Accelerating Cosmological Analysis Using Neural Networks

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

May 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, Deparment of Physics and Astronomy, Yang Institue for Theoretical Physics

Anja von der Linden – Chairperson of Defense
Associate Professor, Department of Physics and Astronomy

Navid Vafaei-Najafabadi
Assistant Professor, Department of Physics and Astronomy

This thesis is accepted by the Graduate School.

Nicole S. Sampson
Dean of the Graduate School

Abstract of the Thesis

Accelerating Cosmological Analysis Using Neural Networks

by

Evan Saraivanov

Master of Arts

in

Physics

Stony Brook University

2023

Despite many efforts, no cosmological model has been proposed that can lower the H_0 and S_8 tension and also agree with ongoing experiments. Fortunately, however, we are approaching a great age for cosmology. In the near future, many telescopes will begin to operate, taking more detailed data than ever before. Current methods of analyzing models on real data are slow, taking multiple days or even weeks to generate the necessary data. With the number of proposed models, analyzing each of them with traditional means is futile. This thesis presents one model independent extension to the accepted model of cosmology, ΛCDM, which splits the matter density into two components. This highlights the traditional analysis pipeline as well as presents a consistency check within cosmological analysis. This thesis then presents novel methods for using neural networks to learn the time-consuming parts of this pipeline, greatly accelerating the analysis of cosmological models. Finally, this new neural network design is applied to computing tension metrics between the upcoming LSST experiment and Planck 2018

data. Tension metrics are frequentist estimators, and thus have uncertainty. Using traditional means, the computation of these errors is intractable, but using the novel neural network the errors can be estimated in a matter of weeks.

Contents

List of Figures viii

Acknowledgements x

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 9
2.1 Weak Lensing . . . 9
2.1.1 The Linear Matter Power Spectrum . . . 10
2.1.2 Galaxy Density . . . 11
2.1.3 Galaxy Shear . . . 12
2.1.4 Weak Lensing Correlation Functions . . . 15
2.1.5 Systematics . . . 16
2.2 CMB . . . 19
2.2.1 TT Power Spectrum . . . 20
2.2.2 TE and EE Power Spectra . . . 21
2.3 Tensions Between Weak Lensing and the CMB . . . 23

3 Computing Techniques 26
3.1 Markov Chain Monte Carlo . . . 26
3.1.1 The Metropolis-Hastings Algorithm . . . 27
3.2 Tension Metrics . . . 28
3.2.1 Parameter-Based Metrics . . . 29
3.2.2 Evidence-Based Metrics . . . 30
3.2.3 Interpreting the Results . . . 31

3.2.4 DES Versus Planck Results . . . 32
3.3 Normalizing Flows . . . 33

4 Growth-Geometry Split in DES-Y3 36
4.1 Split Matter Power Spectrum . . . 36
4.2 Data and Analysis Pipeline . . . 38
4.2.1 DES Parameter Priors . . . 38
4.2.2 Software Pipeline . . . 40
4.2.3 Pipeline Validation on Synthetic Data . . . 41
4.3 Results . . . 42
4.3.1 ΛCDM . . . 42
4.3.2 wCDM . . . 44
4.3.3 Tension Analysis . . . 44
4.4 Conclusion . . . 45

5 Data Emulators 49
5.1 Neural Networks . . . 49
5.2 Architecture Choices . . . 53
5.2.1 Multi-Layer Perceptron . . . 54
5.2.2 Residual Network . . . 54
5.2.3 Bottlenecked Residual Network . . . 54
5.2.4 Attention . . . 56
5.3 Results . . . 56
5.3.1 Training and Testing . . . 58
5.4 Application to Tension Calibration . . . 59
5.5 Conclusion . . . 63

Bibliography 64

List of Figures

1.1 Observer in free-fall, with particles experiencing tidal forces within the observer's 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 temperature map, emphasizing the scale of anisotropies [2]. 7

2.1 Image of weak lensing taken by the Hubble Space Telescope [3]. 9
2.2 Planck TT power spectrum [4]. 20
2.3 Planck TE (top) and EE (bottom) power spectra [4]. 22
2.4 Measurements of S_8 from different experiments. 24
2.5 Measurements of H_0 from different experiments. 25

3.1 Result of tension metrics between Planck and DES [5]. A-priori tension is the Gaussian approximation described in equation 3.20. 32
3.2 An example non-Gaussian distribution (non-zero skewness). . 34
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 $\lambda = 1$. 35

4.1 Non-linear power spectrum from Euclid emulator. When either the geometry or growth parameters are varied, the other is kept fixed at $\Omega^X = 0.3$. 37
4.2 Effect of changing growth and geometry parameters on $\sigma_8(z)$. 38
4.3 Relative difference between Euclid emulator and Halofit. . . . 41

4.4 Cosmic shear, 2×2pt combinations, and 3×2pt posteriors for DES-Y3 with the ALL prior. 42
4.5 Comparison of DES-Y1 and DES-Y3 posteriors. 43
4.6 Split posteriors for DES-Y3 3×2pt and ALL prior. 44
4.7 DES-Y1 and DES-Y3 posteriors for 2-point correlation functions. 45
4.8 Full 1d results for ΛCDM split 46
4.9 wCDM posteriors. 47
4.10 Full 1d results for ΛCDM split 47
4.11 Summary of internal tensions between each probe in the ΛCDM and wCDM models. 48

5.1 Arrangement of neurons. 50
5.2 Neurons with synapses. 50
5.3 Neurons, synapses, and the input/output. 51
5.4 A multi-layer perceptron. 54
5.5 A residual block. 55
5.6 A bottlenecked residual block. 55
5.7 A transformer block 57
5.8 Average $\Delta\chi^2$ as a function of the number of parameters for each architecture. 59
5.9 Average $\Delta\chi^2$ of a $T = 64$ chain using emulators trained on a $T = 128$ chain. 60
5.10 Marginalized distribution of $\Delta\chi^2$ of a $T = 64$ chain using the model in figure 5.11 trained on a $T = 128$ chain. 60
5.11 Full NN architecture used for training on high-temperature chains. Comprised of three ResBlocks of size 256, and a transformer block with 8 heads, each with a size of 128. 61
5.12 Parameter-based tension metrics with error. 62

Acknowledgements

I would like to give my utmost gratitude to my family, friends, and peers, for their great assistance in my studies. My parents, Carrie and Miroslav Saraivanov, have always supported me through any decision in my life, and provided me with the emotional support one needs to achieve such a difficult and stressful feat. I am grateful for their aid in editing and revising my thesis, making sure it is formatted well and it is readable.

My advisor, professor Vivian Miranda, has been an outstanding source of knowledge and ideas. The other masters students I have worked with during my master's degree, Kunhao Zhong and Xiao Huangfei, have also been great for helping me to get past difficulties in my research, as well as sharing their great wealth of knowledge in cosmology. Additionally, I would like to thank Jonathan Gordon and Joshua Kable, our group's PhD student and post-doctoral researcher, for their aid in revising my thesis. Before joining this group, my background in cosmology was minimal and without each of these people, this thesis would not be possible.

Last, but far from least, I would like to thank my girlfriend, Xueting Deng. Not only did she help proofread my thesis, she helped me deal with death at the tail end of this process. Without her, I do not know if I would be able to finish this thesis during such a difficult time of my life.

Although I am a man of few words, I do not want to understate the support I received from others. To each person mentioned, I am eternally grateful for the experience I have had at Stony Brook University, in the cosmology group, and while writing.

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 are felt instantaneously everywhere. 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. Nevertheless, 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 much lower than the speed of light.

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 pointing toward the center. With nothing in view, this person would not be able to determine they were in a gravitational field rather than just linearly accelerating. Suppose this observer has point particles nearby, one to each side, one above, and one below (figure 1.1). The observer will see the particles below and above them move farther away

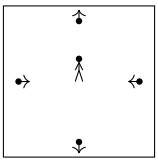


Figure 1.1: Observer in free-fall, with particles experiencing tidal forces within the observer's reference frame.

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 *tidal forces*, as they also give rise to the tides 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 the speed of light $c = 1$ (which is done in the rest of this thesis), the metric is given by

$$g = \mathrm{diag}(-1, 1, 1, 1)\,. \tag{1.1}$$

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.

$$-d\tau^2 = -dt^2 + dx^2 + dy^2 + dz^2 \tag{1.2}$$

Proper time represents the amount of time 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 in the spatial displacement as seen by equation 1.2. The set of transformations that leave τ invariant are the *Lorentz transformations*.

Suppose there are 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

$$\frac{d^2x^\mu}{d\tau^2} = 0\,. \tag{1.3}$$

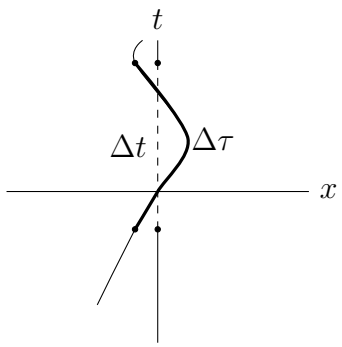


Figure 1.2: Definition of proper time (bold line) vs. coordinate time (dashed line).

One can look at how this changes for the observer in a non-inertial frame:

$$\frac{d^2x'^\lambda}{d\tau^2} + \underbrace{\frac{\partial x'^\lambda}{\partial x^\rho}\frac{\partial^2 x^\rho}{\partial x'^\mu \partial x'^\nu}}_{\equiv \Gamma^\lambda_{\mu\nu}} \frac{dx'^\mu}{d\tau}\frac{dx'^\nu}{d\tau} = 0\,. \tag{1.4}$$

Equation 1.4 looks suspiciously like the geodesic equation from differential geometry. The exception is that the 'Chirstoffel symbols', as defined here, are not the most general form of the Christoffel symbols. This gives 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, which results from curvature of the space itself. Special relativity promotes the classical $\mathbb{R}^3$ space to $\mathbb{M}^4$. However, the metric must allow for curved space, not simply Minkowski space. Instead of $\mathbb{M}^4$, let space be an arbitrary semi-Riemannian 4-manifold M with a metric with one negative eigenvalue.[1]

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 at p, denoted T_pM. The

[1] In very loose terms, a manifold is a locally Euclidean space, which is a sufficient definition for this application.

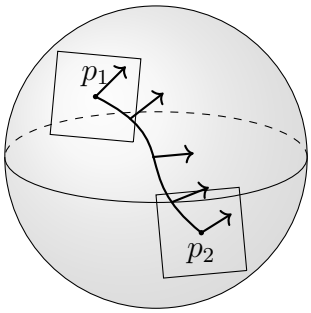


Figure 1.3: Example of parallel transport along a connection on a sphere.

generalization of the derivative is called a *connection* [6], which tells one how to connect the tangent spaces at different points infinitesimally (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 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

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

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

$$\nabla_\mu \partial_\nu = \Gamma^\rho_{\mu\nu} \partial_\rho\,. \tag{1.6}$$

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

$$R^\sigma_{\mu\nu\rho} = \nabla_\mu \Gamma^\sigma_{\nu\rho} - \nabla_\nu \Gamma^\sigma_{\mu\rho}\,. \tag{1.8}$$

To emphasize the elegance of Einstein's equation, it is convenient to write the Riemann curvature as a matrix valued differential 2-form F and the Christoffel symbols as a matrix valued differential 1-form[2] $\mathcal{G}$ [6]. With this notion,

[2]Differential forms are antisymmetric contravariant tensors.

it is easy to see the Bianchi identity for F with respect to the Levi-Civita connection is

$$d_\nabla F = \nabla_{[\tau} R^\sigma_{\mu\nu]\rho} = 0\,, \tag{1.9}$$

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

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

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,

$$\nabla^\mu T_{\mu\nu} = 0\,. \tag{1.11}$$

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

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi\kappa T_{\mu\nu}\,. \tag{1.12}$$

1.3 Cosmology

Einstein's equation includes the constant Λ, which is 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 ΛCDM, 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 [7].

When discussing ΛCDM (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 clearly does not follow them. If one looks into the night sky, one 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, one arrives at the Friedmann-Lemaître-Robertson-Walker (FLRW) metric in the Newtonian gauge,

$$g = -dt^2 + a(t)^2 ds^2 \,, \tag{1.13}$$

where s represents spatial coordinates. Working in the Newtonian gauge is chosen to best illuminate the physics discussed throughout this thesis. To perturb this metric around a homogeneous 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

$$g = -(1 + 2\Psi)dt^2 + a^2(1 - 2\Phi)ds^2 \,. \tag{1.14}$$

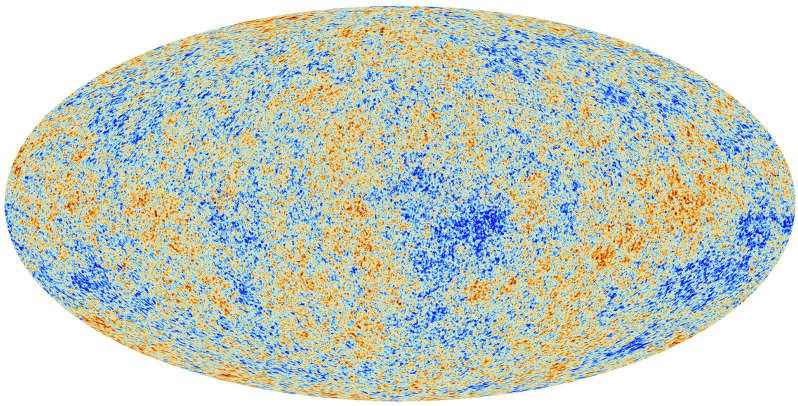


Figure 1.5: CMB temperature map, emphasizing the scale of anisotropies [2].

By computing the Christoffel symbols, one finds

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

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:

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

One can 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 the total density ρ, one finds

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

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

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

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

densities today by

$$\Omega_{i,0} = \frac{\rho_{i,0}}{\rho_{\text{crit}}}\,, \tag{1.19}$$

the relation becomes

$$\Omega + \Omega_k = 1\,, \tag{1.20}$$

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

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P)\,. \tag{1.21}$$

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. This can be avoided by defining the *comoving distance*, in which distances remain fixed through time. If one looks at a coordinate function x^μ at $t = t_0$, at a later time the coordinate function can be written as $x^\mu \to a(t)x^\mu$, thus by dividing by the scale factor $a(t)$ one can define the comoving coordinates as

$$\chi = \int_{t_0}^{t} \frac{1}{a(t')} dt' \tag{1.22}$$

with the standard Minkowski metric multiplied by the scale factor a. This can be taken a step further by determining the distance light has travelled since $t = 0$, called the *comoving horizon*, by

$$\eta = \int_{0}^{t} \frac{1}{a(t')} dt'\,. \tag{1.23}$$

This chapter provides a brief overview of the relevant pieces of cosmology that will be used throughout the rest of this thesis. The next two chapters will use these pieces to see how one 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 ΛCDM. 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 was seen before, deviations from the FLRW metric are represented by the scalar fields Φ and Ψ. In the absence of anisotropic stress, $\Phi = -\Psi$, which is a good approximation for the late universe[1].

Throughout time in the universe, there are three distinct epochs [7]:

- The super-horizon region. At early times, the horizon η is small, and the Fourier modes k for all except the smallest scales are beyond the horizon ($k\eta \ll 1$).
- The horizon-crossing region. At intermediate times with rapid inflation, the horizon grows quickly. Fourier modes which were previously beyond the horizon cross into the horizon ($k\eta \approx 1$).
- The sub-horizon region. At late times, all Fourier modes except the larges scales remain within the horizon ($k\eta \gg 1$).

Additionally, the formation of matter inhomogeneity comes from gravitational instability. Primordial curvature perturbations sourced initial inhomogeneity, which evolved under gravitational forces to become the galaxies and clusters seen today.

This description provides a blueprint for writing the matter power spectrum. There is the primordial curvature perturbation $\mathcal{R}$ which source the initial matter inhomogeneity that formed late-time large-scale structure. Thus, it can be written as

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

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

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

And there is a growth factor, defined as the linear growth of matter density contrast during matter domination, defined as

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

[1] One can use codes such as CAMB [8] to do computations including anisotropy.

Thus the gravitational potential at all times can be written as

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

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

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

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

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

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

Thus, the linear matter power spectrum[2] is given by

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

2.1.2 Galaxy Density

In practice, it is impossible to directly measure the power spectrum; dark matter prohibits direct measurement. Instead, one must determine observables which trace the matter distribution in the universe. The most natural option is to observe the galaxy distribution, where galaxies also form due to gravitational interaction with dark matter. Galaxy density is a biased tracer of the full matter density, so the galaxy and matter power spectra are related by a linear bias factor.

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

Realistically, galaxy surveys 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 measurement is called photometric redshift. To account for this, one can attempt to redefine the galaxy power

[2]Linear comes from the fact that only linear terms in the equations of motion for the matter density are kept

spectrum. First, define the distribution of distances as

$$W(\chi) = \frac{1}{N_g}\frac{dN_g}{d\chi}\,. \tag{2.10}$$

The procedure for determining this distribution will be discussed in section 2.1.5. This allows one to simply write the projected density contrast as

$$\Delta_g = \int_0^\infty W(\chi)\delta_g(x,\tau)d\chi\,. \tag{2.11}$$

To complete this discussion, some approximations must be made. The first is that only consider small scales for which $\sin(\theta) \sim \theta \sim 1/l$ are considered, where l is the Fourier conjugate of θ. The second is the limber approximation, which states that, as long as the galaxy power spectrum P_g slowly varies over the region $\Delta k \sim 1/l\chi$, P_g can be approximated as a constant. These approximations greatly simplify the integrals required when computing the power spectrum of the Fourier transformed Δ_g. This power spectrum is called the angular power spectrum. Considering the approximations above and computing the Fourier transform, one finds an angular power spectrum of

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

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, through some changes of variables, $dx^i/d\chi = -\hat{p}^i$ (for photons and other massless particles). Thus, under lensing, the difference between the observed and true positions within the source plane is given by

$$\chi\theta^i = x^i_\perp = -\int_0^\chi \hat{p}^i_\perp(\chi'')d\chi''\,. \tag{2.13}$$

The goal is to write this as a function of the potential only. Thus, one must compute $d\hat{p}^i/dt$.

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

By computing $\dot{p^i}$, one can find $\dot{p}$ as well:

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

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

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

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

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

Plugging everything back in, and only keeping terms to linear order, one gets

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

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

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

Now, putting everything together, plugging these results into equation 2.14 gives

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

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

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

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

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

$$\frac{d\hat{p}^i_\perp}{d\chi} = -a\frac{d\hat{p}^i_\perp}{dt} = -\partial_i(\Phi - \Psi)\,. \tag{2.22}$$

In the absence of anisotropic stress, $\Phi = -\Psi$ and

$$\frac{d\hat{p}^i_\perp}{d\chi} - 2\partial_i\Phi\,. \tag{2.23}$$

Integrating this equation gives

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

where the constant C_i comes from integrating a derivative, where it is only defined up to a constant. Plugging this back in one finds

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

In the absence of lensing, this should reduce to $\theta^i = \theta_0$, the true source position. Hence $C^i = \theta^i_0$ and the total deflection angle is given by

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

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

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

This result is great, but it's not very experimentally illuminating. Similar to the galaxy versus matter power spectrum, the inability to directly measure the dark matter distribution means the potential Φ cannot be determined. Again, an attempt to relate the lensing potential to some property of galaxies can give a measurable quantity. However, it can be seen that lensing distorts the shape of the galaxies (figure 2.1), where it is assumed all galaxies are elliptical. Define the galaxy shape tensor by

$$q_{ij} \equiv \frac{1}{F} \int I \theta^i \theta^j d\theta^i d\theta^j \,. \tag{2.28}$$

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

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

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

$$A_{ij} \equiv \frac{\partial \theta^i_S}{\partial \theta^j} = \begin{pmatrix} 1+\kappa-\gamma_+ & -\gamma_\times \\ -\gamma_\times & 1+\kappa+\gamma_+ \end{pmatrix} = \delta_{ij} + \psi_{ij} \,. \tag{2.30}$$

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

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

Which relates to the shear components by

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

This connects the observable shear to the theoretical lensing potential.

2.1.4 Weak Lensing Correlation Functions

In the previous two sections, two observables directly connected to cosmological theories were defined. Observing the galaxies can determine Ω_b, A_s, n_s and H_0. Observing shear can determine Ω_m (the details of how to determine these parameters will be discussed in chapter 3). Together, these two can fully determine the values of the 5 ΛCDM parameters. These two quantities, however, are not measurable. Using these two quantities, however, one can construct

three different two-point correlation functions which are measurable.

The first correlation function is the autocorrelation of galaxy density, called galaxy clustering and denoted w_θ. 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

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

where J_0 is the zeroth 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 $\gamma_\times$. The cross correlation $\langle \gamma_+ \gamma_\times \rangle$ vanishes, and the remaining combinations are

$$\langle \gamma_+ \gamma_+ \rangle \pm \langle \gamma_\times \gamma_\times \rangle = \xi_\pm \tag{2.34}$$

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

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

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

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

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

2.1.5 Systematics

Up to now, any systematics of the weak lensing observables have been ignored. 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

the ones which affect the correlation functions after they have been computed. The latter type of parameters are usually referred to as 'fast' parameters, and it will become clear why that is in the next chapter. Meanwhile, the ways the systematics are modelled is described here.

Photometric Redshift Uncertainty

To compute the two-point correlation functions, 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.

$$n_g(z) \mapsto n_g(z + \Delta z) \tag{2.38}$$

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

Intrinsic Alignment of Galaxies

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 [7, 9]. There are two models of intrinsic alignment that will be considered in this thesis: the Tidal Alignment Tidal Torquing (TATT) model [10, 11] 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 by stretching matter along the radial direction from the gravitational source. The model is given by

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

where the coefficients C_a are given by

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

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

Galaxy Bias

The galaxy and matter density are related by a bias parameter b. However, a fiducial value can be chosen to relate the two:

$$\delta_{g,\mathrm{fid}} = b_{\mathrm{fid}}\delta_m \,, \tag{2.41}$$

then consider a multiplicative deviation:

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

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.

$$\begin{aligned} w^i(\theta) =& b^2_{(i)} w^i(\theta) \\ \gamma^i_t(\theta) =& b_{(i)} \gamma^i_t(\theta) \,. \end{aligned} \tag{2.43}$$

where i denotes the redshift bin.

Shear Calibration

Observed shear is not completely caused by cosmological interactions. For example, a major source of additional shear comes from the observation lens itself, where the observed galaxy shape is given by the true galaxy shape convolved with the point spread function of the lens [12, 13]. This is usually modelled as a multiplicative correction to the shear components.

$$\begin{aligned} \xi^{ij}_{\pm} =& (1 + m^i)(1 + m^j)\xi^{ij}_{\pm} \\ \gamma^{ij}_t =& (1 + m^j)\gamma^{ij}_t \,, \end{aligned} \tag{2.44}$$

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

Point Mass

Currently, it is not possible to model the contribution of small scales to the tangential shear [14]. The tangential shear can be written in terms of the

surface mass density Σ tangential to the line of sight:

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

The tangential shear can only be modelled down to scales of $r_{\rm min}$, which allows one to write the point mass contribution (named for the R^{-2} dependence) as

$$\Delta\Sigma(R) = \Delta\Sigma^{\rm model}(R) + \frac{B}{R^2}\,. \tag{2.46}$$

Lensing Magnification

In weak lensing, 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 sparsely in this thesis, it is worth discussing the measurements qualitatively [15–18] The consensus among physicists and astronomers that the early universe was a hot, dense plasma. As such, the mean free path of photons was much shorter than it is today, and thus Compton scattering dominated photon dynamics in the early universe. As the universe cooled, atomic nuclei formed, and electrons began to bind to the nuclei in a process called recombination [7, 15]. As the universe began to expand, the mean free path of photons increased and inter-atomic interactions decreased, meaning the amount of scattering was decreasing, until a point where photons are able to move freely through the universe.

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

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

The CMB is an almost-perfect black body, meaning the intensity of the light can be used to determine the temperature according to the black body equation where $I \propto T^4$. However, the imperfections source anisotropies in the temperature on the scale of 10^{-5} K. The temperature, along with the curl-free E-mode polarization, gives three possible two point correlation functions: TT, TE, and EE.

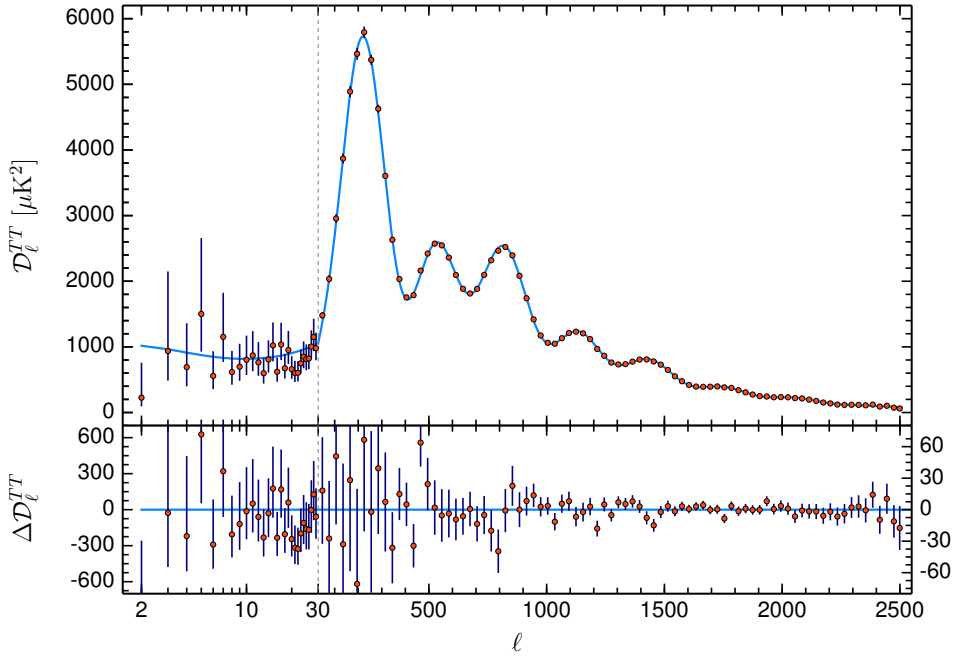


Figure 2.2: Planck TT power spectrum [4].

2.2.1 TT Power Spectrum

The TT power spectrum measures the anisotropies of the temperature field. Thus, it is better to consider 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, consider Ω_b, the baryon density. If the density is increased, there will be an increase in Compton scattering and a 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 because an increase in baryon density causes an increase in the gravitational interactions between the baryons. This shifts down the oscillations in temperature contrast so that the acoustic oscillations

occur around some negative value. Thus, in the TT power spectrum, every other peak is enhanced.

Next is to consider a change is 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: $\mathcal{A}_s$ and τ_{rei}. Naturally, τ_{rei} will have an overall suppression effect. The amplitude is given by $\mathcal{A}_s e^{-2\tau_{\text{rei}}}$. Thus, the effect of $\mathcal{A}_s$ can look identical to changes in τ_{rei}. n_s, as before, defines the slope of the primordial power spectrum. Thus, changes to n_s have a stronger effect on small scales than on large scales.

Finally, is a consideration of H_0, the most simple case. H_0 affects the expansion rate of the universe, thus a changing the value of H_0 will shift the position of the peaks in the power spectrum. Increases in H_0 will shift the peaks to smaller scales (found by propagating H_0 backward through time), and inversely, lowering H_0 will shift the peaks to larger scales. Additionally, the peaks at small scales will have a lower shift than the peaks at large scales. Hence, H_0 can also be determined by the distance between the peaks.

As a final remark, one can consider the curvature density Ω_k. Changes in Ω_k affect the angular distance to recombination, resulting in an overall shift of the TT power spectrum. Measurements strongly constrain this to $\Omega_k = 0$, suggesting a flat universe.

2.2.2 TE and EE Power Spectra

The other two power spectra are shown in figure 2.3. The main consideration with these two power spectra is to observe the way it breaks the degeneracy between τ_{rei} and $\mathcal{A}_s$. τ_{rei} suppresses the power spectrum as it suppresses the fraction of light that reaches an observer. However, the additional scattering actual 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, using all three power spectra will give the strongest constraints on the cosmological parameters.

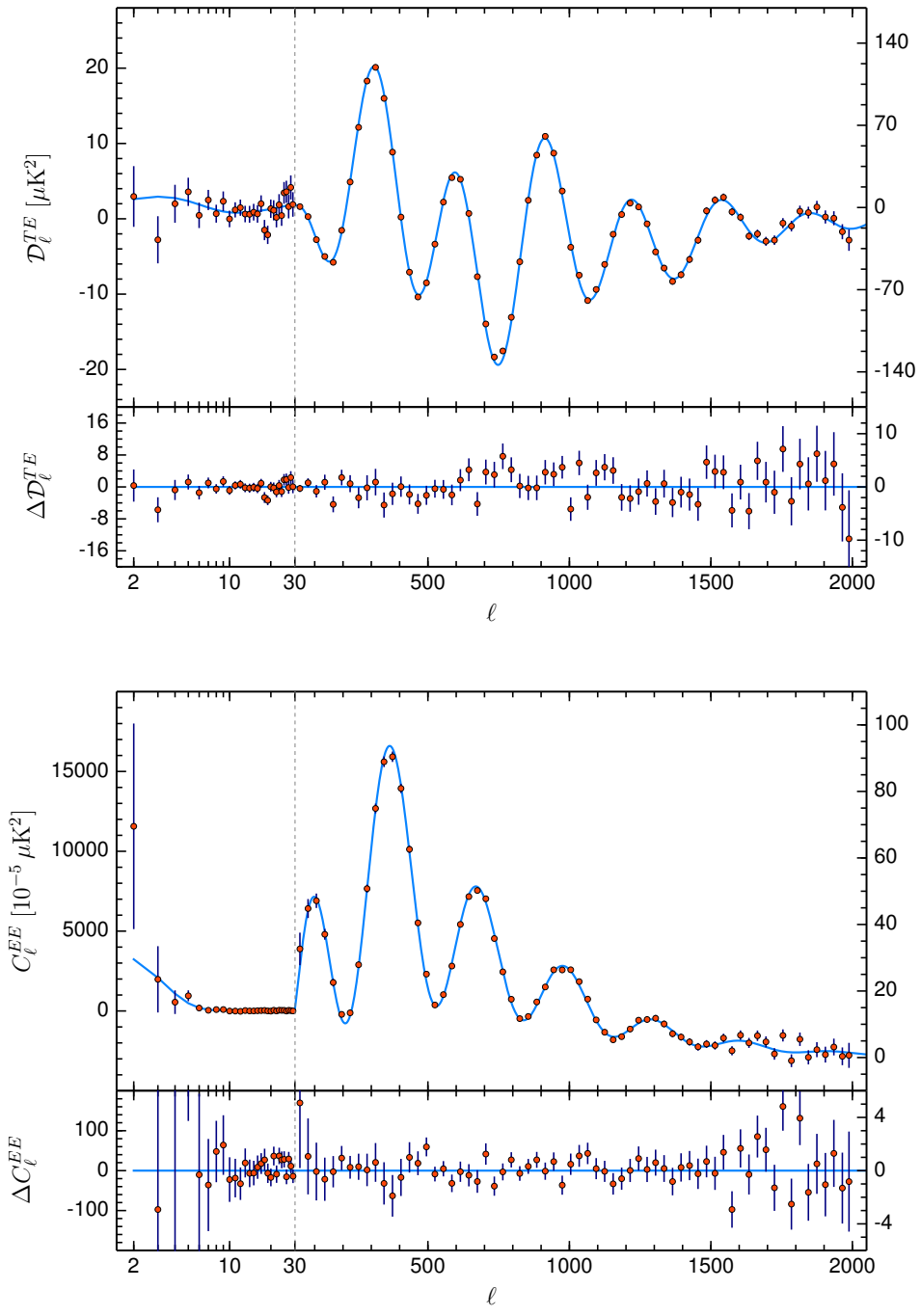


Figure 2.3: Planck TE (top) and EE (bottom) power spectra [4].

2.3 Tensions Between Weak Lensing and the CMB

As seen, there are a total of five parameters needed to describe ΛCDM in weak lensing: Ω_m, Ω_b, H_0, $\mathcal{A}_s$, and n_s. This parameterization is not unique. Another common parameter, which tracks $\mathcal{A}_\int$, is σ_8

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

which can be described as the amplitude of matter fluctuations at the scale $8h^{-1}$Mpc. With this parameter instead of $\mathcal{A}_s$, weak lensing gives strong 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 (precise results can be found in [4, 19, 20]). The main tension between CMB and weak lensing is the S_8 tension at about a 3σ difference (figure 2.4). The most famous discrepancy is the H_0 tension (figure 2.5). The tension is produced by a discrepancy between CMB measurements and the local cosmic distance ladder, producing a nearly 5σ difference. There have been extensive efforts to propose a model that can resolve these tensions (early dark energy[21], quintessence [22], etc.), but as of now no model can 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.

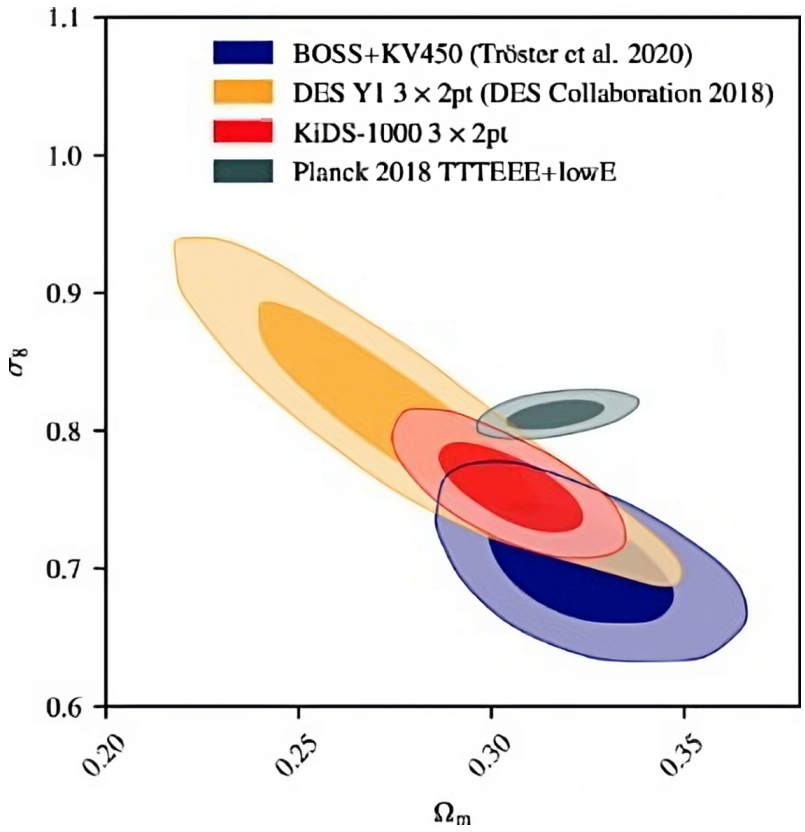


Figure 2.4: Measurements of S_8 from different experiments.

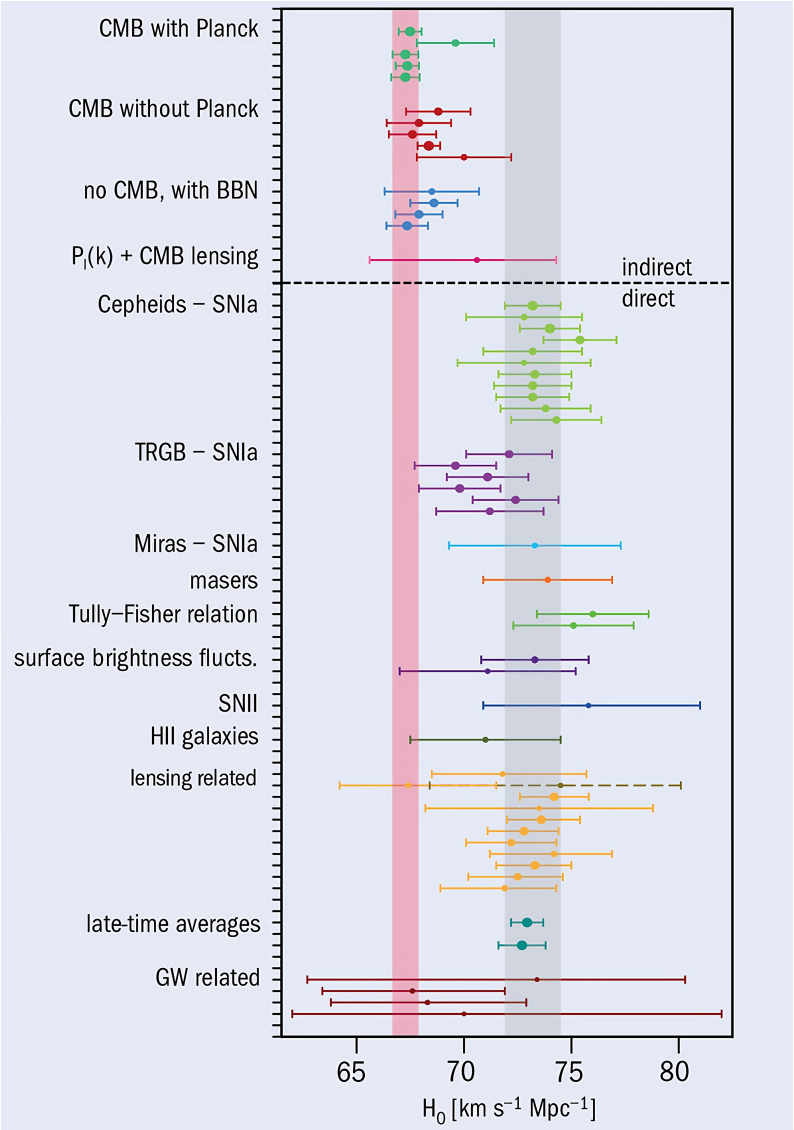


Figure 2.5: Measurements of H_0 from different experiments.

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? From the complexity of the calculations, it's clear 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 the observables given the set of parameters. Using this language, it's clear that this computation involves conditional probability densities, so one can use Bayes' theorem to determine the parameters.

3.1 Markov Chain Monte Carlo

Markov chains are a step process in which the next step depends solely on the current step [23]. Formally, a sequence $X_1, \ldots, X_n$ of random elements is a *Markov Chain* if the conditional distribution of X_{n+1} depends solely 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 word to follow? For a cosmological application, one starts with some set of parameters, and slowly moves towards the most probable set of parameters.

Bayes' theorem states

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)}\,. \tag{3.1}$$

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

$$P(\theta|d) = \frac{P(d|\theta)P(\theta)}{P(d)}\,. \tag{3.2}$$

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 [24], this is usually written

$$\mathcal{P}(\theta) = \frac{\mathcal{L}(\theta)\Pi(\theta)}{\mathcal{E}}\,. \tag{3.3}$$

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 [25], however these usually take a very long time). Thus, with a given likelihood and prior, it is not feasible to compute the posterior exactly. Markov chain Monte Carlo (MCMC) describes a probabilistic 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 [26–28] is the most common algorithm for doing MCMC. Suppose one wants to sample from a distribution $p(x)$. First, 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*

$$p(x_n)g(x_{n+1}|x_n)A(x_n \to x_{n+1}) = p(x_{n+1})g(x_n|x_{n+1})A(x_{n+1} \to x_n)\,, \tag{3.4}$$

where A is an *acceptance probability*, which will be defined 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 rely solely on the likelihood and prior of $p(x)$, which is denoted π and $\mathcal{L}$.

$$\begin{aligned}\pi(x_n)\mathcal{L}(x_n)g(x_{n+1}|x_n)A(x_n \to x_{n+1}) = \\ \pi(x_{n+1})\mathcal{L}(x_{n+1})g(x_n|x_{n+1})A(x_{n+1} \to x_n)\end{aligned} \tag{3.5}$$

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

This allows one to define the acceptance probability as

$$A(x_n \to x_{n+1}) = \min(1, R_{n,n+1})\,. \tag{3.7}$$

This probability is used to determine whether the chain moves to x_{n+1} or stays at x_n.

To assess whether the chain has converged, there are two common methods. The first is the Gelman-Rubin method. This defines the number R as the ratio between the variance of a walker to the variance across all walkers. R approaches one as the chain converges. The second is the autocorrelation time τ. Although Markov chains depend explicitly on the previous state, there is an implicit dependence of all previous states. The autocorrelation time attempts to measure how many steps are required for a walker to forget its initial position. τ thus tells one how many samples are required to get independent samples of the posterior.

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 . 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 [5, 29], 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, the difference posterior will be centered at 0, so the 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:

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

Thus, the tension can be approximated as the volume of the posterior contours with $\mathcal{P}(\Delta\theta) > \mathcal{P}(0)$:

$$\text{PTE} = \int\limits_{\mathcal{P}(\Delta\theta)>\mathcal{P}(0)} \mathcal{P}(\Delta\theta)d\Delta\theta \,. \tag{3.9}$$

This comes with 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

Q_{UDM} is defined as

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

The difference of means is precisely the mean of the parameter difference distribution, and they are weighted by their 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, one can 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 $A + B$.

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

$Q_{\rm DMAP}$ is again χ^2-distributed. The 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 the likelihood is given a set of cosmological parameters which belong to a 5 dimensional parameter space. The number of 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 in not constrained by the likelihood and a ratio $\ll 1$ means a parameter is well constrained by the likelihood. Thus, the number of degrees of freedom is

$$d = N - \mathrm{tr}(C_\Pi^{-1} C_\mathcal{P}) \,. \tag{3.12}$$

Metric 4: Eigentension

This metric differs from the others by restricting our parameters to ones that are well-measured. This is done 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 motivation for removing 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

$$R = \frac{\mathcal{E}_{AB}}{\mathcal{E}_A \mathcal{E}_B} \,, \tag{3.13}$$

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

$$\log S = \log R - \log I \,. \tag{3.14}$$

I is the information ratio, which is defined in terms of the KL divergence:

$$\log I = \mathcal{D}_A + \mathcal{D}_B - \mathcal{D}_{AB} \,, \tag{3.15}$$

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

3.2.3 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,

$$n_\sigma = \sqrt{2}\mathrm{Erf}^{-1}(P) \,. \tag{3.17}$$

There is no concrete reason to believe P comes from a Gaussian distribution in general. It does, however, come from a Gaussian distribution when the posteriors are Gaussian; thus one 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.

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

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

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

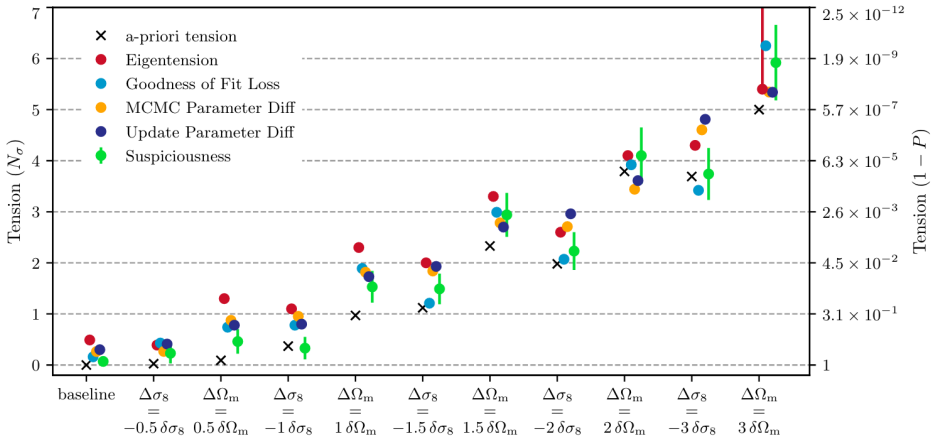


Figure 3.1: Result of tension metrics between Planck and DES [5]. A-priori tension is the Gaussian approximation described in equation 3.20.

One 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.

3.2.4 DES Versus Planck Results

A subset of these exercises have been performed previously by the Dark Energy Survey (DES) against the Planck results. In [5], Lemos et al. generate an a-priori tension on the cosmological parameters by shifting the DES fiducial cosmology along σ_8 and Ω_m. They approximate the a-priori tension using the Gaussian approximation,

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

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 for Density Estimation (MADE) to construct the flow [24, 30, 31]. Suppose the input to the flow is 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 $x_{1:i-1} = (x_1, \ldots, x_{i-1}, 0, \ldots, 0)$. Since the input depends solely on the first $i-1$ inputs, the normalizing flow is *autoregressive* and the Jacobian is triangular.

The implementation in TensorFlow uses *bijectors*, a local diffeomorphism between a manifold M and a target manifold N (which are our parameter spaces), i.e., $\phi : M \to N$ such that ϕ is differentiable and injective. In TensorFlow, bijectors have three operations, forward, inverse, and log_det_jacobian, which are precisely the three needed. 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 $\phi_i : X \to U_i \subset \mathbb{R}^n$ 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 $\psi_j : X \to V_j \subset \mathbb{R}^n$ 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 $\tau : \phi_i(Y) \to \psi_j(Y)$. To ensure our parameterizations are diffeomorphic, τ is required to be differentiable and its inverse differentiable (in fact, the reparameterizations are smooth, 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 though of as a finite sequence of vector fields $v_i : X \to TX$. Each point is moved along its integral curve determined by each v_i. This definition is that immediately implies an MAF is a diffeomorphism.

In formal terms [32], suppose $v : X \to TX$ is a vector field on X and $x \in X$. Then for any $\epsilon > 0$ there exists a unique function $f : B[-\epsilon, \epsilon] \to X$ such that f is an integral curve of with infinitesimal generator v and with $f(0) = x$.

Given a vector $v_x : U \to T_x X$ with $x \in U$, for it to be in the tangent space

T_pX it must have a curve with $f(0) = x$ and $f'(0) = v_x$. This follows from the definition of T_xX.

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 ODEs for $i = 1, \ldots, n$

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

This can be solved by $\gamma^i(t) = \int\limits_{U \subset X} v^i(\gamma(t))\text{vol}.$. 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 $t \in I$.

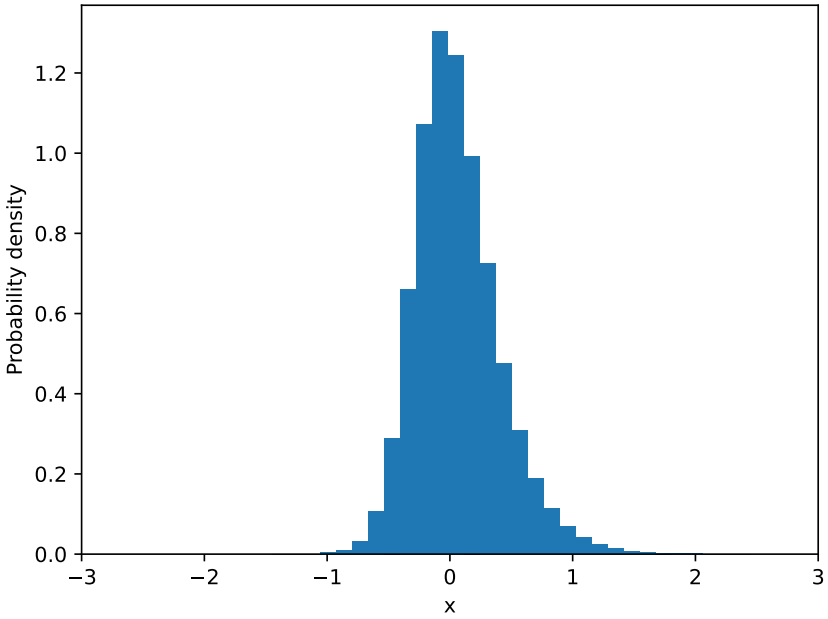


Figure 3.2: An example non-Gaussian distribution (non-zero skewness).

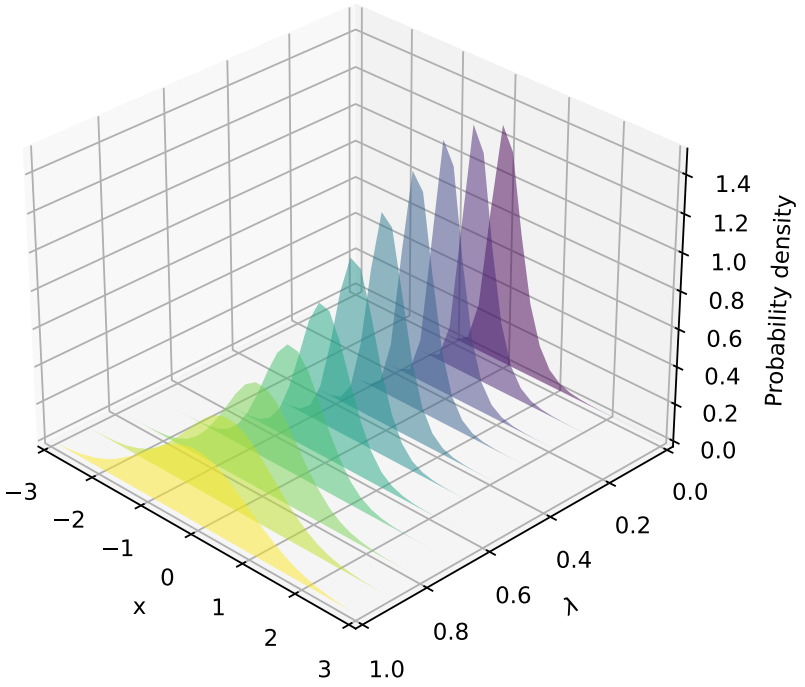


Figure 3.3: An example of normalizing flow as the transformation is 'turned on', represented by the parameter λ. The non-Gaussian distribution at $\lambda = 0$ is transformed into the Gaussian distribution at $\lambda = 1$.

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 [33] and DES-Y3. These results are also available in a paper written by Kunhao Zhong, another Masters student, and myself, on arXiv [34] and are pending publication in 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}, a single growth factor and a single matter power spectrum; these are split as well. From equation 2.8, 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

$$P^L_{\rm split}(k,z) = P^L_{\rm geo}(k,z)\left(\frac{G_{\rm growth}(z)}{G_{\rm geo}(z)}\right)^2 , \tag{4.1}$$

which means σ_8 is also split according to

$$\sigma^2_{8,\rm split} = \sigma^2_{8,\rm geo}\left(\frac{G_{\rm growth}(z)}{G_{\rm geo}(z)}\right)^2 . \tag{4.2}$$

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

$$P(k,z) = P^L(k,z) \times B(k,z)\,. \tag{4.3}$$

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

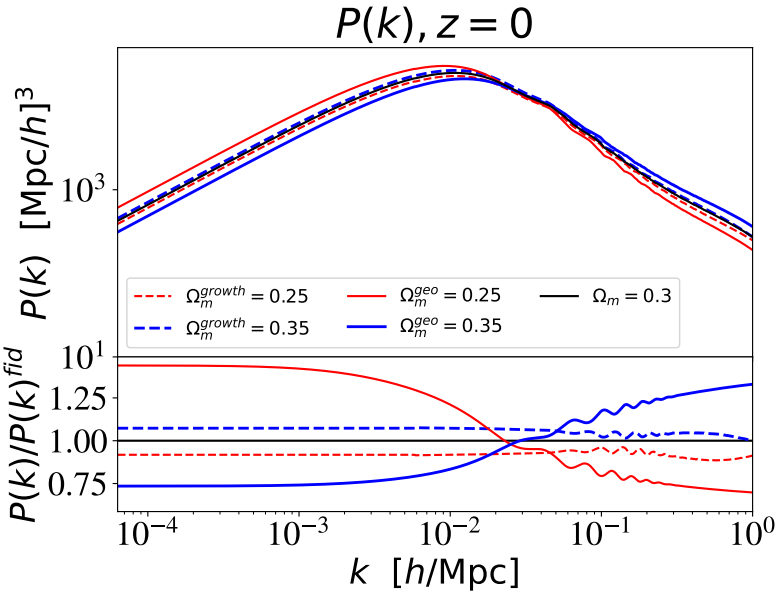
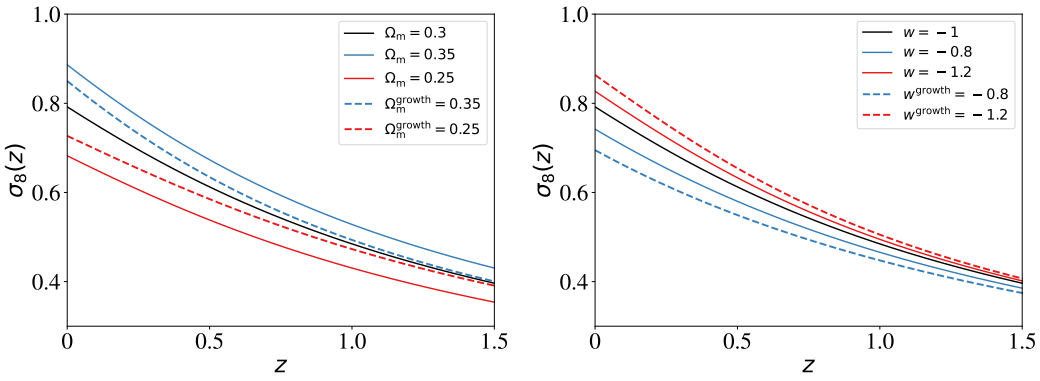


Figure 4.1: Non-linear power spectrum from Euclid emulator. When either the geometry or growth parameters are varied, the other is kept fixed at $\Omega^X = 0.3$.

the geometry parameters. This is due to the enhanced gravitational effects for higher $\Omega_m^{\rm geo}$, or suppressed gravitational effects for lower $\Omega_m^{\rm geo}$. From this, one can predict that the value of σ_8 will change, the change being larger for $\Omega_m^{\rm geo}$ than $\Omega_m^{\rm growth}$, 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



(a) ΛCDM split.

(b) wCDM split.

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.

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 (Ω_m^{geo}, w^{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

- CMBP: The Planck CMB TTTEEE power spectra together with the

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		
$\Delta z_1^1 \times 10^2$	Gauss(0.8, 0.7)	Gauss(0.6, 0.4)
$\Delta z_1^2 \times 10^2$	Gauss(-0.5, 0.7)	Gauss(0.1, 0.3)
$\Delta z_1^3 \times 10^2$	Gauss(0.6, 0.6)	Gauss(0.4, 0.3)
$\Delta z_1^4 \times 10^2$	Gauss(0, 0.01)	Gauss(-0.2, 0.5)
$\Delta z_1^5 \times 10^2$	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.

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 the baryon density from Big Bang Nucleosynthesis as $100\Omega_b h^2 = 2.208 \pm 0.052$.

- 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[1], which interfaces Cobaya [35], the MCMC sampler we use, with CosmoLike [36], 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 $R - 1 < 0.02$.

[1] https://github.com/CosmoLike/cocoa

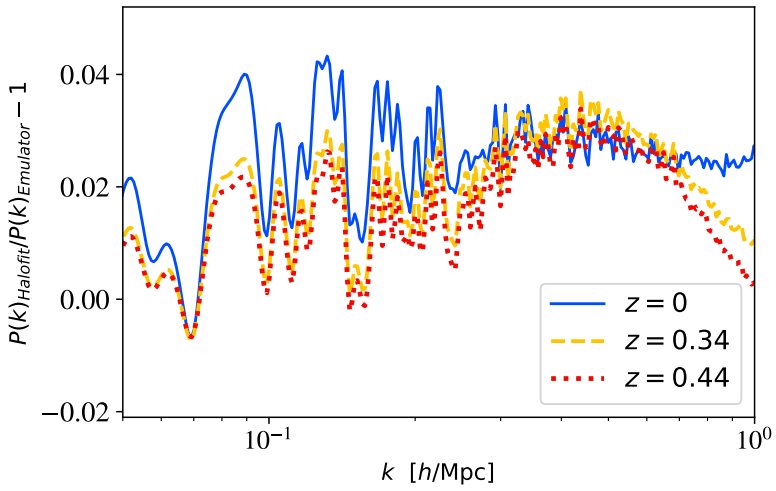


Figure 4.3: Relative difference between Euclid emulator and Halofit.

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 (equation 4.4), and we hope to find no difference between growth and geometry parameters and hope each parameter is equal to the fiducial value.

$$(\mathcal{A}_s \times 10^{-9}, n_s, H_0, \Omega_m, \Omega_b) = (2.101, 0.965, 67.32, 0.317, 0.049) \tag{4.4}$$

Using the ALL prior and external data, there is marginal improvement between the cosmic shear constraints on $\Omega_m^{\rm growth}$ compared to the prior. However, there is significant improvement in the 3×2pt case compared to cosmic shear.

In wCDM, there is poor constraining power on $\Omega_m^{\rm growth}$ and $w^{\rm growth}$, even with the most informative 3×2pt case. For this case, we give constraining power along the first principal component

$$\mathrm{PC}_1 \equiv 0.7071\Delta\Omega_m - 0.7071\Delta w\,. \tag{4.5}$$

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 $\Omega_m^{\rm growth}$ between DES-Y1 and DES-Y3. In the 3×2pt case with the ALL prior gives a smaller ($\sim 17\%$) error bar. Constraints on PC_1 are prior dominated and nearly identical for the wCDM split in DES-Y1 and DES-Y3.

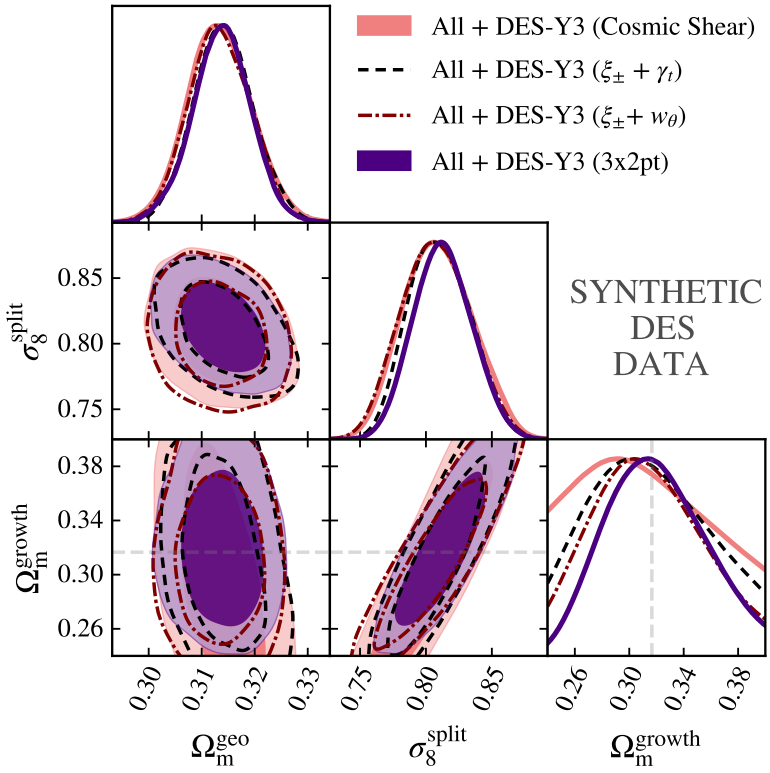


Figure 4.4: Cosmic shear, 2×2pt combinations, and 3×2pt posteriors for DES-Y3 with the ALL prior.

4.3 Results

4.3.1 ΛCDM

For the most constraining case, 3×2pt 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 is 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 clustering posterior. However, these shifts from $\Delta\Omega_m = 0$ are prior dominated,

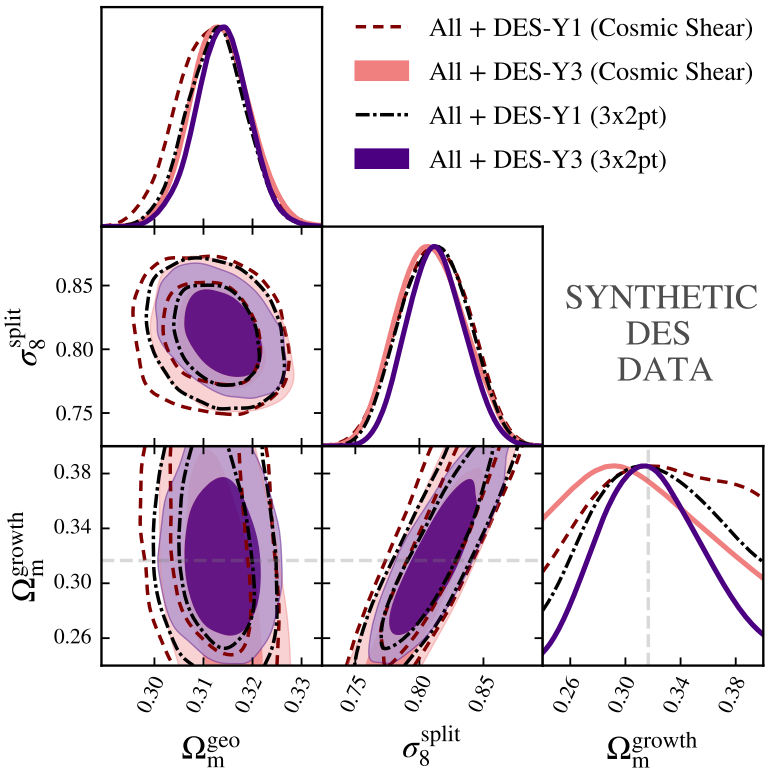


Figure 4.5: Comparison of DES-Y1 and DES-Y3 posteriors.

not data dominated, thus do not constitute evidence for a shift.

The full results of the 1-D marginalizations with each choice of prior is summarized in figure 4.8. Aside from the 2-point correlation functions, every $\Delta\Omega_m$ result is compatible with 0 to within 1-sigma. Similarly to the synthetic chains, the 3×2pt 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.

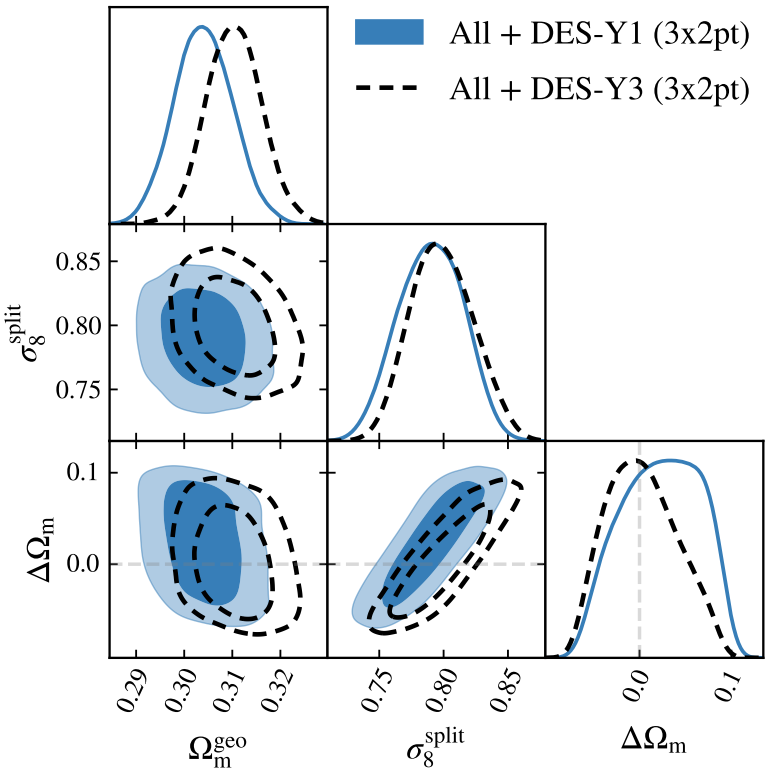


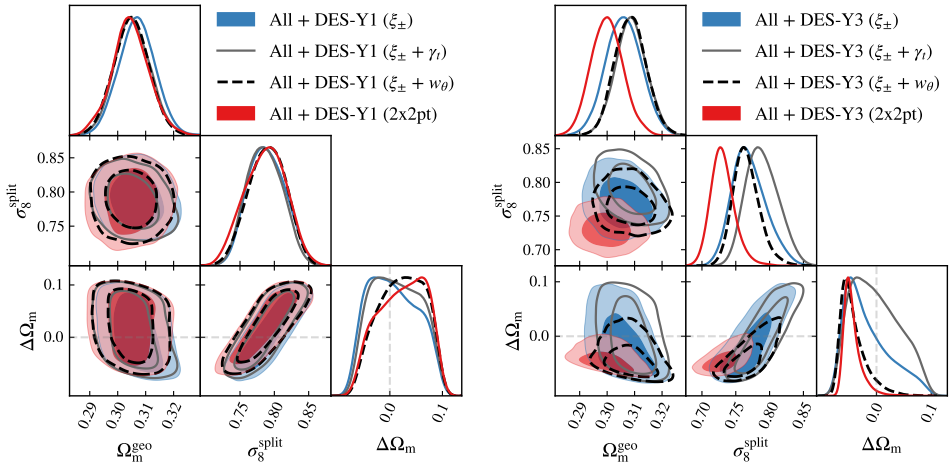
Figure 4.6: Split posteriors for DES-Y3 3 × 2pt and ALL prior.

4.3.2 wCDM

The results for wCDM 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 2 × 2pt 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 $(\mathcal{A}_s, n_s, H_0, \Omega_m^{\rm geo}, \sigma_8^{\rm split})$ (with $w^{\rm geo}$ added in the wCDM case). Interestingly, the tension appears to be generated predominantly by $\mathcal{A}_s$. Additionally, the 2 × 2pt correlation functions have a significant degradation in the goodness of fit when CMB priors are added. This reflects the RedMaGiC inconsistency



(a) DES-Y1 (b) DES-Y3

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

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

In the wCDM case, we once again see the 2×2pt has a significant shift. Contrary to the split ΛCDM case, the $\mathcal{A}_s$ tension is much smaller, making the detection of a shift in PC_1 more meaningful. Once again, an analysis involving $X_{\rm lens}$ 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 2×2pt chains, however, there is an internal inconsistency with the RedMaGiC samples that may source this result. Because we fixed $X_{\rm lens} = 1$, an additional analysis can be done investigating if $X_{\rm lens}$ can source a significant shift. Additionally, one can repeat the analysis with the MagLim lensing samples. In the

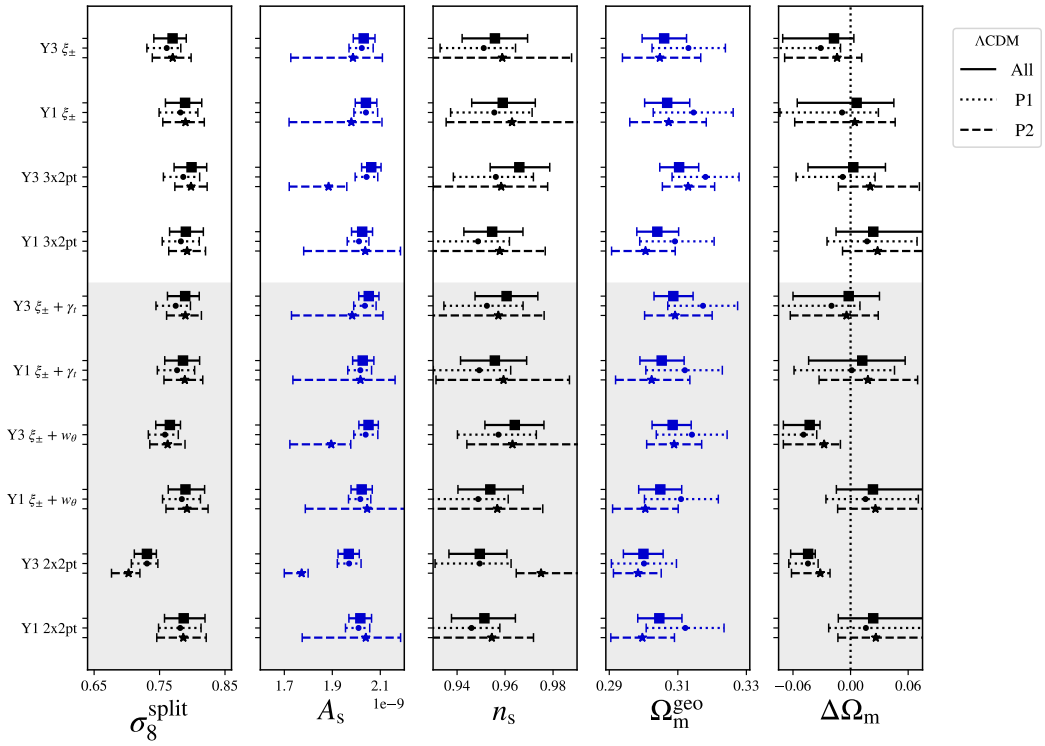
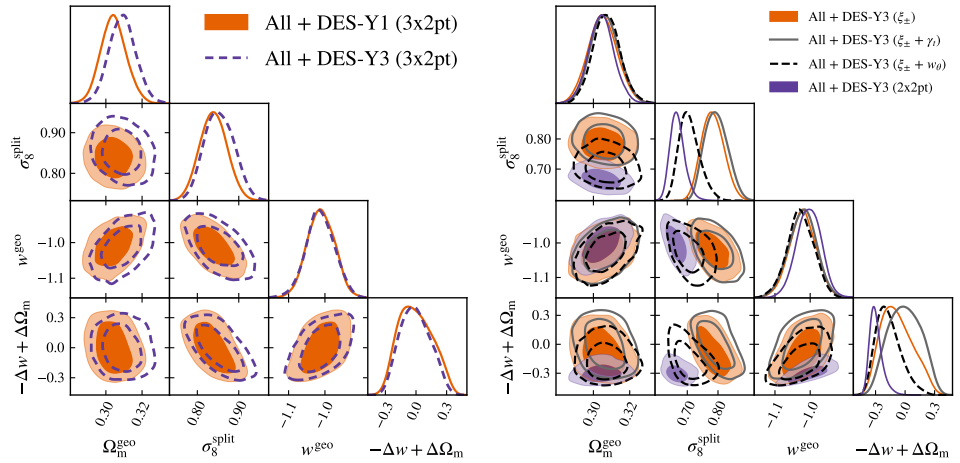


Figure 4.8: Full 1d results for ΛCDM split

future, 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 3×2pt wCDM posteriors.

(b) DES-Y3 2-point correlation function posteriors

Figure 4.9: wCDM posteriors.

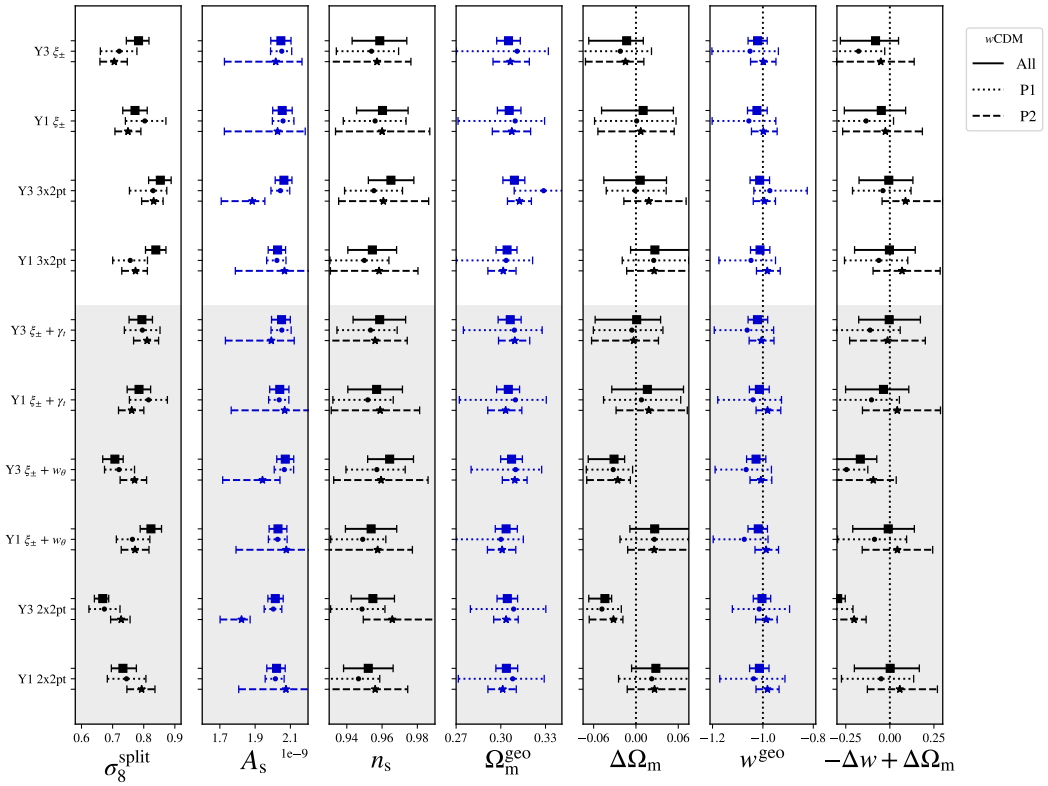
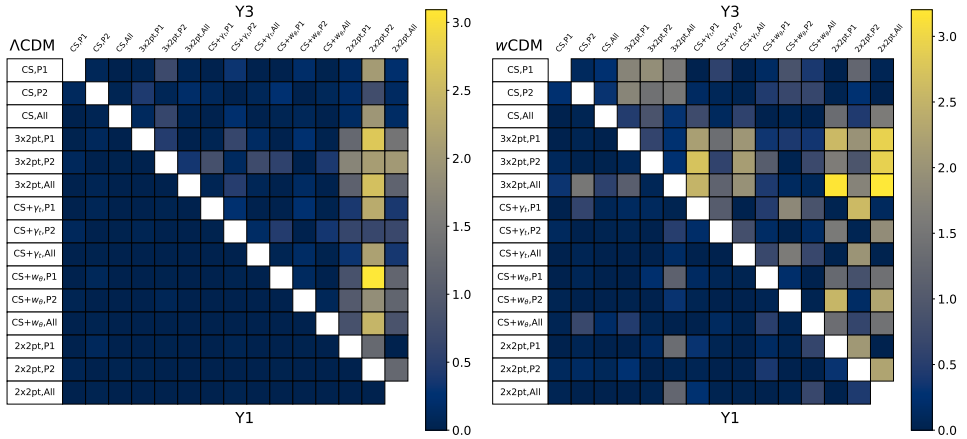


Figure 4.10: Full 1d results for ΛCDM split



(a) Internal tension in the ΛCDM split.

(b) Internal tension in the wCDM split

Figure 4.11: Summary of internal tensions between each probe in the ΛCDM and wCDM models.

Chapter 5

Data Emulators

As can be seen from the growth-geometry split analysis, an enormous amount of time and computing resources are needed to do a comprehensive analysis of novel cosmological models and extensions to ΛCDM. 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. There is a 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

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

In this exceptional case, there is an analytic way to minimize L, but in general, it is a highly non-trivial task. 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, as humans, have a tool that can do this already: the brain [37]! 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, one can attempt to construct a general model, starting with the neurons 5.1. Next, add in the synapses, the connections between neurons 5.2.

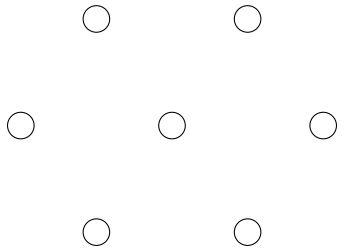


Figure 5.1: Arrangement of neurons.

In general, 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. Furthermore, each synapse has a weight that represents the strength of the connection. Next, define which neurons receive an external

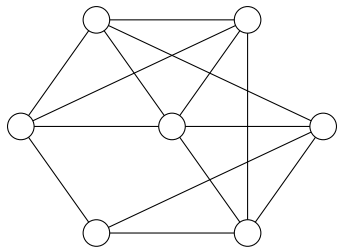


Figure 5.2: Neurons with synapses.

signal, and which give a signal to the outside 5.3. These act as the input and output. Finally, there is a choice to make in the activation function. This is the chance to greatly increase the modelling capability of NNs. The only requirement is that the function is continuous; thus one can use non-linear activation functions, allowing modelling of highly non-linear outputs.

The goal is to define a loss function that will be minimized (in the linear regression, the loss is the square of the residual). The loss is a function of the output of the NN, and the output is parameterized by the parameters of the

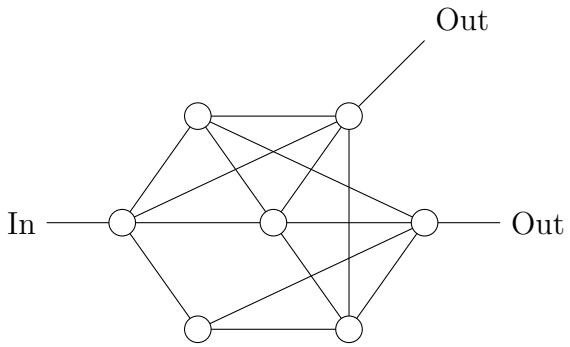


Figure 5.3: Neurons, synapses, and the input/output.

NN. Thus, if $y = \phi(\alpha_i, x)$ represents the output of the neural network with parameters α_i and input x, the loss function is, in fact, parameterized by the neural network as well.

$$L(y) = L(\phi(\alpha_i, x)) = \mathfrak{L}(\alpha_i, x) \tag{5.2}$$

Thus, a neural network is, in simple terms, an optimization problem. The issue, however, is that neural networks generally have numerous parameters (thousands, millions, even billions!), so this optimization is difficult in practice. Moreover, 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 the gradient is 0.

This simple algorithm has one glaring flaw: it is susceptible to converging in a local minimum. Ideally, gradient descent will converge to the global minimum of 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 vanishes, or a gradient that explodes to large values. This is the problem of *vanishing gradients*, and methods to avoid this are discussed in the next section.

The last issue to mention is that, given numerous parameters, it is possible to fit the training data very well, but the model cannot fit external data sets. This is called *overfitting*. There are many strategies to resolve this. The two used in this project are 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.

$$L(y) \mapsto L(y) + \sum_i \alpha_i^2 \tag{5.3}$$

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. It is important to pick a smart loss function depending on one's 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 also important to implement strategies to avoid overfitting on the training data by penalizing the model for relying too heavily on certain regions of the graph.

Using NNs to compute data vectors is not a new technique [38, 39]. The previous methods typically use an iterative approach, where they start by training on a Latin hypercube, and use the trained model to generate new training points from an MCMC that are added to the hypercube points. This process is repeated until the loss function passes some threshold. The claim

that this method gives satisfactory results on the entire Latin hypercube is somewhat misleading. While the hypercube points remain in the training data, the points become diluted until they have minimal contribution to the training loss. Testing with the emulator in [38], the emulator begins to fail at around 4σ shifts. Our goal is to greatly improve the region where the emulators are useful. This will allow other scientists to use emulators to analyze real data, where the fiducial cosmology is not known.

5.2 Architecture Choices

As discussed before, the weights, activation functions, and connections are choices one can make when designing a neural network. Together, they comprise the architecture. This section discusses 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:

$$x^i = \frac{\theta^i - \bar{\theta}^i}{\sigma_i}\,. \tag{5.4}$$

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

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

where C is the covariance of the data and P is the change of basis matrix to the eigenbasis of C. This generally means we will have many negative parameters, and the input and output are both symmetric about 0, and 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 in 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).

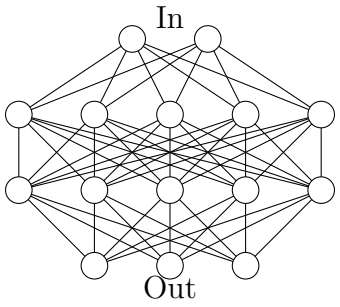


Figure 5.4: A multi-layer perceptron.

5.2.2 Residual Network

As mentioned above, there is an issue of vanishing gradients with deep neural networks. To resolve this, one can add the input and output together so that gradient information is allowed to skip layers [40](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 layers, whose map is denoted ϕ, will be

$$\phi(\alpha_i, \chi(\beta_i, x)) = \psi(\alpha_i, \beta_i, x) + \chi(\beta_i, x)\,. \tag{5.6}$$

This section of the model is going to attempt to learn

$$\psi(\alpha_i, \beta_i, x) - \chi(\beta_i, x)\,. \tag{5.7}$$

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).

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

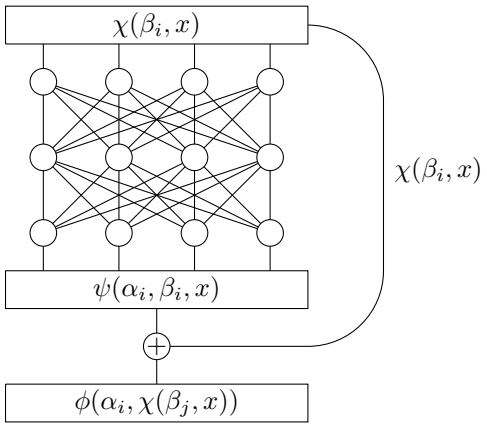


Figure 5.5: A residual block.

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, it is referred to as a *bottlenecked residual network* (ResBottle) (figure 5.6). In the examples

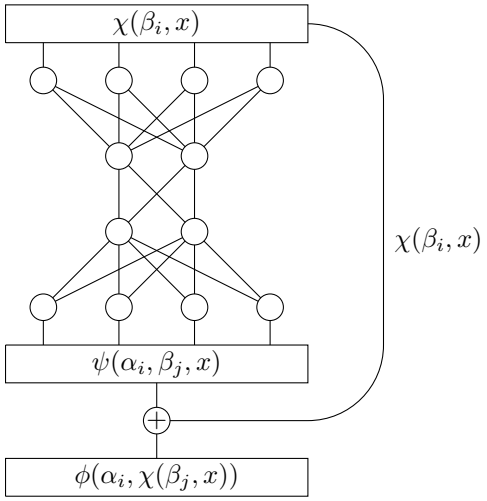


Figure 5.6: A bottlenecked residual block.

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 one examines the structure of the data vector, there is a certain order to the data and correlation between the components. The data vectors are separated into redshift bins, and within each redshift bin are smaller bins representing different scales. Thus, a NN that can learn sequences may be useful. Fortunately, there are NNs encountered every day that use sequences: chatbots [41]! The basis for this is usually attention networks [42]. These work by examining dot products between vectors so that the output is directly related to the angle between the two vectors.

Suppose there is a given sequence of input vectors, $x = (x_1, x_2, \ldots, x_n)$, then one can construct dot products of each x_i and use them as coefficients for the output.

$$y_k = (x_i \cdot x_j)x_k \tag{5.8}$$

One can then introduce a generalized dot product using linear transformations. That is, one 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 the dot product can be written as

$$y_k = (q_i^T k_j)v_k \,. \tag{5.9}$$

Finally, one can introduce a non-linearity, which also keeps the output sequence y normalized, by introducing the *softmax* function,

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

where the softmax function is an exponential normalization, given by

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

and d is the embedding dimension (i.e $W_K \in \mathbb{R}^{d_x \times d}$). The linear matrices W_Q, W_K, and W_V, contain learnable components. Each output y_i is then passed to its own MLP. We also keep the skip connections for residual learning. 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"

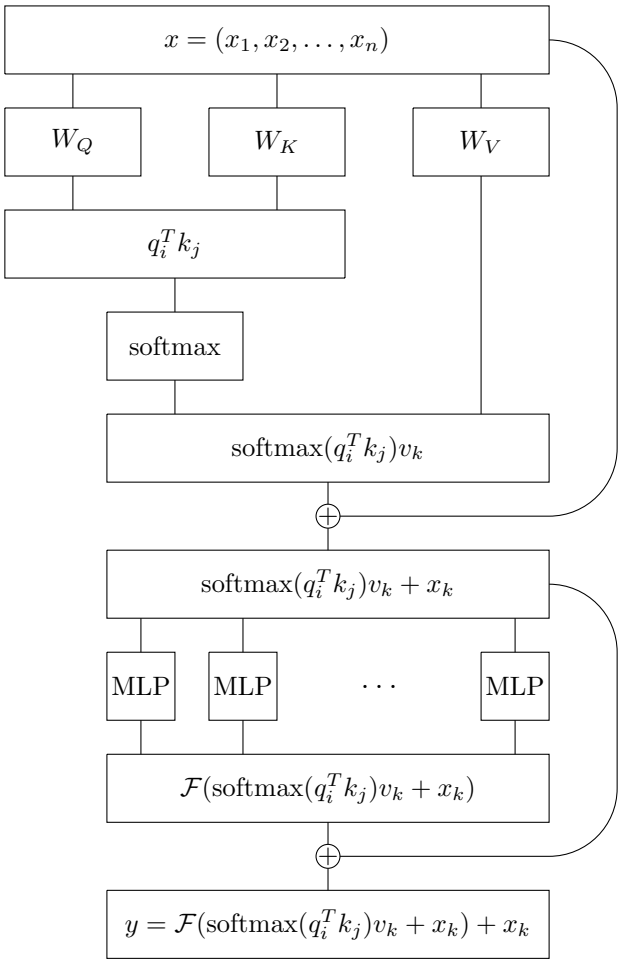


Figure 5.7: A transformer block

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 can 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 covariance of 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 tested the different architectures by training them on a $T = 16$ chain. The testing was done on a $T = 8$ chain. we find that the ResNet is able to consistently perform the best of the three basic architectures. The number of layers and width of the layers makes minimal difference. An interesting result comes from the wide MLP architectures, where the testing loss suddenly spikes. This is likely due to either vanishing gradients or some poorly chosen hyperparameters of the model. Regardless of the issue, this instability is not desirable, as we would like to provide a framework for others to build emulators of their own, and fine-tuning hyperparameters adds significant difficulty to this process. The ResBottle consistently performs worse than the others, meaning the information the model learns in necessary to correctly model the final data vector. This is perhaps unsurprising for our relatively small models, but if one needs to create much wider networks, the ResBottle may be the only way to keep the number of parameters reasonable. Nevertheless, the ResBottle has a satisfactory performance overall.

Additionally, we can test how of a broad of a region in parameter space we can train on and still get satisfactory performance. We do this by scaling the covariance by a temperature T, i.e $C \mapsto TC$. We noticed the MLP, ResNet,

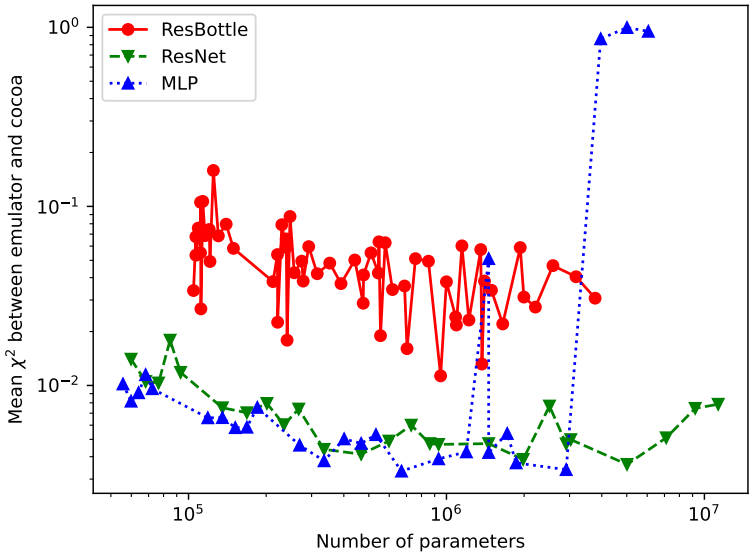


Figure 5.8: Average $\Delta\chi^2$ as a function of the number of parameters for each architecture.

and ResBottle models fail to train on a $T = 128$ chain (figure 5.9). We added the transformer block to the $T = 128$ training to see the gain in performance attributed to the complex model. We find adding a transformer block to the end of the 3 ResBlock model (figure 5.11) can give satisfactory performance when training on $T = 128$ and testing on $T = 64$ chains.

This demonstrates the importance of making an intelligent choice for emulator architecture. By implementing the self attention, we were able to greatly increase the amount of the ΛCDM parameter space covered by the emulator. The additional volume allows one to use emulators, even on real data, where the data vectors can be shifted greatly from things such as baryon contamination. We would like to continue this to see how this architecture will hold for more complex extensions to ΛCDM, such as early dark energy, which adds ~ 10 parameters. We are also working with a team on using this architecture to model much larger data vectors.

5.4 Application to Tension Calibration

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

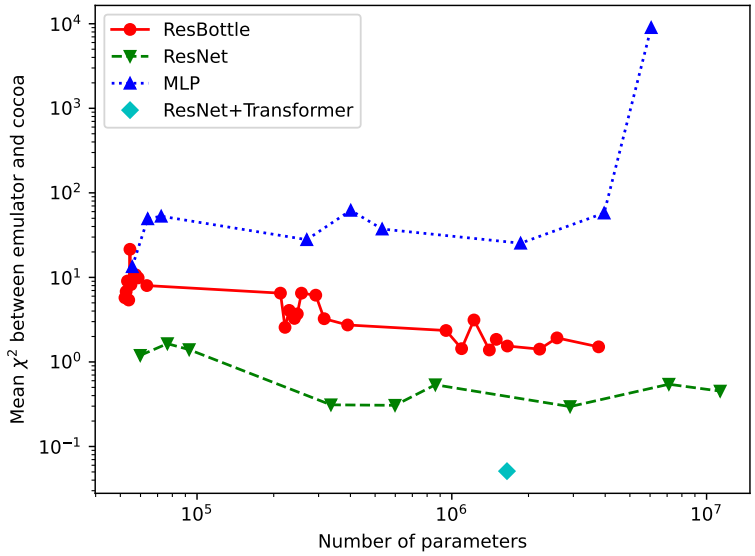


Figure 5.9: Average $\Delta\chi^2$ of a $T = 64$ chain using emulators trained on a $T = 128$ chain.

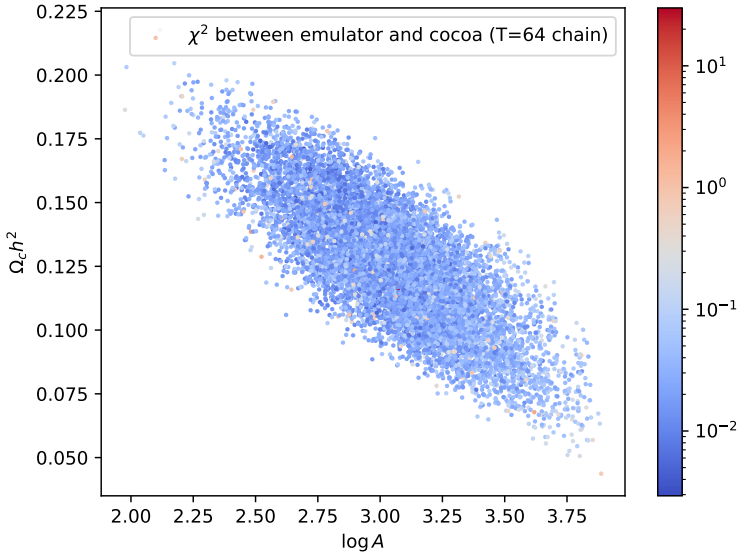


Figure 5.10: Marginalized distribution of $\Delta\chi^2$ of a $T = 64$ chain using the model in figure 5.11 trained on a $T = 128$ chain.

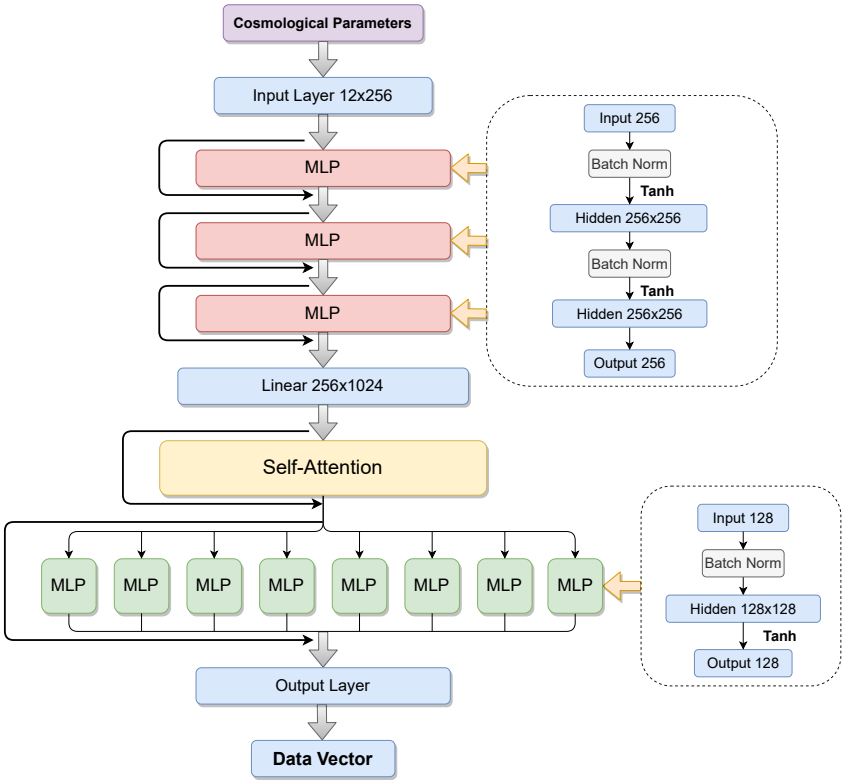


Figure 5.11: Full NN architecture used for training on high-temperature chains. Comprised of three ResBlocks of size 256, and a transformer block with 8 heads, each with a size of 128.

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 results of the tension metrics will allow us to gauge how reasonable it may be for the tension to be resolved by additional identical measurements.

We will be shifting the LSST fiducial cosmology to $\pm 5\sigma$ in both the σ_8 and Ω_m directions. At each of these shifts, we will compute 256 noise realizations on LSST cosmic shear and Planck TTTEEE using their respective data covariance. Finally, the tension will be computed using the methods described in section 3.2. Due to time constraints, the chains for the shifts will not be presented in this thesis. We will, however, demonstrate this method using the fiducial cosmologies of LSST and Planck.

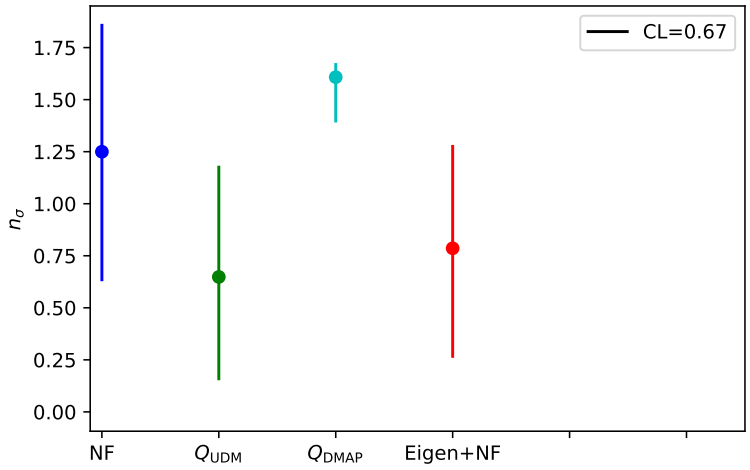


Figure 5.12: Parameter-based tension metrics with error.

The autocorrelation τ for the MCMC sampler was measured to be $\tau \sim 1000$, thus we need to run the chain for about 50000 steps (50τ) using the ensemble sampler in EMCEE. Each chain has 120 walkers, resulting in a total runtime of just 8 hours on a single 40 core CPU. The joint chains used for $Q_{\rm UDM}$ were run for just 5000 steps, since that is sufficient for the mean and covariance to converge.

We generate noise realizations for Planck around the Planck 2018 baseline low-ℓ TTTEEE best-fit cosmology [20]. We utilize COSMOPOWER to emulate the CMB power spectra [43, 44]. The Planck chains are run with 120 walkers for 50000. Although the autocorrelation is much lower than the LSST autocorrelation, there needs to be enough data in the tails to accurately estimate the tension.

We see that, up to noise realizations, the different tension metrics are nearly indistinguishable. Both Planck and LSST are nearly Gaussian, but for more non-Gaussian posteriors (such as DES), the difference between the parameter difference and other metrics may be more significant. We also see that the tension can be reduced by the experimental noise, but cannot be reduced to 0. It remains to compute the tension for noise realizations on the cosmologies with shifted Ω_m and σ_8.

The result for $Q_{\rm UDM}$ is significantly lower than the other metrics. This is likely

because it makes the assumption that the posteriors are Gaussian, whereas the tension can hide in the non-Gaussian parts of the posterior. Additionally, the error bar on $Q_{\rm DMAP}$ is much smaller. This requires some investigation into the distribution of the maximum a posteriori of the noise realizations. While noise realizations on the data generate large shifts in the posterior, the same is not necessarily true for the maximum a posteriori.

5.5 Conclusion

We are able to implement a NN emulator that works on temperatures as high as 128 by using the ChatGPT inspired transformer block. Using the emulators, we can ensure that the posteriors are robust for noise realizations on the data. In fact, using a temperature of 128 ensures the emulator should be robust up to 10σ shifts in the data. Using these noise realizations, we can compute the error on the tension metrics.

A future study will be done creating emulators for different extension models which push the number of input parameters, as well as different data sets which push the number of output parameters. Using this baseline framework, we can also rapidly analyze smaller extensions, such as the geometry and growth split detailed in chapter 4.

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] Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, S. Basak, R. Battye, K. Benabed, J.-P. Bernard, M. Bersanelli, P. Bielewicz, J. J. Bock, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, J.-F. Cardoso, J. Carron, A. Challinor, H. C. Chiang, J. Chluba, L. P. L. Colombo, C. Combet, D. Contreras, B. P. Crill, F. Cuttaia, P. de Bernardis, G. de Zotti, J. Delabrouille, J.-M. Delouis, E. Di Valentino, J. M. Diego, O. Doré, M. Douspis, A. Ducout, X. Dupac, S. Dusini, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, Y. Fantaye, M. Farhang, J. Fergusson, R. Fernandez-Cobos, F. Finelli, F. Forastieri, M. Frailis, A. A. Fraisse, E. Franceschi, A. Frolov, S. Galeotta, S. Galli, K. Ganga, R. T. Génova-Santos, M. Gerbino, T. Ghosh, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gruppuso, J. E. Gudmundsson, J. Hamann, W. Handley, F. K. Hansen, D. Herranz, S. R. Hildebrandt, E. Hivon, Z. Huang, A. H. Jaffe, W. C. Jones, A. Karakci, E. Keihänen, R. Keskitalo, K. Kiiveri, J. Kim, T. S. Kisner, L. Knox, N. Krachmalnicoff, M. Kunz, H. Kurki-Suonio, G. Lagache, J.-M. Lamarre, A. Lasenby, M. Lattanzi, C. R. Lawrence, M. Le Jeune, P. Lemos, J. Lesgourgues, F. Levrier, A. Lewis, M. Liguori, P. B.

Lilje, M. Lilley, V. Lindholm, M. López-Caniego, P. M. Lubin, Y.-Z. Ma, J. F. Macías-Pérez, G. Maggio, D. Maino, N. Mandolesi, A. Mangilli, A. Marcos-Caballero, M. Maris, P. G. Martin, M. Martinelli, E. Martínez-González, S. Matarrese, N. Mauri, J. D. McEwen, P. R. Meinhold, A. Melchiorri, A. Mennella, M. Migliaccio, M. Millea, S. Mitra, M.-A. Miville-Deschênes, D. Molinari, L. Montier, G. Morgante, A. Moss, P. Natoli, H. U. Nørgaard-Nielsen, L. Pagano, D. Paoletti, B. Partridge, G. Patanchon, H. V. Peiris, F. Perrotta, V. Pettorino, F. Piacentini, L. Polastri, G. Polenta, J.-L. Puget, J. P. Rachen, M. Reinecke, M. Remazeilles, A. Renzi, G. Rocha, C. Rosset, G. Roudier, J. A. Rubiño-Martín, B. Ruiz-Granados, L. Salvati, M. Sandri, M. Savelainen, D. Scott, E. P. S. Shellard, C. Sirignano, G. Sirri, L. D. Spencer, R. Sunyaev, A.-S. Suur-Uski, J. A. Tauber, D. Tavagnacco, M. Tenti, L. Toffolatti, M. Tomasi, T. Trombetti, L. Valenziano, J. Valiviita, B. Van Tent, L. Vibert, P. Vielva, F. Villa, N. Vittorio, B. D. Wandelt, I. K. Wehus, M. White, S. D. M. White, A. Zacchei, and A. Zonca. Planck 2018 results. VI. Cosmological parameters. *A&A*, 641:A6, September 2020. ISSN 0004-6361, 1432-0746. doi: 10.1051/0004-6361/201833910. URL `http://arxiv.org/abs/1807.06209`. arXiv:1807.06209 [astro-ph].

[5] 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].

[6] 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.

[7] Scott Dodelson and Fabian Schmidt. *Modern Cosmology*. Academic Press, 2021. ISBN 978-0-12-815948-4.

[8] Antony Lewis. CAMB Notes. URL https://cosmologist.info/notes/CAMB.pdf.

[9] M.A. Troxel and Mustapha Ishak. The intrinsic alignment of galaxies and its impact on weak gravitational lensing in an era of precision cosmology. *Physics Reports*, 558:1–59, February 2015. ISSN 03701573. doi: 10.1016/j.physrep.2014.11.001. URL https://linkinghub.elsevier.com/retrieve/pii/S0370157314003974.

[10] 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. Petravick, 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].

[11] 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].

[12] Christopher M. Hirata and Uros Seljak. Shear calibration biases in weak lensing surveys. *Monthly Notices of the Royal Astronomical Society*, 343(2):459–480, August 2003. ISSN 0035-8711, 1365-2966. doi: 10.1046/j.1365-8711.2003.06683.x. URL http://arxiv.org/abs/astro-ph/0301054. arXiv:astro-ph/0301054.

[13] Bryan R. Gillis and A. N. Taylor. The Effects of Calibration on the Bias of Shear Measurements. *Monthly Notices of the Royal Astronomical Society*, 482(1):402–415, January 2019. ISSN 0035-8711, 1365-2966. doi: 10.1093/mnras/sty2637. URL http://arxiv.org/abs/1809.09540. arXiv:1809.09540 [astro-ph].

[14] 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 http://arxiv.org/abs/1903.07101. arXiv:1903.07101 [astro-ph].

[15] Wayne Hu. Animations, . URL http://background.uchicago.edu/~whu/metaanim.html.

[16] Joshua Frieman. CMB: Sound Waves in the Early Universe. URL https://astro.uchicago.edu/~frieman/Courses/A411-old/Lecture8/Lecture8-2010.pdf.

[17] Wayne Hu. Thermal History and Acoustic Kinematics, . URL http://background.uchicago.edu/~whu/Presentations/canaries1.pdf.

[18] Wayne Hu. Astro 448 Cosmic Microwave Background.

[19] A. Amon, D. Gruen, M. A. Troxel, N. MacCrann, S. Dodelson, A. Choi, C. Doux, L. F. Secco, S. Samuroff, E. Krause, J. Cordero, J. Myles, J. DeRose, R. H. Wechsler, M. Gatti, A. Navarro-Alsina, G. M. Bernstein, B. Jain, J. Blazek, A. Alarcon, A. Ferté, M. Raveri, P. Lemos, A. Campos, J. Prat, C. Sánchez, M. Jarvis, O. Alves, F. Andrade-Oliveira, E. Baxter, K. Bechtol, M. R. Becker, S. L. Bridle, H. Camacho, A. Campos, A. Carnero Rosell, M. Carrasco Kind, R. Cawthon, C. Chang, R. Chen, P. Chintalapati, M. Crocce, C. Davis, H. T. Diehl, A. Drlica-Wagner, K. Eckert, T. F. Eifler, J. Elvin-Poole, S. Everett, X. Fang, P. Fosalba, O. Friedrich, G. Giannini, R. A. Gruendl, I. Harrison, W. G. Hartley, K. Herner, H. Huang, E. M. Huff, D. Huterer, N. Kuropatkin, P.-F. Leget, A. R. Liddle, J. McCullough, J. Muir, S. Pandey, Y. Park, A. Porredon, A. Refregier, R. P. Rollins, A. Roodman, R. Rosenfeld, A. J. Ross, E. S. Rykoff, J. Sanchez, I. Sevilla-Noarbe, E. Sheldon, T. Shin, A. Troja, I. Tutusaus, T. N. Varga, N. Weaverdyck, B. Yanny, B. Yin, Y. Zhang, J. Zuntz, M. Aguena, S. Allam, J. Annis, D. Bacon, E. Bertin, S. Bhargava, D. Brooks, E. Buckley-Geer, D. L. Burke, J. Carretero, M. Costanzi, L. N. da Costa, M. E. S. Pereira, J. De Vicente, S. Desai, J. P. Dietrich, P. Doel, I. Ferrero, B. Flaugher, J. Frieman, J. García-Bellido, E. Gaztanaga, D. W. Gerdes, T. Giannantonio, J. Gschwend, G. Gutierrez, S. R. Hinton, D. L. Hollowood, K. Honscheid, B. Hoyle, D. J. James, R. Kron, K. Kuehn, O. Lahav, M. Lima, H. Lin, M. A. G. Maia, J. L. Marshall, P. Martini, P. Melchior, F. Menanteau, R. Miquel, J. J. Mohr, R. Morgan, R. L. C. Ogando, A. Palmese, F. Paz-Chinchón, D. Petravick, A. Pieres, A. A. Plazas Malagón, A. K. Romer, E. Sanchez, V. Scarpine, M. Schubnell, S. Serrano, M. Smith, M. Soares-Santos, E. Suchyta, G. Tarle, D. Thomas, C. To, and J. Weller. Dark Energy Survey Year 3 Results: Cosmology from Cosmic Shear and Robustness to Data Calibration. *Phys. Rev. D*, 105(2):023514, January 2022. ISSN 2470-0010, 2470-0029. doi: 10.1103/PhysRevD.105.023514. URL `http://arxiv.org/abs/2105.13543`. arXiv:2105.13543 [astro-ph].

[20] Planck 2018 Results: Cosmological Parameter Tables. *Phys. Rev. Lett.*, 121(22):221301, November 2018. ISSN 0031-9007, 1079-7114. doi: 10.1103/PhysRevLett.121.221301. URL `http://arxiv.org/abs/1810.05216`. arXiv:1810.05216 [astro-ph].

[21] Marc Kamionkowski and Adam G. Riess. The Hubble Tension and Early Dark Energy, November 2022. URL `http://arxiv.org/abs/2211.04492`. arXiv:2211.04492 [astro-ph, physics:gr-qc, physics:hep-ph].

[22] Shinji Tsujikawa. Quintessence: A Review. *Class. Quantum Grav.*, 30 (21):214003, November 2013. ISSN 0264-9381, 1361-6382. doi: 10.1088/0264-9381/30/21/214003. URL http://arxiv.org/abs/1304.1961. arXiv:1304.1961 [astro-ph, physics:gr-qc, physics:hep-ph, physics:hep-th].

[23] Shivam Agrahari. Monte Carlo Markov Chain (MCMC) explained, July 2021. URL https://towardsdatascience.com/monte-carlo-markov-chain-mcmc-explained-94e3a6c8de11.

[24] 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].

[25] W. J. Handley, M. P. Hobson, and A. N. Lasenby. PolyChord: nested sampling for cosmology. 450(1):L61–L65. ISSN 1745-3933, 1745-3925. doi: 10.1093/mnrasl/slv047. URL http://arxiv.org/abs/1502.01856.

[26] Justin L Tobias. The Metropolis-Hastings Algorithm.

[27] Boyang Zhao. Bayesian inference using Markov Chain Monte Carlo with Python (from scratch and with PyMC3), January 2021. URL https://boyangzhao.github.io/posts/mcmc-bayesian-inference.

[28] Jennifer Helsby. Monte Carlo Markov Chains: A Brief Introduction and Implementation. (5).

[29] Sebastian Seehars, Sebastian Grandis, Adam Amara, and Alexandre Refregier. Quantifying concordance in cosmology. *Phys. Rev. D*, 93(10): 103507, May 2016. ISSN 2470-0010, 2470-0029. doi: 10.1103/PhysRevD.93.103507. URL https://link.aps.org/doi/10.1103/PhysRevD.93.103507.

[30] 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].

[31] 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].

[32] John Lee. Smooth Manifolds.

[33] 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. Sáchez, 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 ΛCDM. *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].

[34] 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].

[35] Jesus Torrado and Antony Lewis. Cobaya: Code for bayesian analysis of hierarchical physical models. 2021(5):057. ISSN 1475-7516. doi: 10.1088/1475-7516/2021/05/057. URL `http://arxiv.org/abs/2005.05290`.

[36] Elisabeth Krause and Tim Eifler. CosmoLike - cosmological likelihood analyses for photometric galaxy surveys. 470(2):2100–2112. ISSN 0035-8711, 1365-2966. doi: 10.1093/mnras/stx1261. URL `http://arxiv.org/abs/1601.05779`.

[37] What are Neural Networks? | IBM. URL `https://www.ibm.com/topics/neural-networks`.

[38] Supranta S. Boruah, Tim Eifler, Vivian Miranda, and Sai Krishanth P.

M. Accelerating cosmological inference with gaussian processes and neural networks – an application to LSST y1 weak lensing and galaxy clustering. URL http://arxiv.org/abs/2203.06124.

[39] Chun-Hao To, Eduardo Rozo, Elisabeth Krause, Hao-Yi Wu, Risa H. Wechsler, and Andrés N. Salcedo. LINNA: Likelihood inference neural network accelerator. 2023(1):016. ISSN 1475-7516. doi: 10.1088/1475-7516/2023/01/016. URL http://arxiv.org/abs/2203.05583.

[40] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition, December 2015. URL http://arxiv.org/abs/1512.03385. arXiv:1512.03385 [cs].

[41] Molly Ruby. How ChatGPT Works: The Models Behind The Bot, February 2023. URL https://towardsdatascience.com/how-chatgpt-works-the-models-behind-the-bot-1ce5fca96286.

[42] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention Is All You Need, December 2017. URL http://arxiv.org/abs/1706.03762. arXiv:1706.03762 [cs].

[43] A. Spurio Mancini, D. Piras, J. Alsing, B. Joachimi, and M. P. Hobson. COSMOPOWER: emulating cosmological power spectra for accelerated Bayesian inference from next-generation surveys. *Monthly Notices of the Royal Astronomical Society*, 511(2):1771–1788, February 2022. ISSN 0035-8711, 1365-2966. doi: 10.1093/mnras/stac064. URL http://arxiv.org/abs/2106.03846. arXiv:2106.03846 [astro-ph].

[44] Heather Prince and Jo Dunkley. Data compression in cosmology: A compressed likelihood for Planck data. *Phys. Rev. D*, 100(8):083502, October 2019. ISSN 2470-0010, 2470-0029. doi: 10.1103/PhysRevD.100.083502. URL http://arxiv.org/abs/1909.05869. arXiv:1909.05869 [astro-ph].