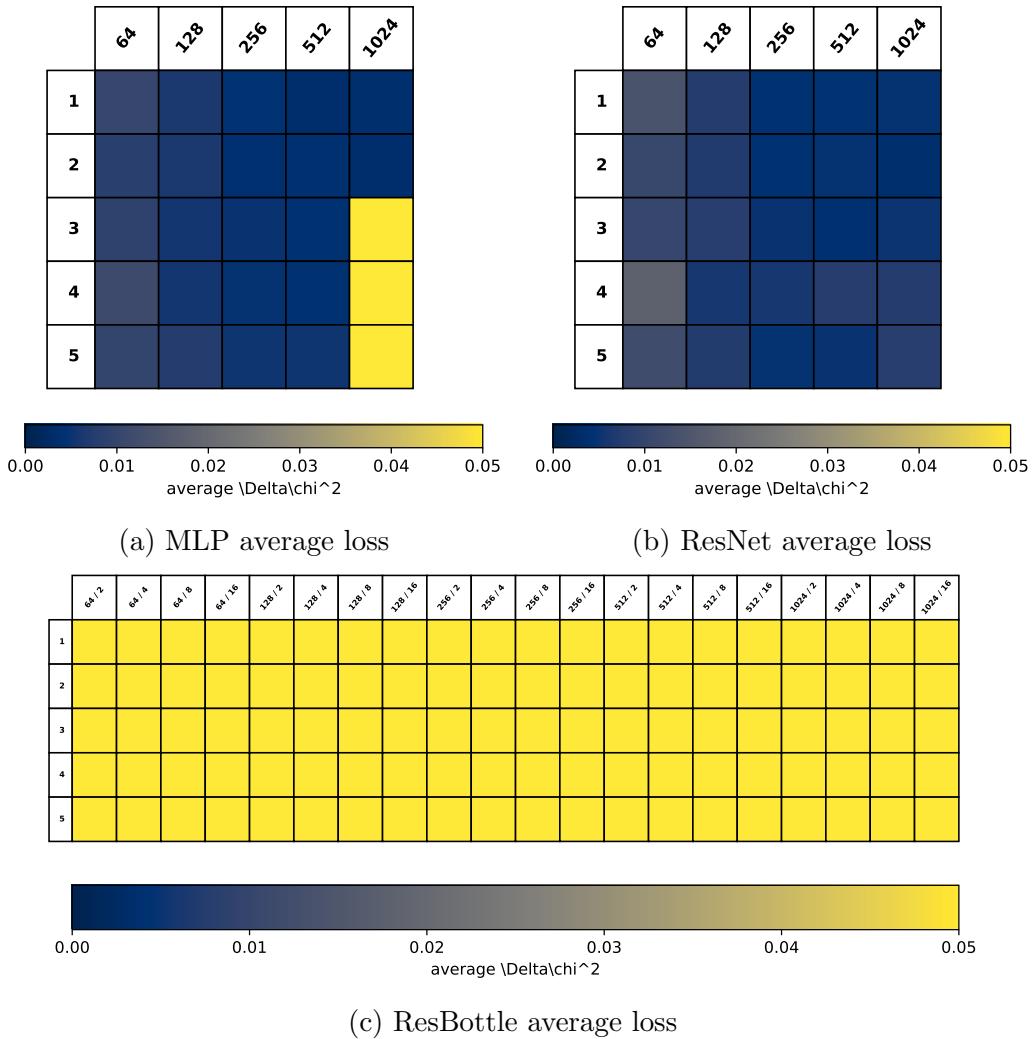


Figure 5.9: Testing loss on a $T = 8$ chain.



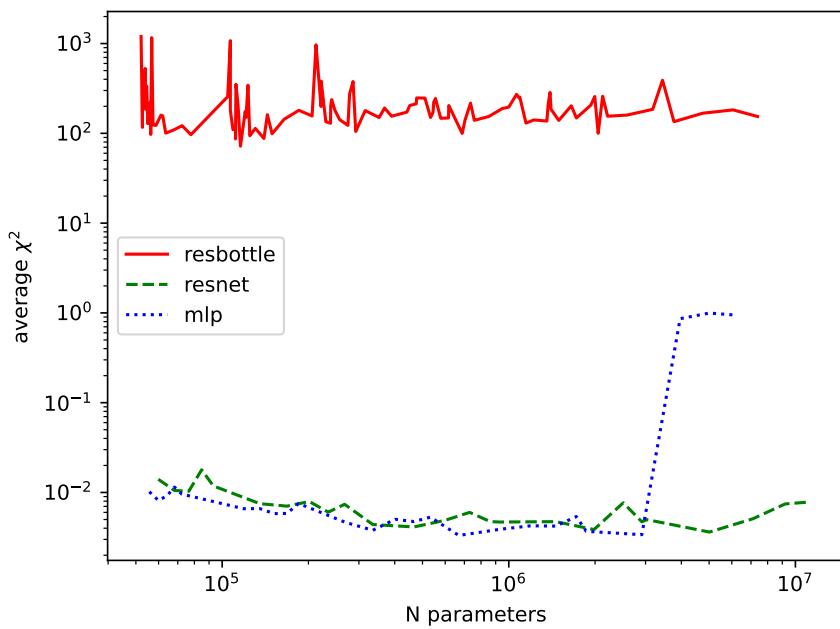


Figure 5.11: Average χ^2 as a function of the number of parameters for each architecture.

Bibliography

- [1] ESA, NASA, James Webb Space Telescope. Webb uncovers new details in pandora's cluste. URL <https://esawebb.org/images/weic2305a/>.
- [2] ESA and Planck collaboration. Planck cmb. URL https://www.esa.int/ESA_Multimedia/Images/2013/03/Planck_CMB.
- [3] NASA, ESA, and J. Lotz and the HFF Team. Hubble lensing, release id: 2017-20. URL <https://hubblesite.org/contents/news-releases/2017/news-2017-20.html>.
- [4] P. Lemos, M. Raveri, A. Campos, Y. Park, C. Chang, N. Weaverdyck, D. Huterer, A. R. Liddle, J. Blazek, R. Cawthon, A. Choi, J. DeRose, S. Dodelson, C. Doux, M. Gatti, D. Gruen, I. Harrison, E. Krause, O. Lahav, N. MacCrann, J. Muir, J. Prat, M. M. Rau, R. P. Rollins, S. Samuroff, J. Zuntz, W. G. Hartley, B. Hoyle, I. Sevilla-Noarbe, M. A. Troxel, M. Aguena, S. Allam, J. Annis, S. Avila, D. Bacon, G. M. Bernstein, E. Bertin, D. Brooks, D. L. Burke, A. Carnero Rosell, M. Carrasco Kind, J. Carretero, C. Conselice, M. Costanzi, M. Crocce, M. E. S. Pereira, J. De Vicente, S. Desai, H. T. Diehl, P. Doel, K. Eckert, T. F. Eifler, J. Elvin-Poole, S. Everett, A. E. Evrard, I. Ferrero, A. Ferté, B. Flaugher, P. Fosalba, J. Frieman, J. García-Bellido, E. Gaztanaga, D. W. Gerdes, T. Giannantonio, R. A. Gruendl, J. Gschwend, G. Gutierrez, S. R. Hinton, D. L. Hollowood, K. Honscheid, E. M. Huff, D. J. James, M. Jarvis, M. Lima, M. A. G. Maia, M. March, J. L. Marshall, P. Martini, P. Melchior, F. Menanteau, R. Miquel, J. J. Mohr, R. Morgan, J. Myles, R. L. C. Ogando, A. Palmese, S. Pandey, F. Paz-Chinchón, A. A. Plazas, M. Rodriguez-Monroy, A. Roodman, E. Sanchez, V. Scarpine, M. Schubnell, L. F. Secco, S. Serrano, M. Smith, M. Soares-Santos, E. Suchyta, M. E. C. Swanson, G. Tarle, D. Thomas, C. To, T. N. Varga, J. Weller, and W. Wester. Assessing tension metrics with Dark Energy Survey and

- Planck data. *Monthly Notices of the Royal Astronomical Society*, 505(4): 6179–6194, July 2021. ISSN 0035-8711, 1365-2966. doi: 10.1093/mnras/stab1670. URL <http://arxiv.org/abs/2012.09554>. arXiv:2012.09554 [astro-ph].
- [5] Baez, John and Muniain, Javier. *Gauge Fields, Knots, and Gravity*, volume Vol 4 of *Series on Knots and Everything*. World Scientific, 1994. ISBN 981-02-1729-3.
- [6] Scott Dodelson and Fabian Schmidt. *Modern Cosmology*. Academic Press, 2021. ISBN 978-0-12-815948-4.
- [7] E. Krause, X. Fang, S. Pandey, L. F. Secco, O. Alves, H. Huang, J. Blazek, J. Prat, J. Zuntz, T. F. Eifler, N. MacCrann, J. DeRose, M. Crocce, A. Porredon, B. Jain, M. A. Troxel, S. Dodelson, D. Huterer, A. R. Liddle, C. D. Leonard, A. Amon, A. Chen, J. Elvin-Poole, A. Ferté, J. Muir, Y. Park, S. Samuroff, A. Brandao-Souza, N. Weaverdyck, G. Zacharegkas, R. Rosenfeld, A. Campos, P. Chintalapati, A. Choi, E. Di Valentino, C. Doux, K. Herner, P. Lemos, J. Mena-Fernández, Y. Omori, M. Paterno, M. Rodriguez-Monroy, P. Rogozenski, R. P. Rollins, A. Troja, I. Tutusaus, R. H. Wechsler, T. M. C. Abbott, M. Aguena, S. Allam, F. Andrade-Oliveira, J. Annis, D. Bacon, E. Baxter, K. Bechtol, G. M. Bernstein, D. Brooks, E. Buckley-Geer, D. L. Burke, A. Carnero Rosell, M. Carrasco Kind, J. Carretero, F. J. Castander, R. Cawthon, C. Chang, M. Costanzi, L. N. da Costa, M. E. S. Pereira, J. De Vicente, S. Desai, H. T. Diehl, P. Doel, S. Everett, A. E. Evrard, I. Ferrero, B. Flaugher, P. Fosalba, J. Frieman, J. García-Bellido, E. Gaztanaga, D. W. Gerdes, T. Giannantonio, D. Gruen, R. A. Gruendl, J. Gschwend, G. Gutierrez, W. G. Hartley, S. R. Hinton, D. L. Hollowood, K. Honscheid, B. Hoyle, E. M. Huff, D. J. James, K. Kuehn, N. Kuropatkin, O. Lahav, M. Lima, M. A. G. Maia, J. L. Marshall, P. Martini, P. Melchior, F. Menanteau, R. Miquel, J. J. Mohr, R. Morgan, J. Myles, A. Palmese, F. Paz-Chinchón, D. Petracick, A. Pieres, A. A. Plazas Malagón, E. Sanchez, V. Scarpine, M. Schubnell, S. Serrano, I. Sevilla-Noarbe, M. Smith, M. Soares-Santos, E. Suchyta, G. Tarle, D. Thomas, C. To, T. N. Varga, and J. Weller. Dark Energy Survey Year 3 Results: Multi-Probe Modeling Strategy and Validation, May 2021. URL <http://arxiv.org/abs/2105.13548>. arXiv:2105.13548 [astro-ph].
- [8] Jonathan Blazek, Niall MacCrann, M. A. Troxel, and Xiao Fang. Beyond linear galaxy alignments. *Phys. Rev. D*, 100(10):103506, November 2019.

ISSN 2470-0010, 2470-0029. doi: 10.1103/PhysRevD.100.103506. URL <http://arxiv.org/abs/1708.09247>. arXiv:1708.09247 [astro-ph].

- [9] Niall MacCrann, Jonathan Blazek, Bhuvnesh Jain, and Elisabeth Krause. Controlling and leveraging small-scale information in tomographic galaxy–galaxy lensing. *Monthly Notices of the Royal Astronomical Society*, 491(4):5498–5509, February 2020. ISSN 0035-8711, 1365-2966. doi: 10.1093/mnras/stz2761. URL <https://academic.oup.com/mnras/article/491/4/5498/5586592>.
- [10] Marco Raveri and Cyrille Doux. Non-Gaussian estimates of tensions in cosmological parameters. *Phys. Rev. D*, 104(4):043504, August 2021. ISSN 2470-0010, 2470-0029. doi: 10.1103/PhysRevD.104.043504. URL <http://arxiv.org/abs/2105.03324>. arXiv:2105.03324 [astro-ph].
- [11] Justin L Tobias. The Metropolis-Hastings Algorithm.
- [12] Boyang Zhao. Bayesian inference using Markov Chain Monte Carlo with Python (from scratch and with PyMC3), January 2021. URL <https://bo yangzhao.github.io/posts/mcmc-bayesian-inference>.
- [13] Jennifer Helsby. Monte Carlo Markov Chains: A Brief Introduction and Implementation. (5).
- [14] Mathieu Germain, Karol Gregor, Iain Murray, and Hugo Larochelle. MADE: Masked Autoencoder for Distribution Estimation, June 2015. URL <http://arxiv.org/abs/1502.03509>. arXiv:1502.03509 [cs, stat].
- [15] George Papamakarios, Theo Pavlakou, and Iain Murray. Masked Autoregressive Flow for Density Estimation, June 2018. URL <http://arxiv.org/abs/1705.07057>. arXiv:1705.07057 [cs, stat].
- [16] J. Muir, E. Baxter, V. Miranda, C. Doux, A. Ferté, C. D. Leonard, D. Huterer, B. Jain, P. Lemos, M. Raveri, S. Nadathur, A. Campos, A. Chen, S. Dodelson, J. Elvin-Poole, S. Lee, L. F. Secco, M. A. Troxel, N. Weaverdyck, J. Zuntz, D. Brout, A. Choi, M. Crocce, T. M. Davis, D. Gruen, E. Krause, C. Lidman, N. MacCrann, A. Möller, J. Prat, A. J. Ross, M. Sako, S. Samuroff, C. Sánchez, D. Scolnic, B. Zhang, T. M. C. Abbott, M. Aguena, S. Allam, J. Annis, S. Avila, D. Bacon, E. Bertin, S. Bhargava, S. L. Bridle, D. Brooks, D. L. Burke, A. Carnero Rosell, M. Carrasco Kind, J. Carretero, R. Cawthon, M. Costanzi, L. N. da Costa, M. E. S. Pereira, S. Desai, H. T. Diehl, J. P. Dietrich, P. Doel, J. Estrada,

- S. Everett, A. E. Evrard, I. Ferrero, B. Flaugher, J. Frieman, J. García-Bellido, T. Giannantonio, R. A. Gruendl, J. Gschwend, G. Gutierrez, S. R. Hinton, D. L. Hollowood, K. Honscheid, B. Hoyle, D. J. James, T. Jeltema, K. Kuehn, N. Kuropatkin, O. Lahav, M. Lima, M. A. G. Maia, F. Menanteau, R. Miquel, R. Morgan, J. Myles, A. Palmese, F. Paz-Chinchón, A. A. Plazas, A. K. Romer, A. Roodman, E. Sanchez, V. Scarpine, S. Serrano, I. Sevilla-Noarbe, M. Smith, E. Suchyta, M. E. C. Swanson, G. Tarle, D. Thomas, C. To, D. L. Tucker, T. N. Varga, J. Weller, and R. D. Wilkinson. DES Y1 results: Splitting growth and geometry to test Λ CDM. *Phys. Rev. D*, 103(2):023528, January 2021. ISSN 2470-0010, 2470-0029. doi: 10.1103/PhysRevD.103.023528. URL <http://arxiv.org/abs/2010.05924>. arXiv:2010.05924 [astro-ph].
- [17] Kunhao Zhong, Evan Saraivanov, Vivian Miranda, Jiachuan Xu, Tim Eifler, and Elisabeth Krause. Growth and Geometry Split in Light of the DES-Y3 Survey, January 2023. URL <http://arxiv.org/abs/2301.03694>. arXiv:2301.03694 [astro-ph].