

Thesis Notes

Evan Saraivanov

February 7, 2023

Contents

1	Cosmology	2
1.1	Introduction	2
1.2	General Relativity	2
1.2.1	Smooth Manifolds	2
1.2.2	Curvature	3
1.2.3	Einstein's Field Equation	3
1.3	Λ CDM: The Standard Model of Cosmology	4
1.3.1	The FLRW Metric and Dark Energy	4
1.3.2	Perturbations of the FLRW Metric	5
2	Connecting to Experiments: Weak Lensing and the Cosmic Microwave Background	6
2.1	Weak Lensing	6
2.1.1	The Linear Matter Power Spectrum	6
2.1.2	Galaxy Density	7
2.1.3	Galaxy Shear	8
2.1.4	Weak Lensing Correlation Functions	10
2.1.5	Systematics	10
2.2	CMB	12
2.2.1	TT,TE,EE Power Spectra	12
2.2.2	Systematics	12
3	Tension	13
3.1	Tension Metrics	13
3.1.1	Metric 1: Parameter Difference	13
3.1.2	Metric 2: Parameter Difference in Update Form	14
3.1.3	Metric 3: Goodness of Fit Degradation	14
3.1.4	Metric 4: Eigentension	14
3.1.5	Evidence-Based Metrics	14
3.1.6	Interpreting the Results	15
3.1.7	DES v Planck Results	15
3.1.8	Building from previous results	15
3.2	Markov Chain Monte Carlo	15
3.2.1	The Metropolis-Hastings Algorithm	15
3.3	Data Emulators	16
3.3.1	LSST Emulator	16
3.3.2	Planck C_ℓ Emulator	16
3.4	Normalizing Flows	16
4	Approximating Tension Between LSST and Planck	18
4.1	Noise Realizations of the Data	18
4.1.1	Data Emulators	18
4.1.2	Results	18
4.2	Tension From Shifted Cosmologies	18
4.2.1	Approximated s8	18
4.2.2	Results	18

Chapter 1

Cosmology

1.1 Introduction

With Hubble's discovery of the expanding universe, there has been great efforts to understand this expansion and the evolution of this expansion (and of the universe as a whole). A related idea, General Relativity has described the importance of the geometry of the universe. There are, in general, three possible geometries:

- A flat geometry which is equivalent to Euclidean space with zero curvature.
- An open geometry with constant negative curvature (Anti-de Sitter space).
- A Closed geometry with constant positive curvature (de Sitter space)

Given the relationship between curvature and the energy-momentum tensor given by the Einstein field equations, it seems reasonable to assume that any non-zero energy means the universe is not flat. Through quantum field theory (QFT) the Casimir effect shows that the vacuum has non-zero energy density, and through astronomical observation of distant galaxies it can be seen that our universe is flat. At first these two observations contradict each other. However, by introducing another term to the Einstein field equations, this discrepancy can be resolved.

$$G_{\mu\nu} - \underbrace{\Lambda g_{\mu\nu}}_{\text{new term}} = T_{\mu\nu} \quad (1.1)$$

The constant Λ is called the *cosmological constant* which can absorb the contributions from the vacuum energy, allowing for a flat universe and non-zero vacuum energy density. A natural question to ask is 'what is the source of the cosmological constant?'

1.2 General Relativity

To align with my interest I will write this section from a mathematical perspective and give a physics interpretation after. In this section, I will derive Einstein's field equation. First starting with defining smooth structures on a manifold. Afterwards I move into curvature before using the Bianchi identity on the Riemann tensor to derive the equation governing General Relativity. I will also discuss some properties of Einstein's Field Equation.

1.2.1 Smooth Manifolds

The basic structure of our universe according to general relativity is that the universe is a smooth 4-manifold. This already contains a lot of information so let's break it down. To begin, one starts with defining a topological manifold. Given a space X with topology τ , X is an n dimensional topological manifold if

- X is Hausdorff. For each $p, q \in X$ there exists neighborhoods $p \in U$, $q \in V$, $U, V \in \tau$ such that $U \cap V = \emptyset$.

- X is second countable. There exists a countable basis for the topology τ .
- X is locally Euclidean. For all $U \in \tau$ there exists a homeomorphism $\phi : U \rightarrow V$ such that $V \subset \mathbb{R}^n$.

The last condition is one that hints at a differentiable structure because of known calculus in \mathbb{R}^n . To solidify this, consider the pair (U, ϕ) called a chart. Given two charts (U, ϕ) and (V, ψ) with $U \cap V \neq \emptyset$, we can create the transition map $\psi \circ \phi^{-1} : \phi(U \cap V) \rightarrow \psi(U \cap V)$. The two charts are smoothly compatible if $U \cap V = \emptyset$ or the transition map is a homeomorphism.

Manifolds are a generalization of surfaces to arbitrary dimension. To make this a true generalization one needs a way to compute derivatives on a manifold. Let M be a manifold with $x \in M$ and let $\gamma : (-\epsilon, \epsilon) \rightarrow M$ be a curve on M with $\gamma(0) = x$. A tangent vector of M is the equivalence class of curves, $[\gamma(t)]$, with equal $\gamma'(0)$. The tangent space of M at x , denoted $T_x M$ is the set of all tangent vectors at x . The tangent bundle of M is the disjoint union of all tangent spaces of M , denoted $TM = \bigsqcup_{x \in M} T_x M$.

The tangent bundle is constructed as a trivial bundle, so one can easily define $(T_x M, x)$ as fibers of the bundle and a projection $\pi : TM \rightarrow M$ with $\pi(T_x M, x) = x$. Thus sections take x to the tangent space at x .

With this structure, one can now infinitesimally take a vector in a tangent space and transport it to an adjacent tangent space, thus giving rise to the differential structure sought in the previous paragraph. Given a vector field v on M , D_v is a connection on M if the usual derivative properties hold (linearity over $C^\infty(M)$ and Leibniz's law).

Usual definition of vector field: a function $v : C^\infty(M) \rightarrow C^\infty(M)$

1.2.2 Curvature

Suppose one have a (semi-) Riemannian manifold M with metric g and tangent bundle TM . An *affine connection* is a map

$$\nabla : \Gamma(TM) \times \Gamma(TM) \rightarrow \Gamma(TM) \quad (1.2)$$

$$(X, Y) \mapsto \nabla_X Y \quad (1.3)$$

That is, it parallel transports the vector field Y along the connection ∇ in the direction of vector field X . From this, one can write the affine geodesic equation for a path $\gamma(t)$

$$\nabla_{\dot{\gamma}} \dot{\gamma}(t) = 0 \quad (1.4)$$

Thus, a geodesic is a path such that its tangent vector is parallel translated. Since one observes their world with coordinates, in physics it is often more instructive to work this out in a specific set of coordinates x^μ . Thus this can be written as

$$\ddot{\gamma}^\mu + \Gamma_{\rho\lambda}^\mu \dot{\gamma}^\rho \dot{\gamma}^\lambda \quad (1.5)$$

In general, affine connections have three components: a curvature component, a torsion component, and a *metricity* component.

1.2.3 Einstein's Field Equation

In a curved spacetime, conservation of energy is written

$$\nabla^\mu T_{\mu\nu} = 0 \quad (1.6)$$

When deriving Einstein's equation, one wants to find a divergence-less tensor that depends only on the geometry. The following procedure follows closely one would find the classical Yang-Mills equation for gauge fields. The Riemann curvature is anti-symmetric in the first two lower indices, so the Riemann tensor is like a $GL(TM)$ valued differential 2 form, and the Bianchi identity is

$$d_\nabla R = 0 \quad (1.7)$$

Explicitly writing this becomes

$$\nabla_\alpha R_{\beta\gamma\delta}^\alpha + \nabla_\beta R_{\gamma\alpha\delta}^\alpha + \nabla_\gamma R_{\alpha\beta\delta}^\alpha = 0$$

$$\begin{aligned} & \nabla_\alpha R_{\beta\gamma\delta}^\alpha + \nabla_\beta R_{\gamma\alpha\delta}^\alpha - \nabla_\gamma R_{\beta\alpha\delta}^\alpha \\ & \nabla_\alpha R_{\beta\gamma\delta}^\alpha + \nabla_\beta R_{\gamma\delta}^\alpha - \nabla_\gamma R_{\alpha\delta}^\alpha \end{aligned}$$

Multiplying by $g^{\beta\delta}$ and doing some relabelling/contractions of internal indices one finds

$$\nabla^\alpha (R_{\gamma\alpha} - \frac{1}{2}g_{\gamma\alpha}R) \equiv \nabla^\alpha G_{\gamma\alpha} = 0 \quad (1.8)$$

There is one more divergencless tensor, the metric tensor $g_{\mu\nu}$, so one can write

$$G_{\mu\nu} - \Lambda g_{\mu\nu} = 8\pi\kappa T_{\mu\nu} \quad (1.9)$$

1.3 Λ CDM: The Standard Model of Cosmology

1.3.1 The FLRW Metric and Dark Energy

In general, metrics allow one to attribute distances between points in a space as $d = g_{\mu\nu}x^\mu x^\nu$. The ‘flat’ metric for relativity is the *Minkowski metric* given by $\text{diag}(-1, 1, 1, 1)$, however, a metric to describe the expanding universe is given by the *FLRW metric* by $d^2 = t^2 - a^2(t)s^2$ with s^2 the standard Euclidean distance in \mathbb{R}^3 . What is immediately apparent from the FLRW metric is that spacial slices of the $d = 4$ spacetime remain curvature free under the expansion of the universe describe by the scale factor $a^2(t)$.

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

$$\chi = \int_{t_0}^t \frac{1}{a(t')} dt' \quad (1.10)$$

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

$$\eta = \int_0^t \frac{1}{a(t')} dt' \quad (1.11)$$

Since we can’t see anything beyond this distance, it is often called the *comoving horizon*. There is one last useful distance to define, the *angular distance* which is inferred by the angle subtended by two objects. This relates distances to the geometry discuss in the first section where the measured distance will be the radial distance D

$$D_A = \begin{cases} R & K = 0 \\ R \sin(D/R) & K > 0 \\ R \sinh(D/R) & K < 0 \end{cases} \quad (1.12)$$

When describing the structure of the universe, I will make a few assumptions (which hold up to small perturbations):

- Homogeneity. The cosmology describing the universe does not depend on location.
- Isotropy. The cosmology describing the universe does not depend on direction.

These two conditions for what is sometimes referred to as *the cosmological principle*. In general, they don’t hold on small scales, however averaging over a sufficiently large distance these assumptions give an accurate description.

If we examine the 00 component of Einstein’s Field Equation, the result is the *first Friedman equation* (note that ρ is shorthand for $\sum_i \rho_i$ and $R(G)$ to note that R depends on the geometry of spacial slices)

$$H^2(a) + \frac{1}{a^2 R^2(G)} = \frac{8\pi G}{3} \rho \quad (1.13)$$

We can interpret the second term on the left (which is spacial curvature) as some density asociated with the curvature ρ_k . If we divide by ρ_{crit} at $z = 0$ / $a = 1$ we find the usual form.

$$\omega + \omega_k = 1 \quad (1.14)$$

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

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P) \quad (1.15)$$

1.3.2 Perturbations of the FLRW Metric

Upon looking at the sky, one should realize that our universe is not homogenous. We can see clusters of stars and galaxies as well as regions devoid of matter. Thus, the FLRW metric is taken as a baseline approximation which we perturb in order to correct for our observations. In general relativity, the Christoffel symbols are the gravitational potential generating the gravitational field. The definition of the Christoffel symbols in terms of the metric is

$$\Gamma_{\mu\nu\lambda} = \frac{1}{2} (\partial_\lambda g_{\mu\nu} + \partial_\nu g_{\mu\lambda} - \partial_\mu g_{\nu\lambda}) . \quad (1.16)$$

The FLRW metric only has one non-zero partial derivative, the derivative with respect to t . Thus we if $\lambda = 0$ or $\nu = 0$, we have

$$\begin{aligned} \Gamma_{ij0} &= \Gamma_{i0j} = \delta_{ij} a \dot{a} \\ \Gamma_{j0}^i &= \delta_{ij} \frac{\dot{a}}{a} \\ \Gamma_{ij}^0 &= \delta_{ij} a \dot{a} \end{aligned} \quad (1.17)$$

Thus, with two unique Christoffel symbol values, we can perturb the metric using two potentials Φ and Ψ . So the perturbed metric becomes

$$d^2 = t^2(1 + 2\Psi) - a^2(t)s^2(1 - 2\Phi) \quad (1.18)$$

where the perturbations are much smaller than 1. For the remainder of this thesis, calculations involving the geometry of the universe will be based on this perturbed metric.

Chapter 2

Connecting to Experiments: 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.

Weak lensing occurs when light travels with curved trajectories. Contrary to the purely Newtonian approach to gravity, in General Relativity light travels along geodesics with vanishing infinitesimal distances. Thus the time a photon has travelled is given by

2.1.1 The Linear Matter Power Spectrum

As we have seen before, we consider deviations from the FLRW metric using the scalar fields Φ and Ψ . In the absence of anisotropic stress, $\Phi = -\Psi$. Throughout time in the universe, we consider there to be three distinct epochs/regions:

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

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

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

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

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

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

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

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

Thus we can write the gravitational potential at all times as

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

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

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

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

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

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

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

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

2.1.2 Galaxy Density

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

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

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

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

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

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

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

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

2.1.3 Galaxy Shear

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Now, putting everything together, plugging these results into eq (2.14) gives

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

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

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

Interestingly, $\delta^{ij} - \hat{p}^i \hat{p}^j$, is the projection operator of the j -th direction onto the i -th. So, if we orient our axes so that the photon approximately travels along the j -th direction, we can sum of i and k so that the orthogonal components of the momentum are evaluated by

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

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

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

Integrating this equation gives us

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Which relates to the shear components by

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

This connects the observable shear to the theoretical lensing potential.

2.1.4 Weak Lensing Correlation Functions

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

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

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

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

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

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

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

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

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

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

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

2.1.5 Systematics

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

Photometric Redshift Uncertainty

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

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

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

Intrinsic Alignment of Galaxies

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

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

Where the coefficients C_a are given by

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

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

Galaxy Bias

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

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

then consider a multiplicative deviation

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

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

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

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

where i here denotes the redshift bin

Shear Calibration

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

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

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

Point Mass

Lensing Magnification

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

2.2 CMB

2.2.1 TT,TE,EE Power Spectra

2.2.2 Systematics

Chapter 3

Tension

As discussed before (where?), standard Λ CDM can be parameterized by 6 parameters, $(A_s, n_s, H_0, \tau, \Omega_b, \Omega_c)$. As usual, this is not the only parameterization of Λ CDM. A_s describes the amplitude of the matter power spectrum. By integrating the matter power spectrum up to a scale of $8h^{-1}\text{Mpc}$ we can find a parameter denoted σ_8 defined as

$$\sigma_8^2 = \int_0^\infty P(k, r) \left(\frac{3j_1(kr)}{kr} \right)^2 \frac{k^2}{2\pi^2} dk \bigg|_{r=8} \quad (3.1)$$

which can be described as the amplitude of matter fluctuations at the scale $8h^{-1}\text{Mpc}$. In addition, we can add up all of the matter densities to get

$$\Omega_m = \sum_i \Omega_i = 1 - \Omega_k \quad (3.2)$$

which describes the matter density. As it turns out, measurements of the early universe (CMB and high redshift observation) and late universe (galaxy observations and low redshift observation) disagree on measurements of H_0 and $S_8 = \sigma_8 \sqrt{\Omega_m}/0.3$. These individual parameters are in tension between experiments, but the driving questions of this project is: how much tension is there between measurements of the entire Λ CDM parameters? Are there models that can resolve this tension? What is the best/most reliable way to measure the tension?

3.1 Tension Metrics

To evaluate the difference between two posteriors, we establish *tension metrics* which quantify the difference between the posteriors.

3.1.1 Metric 1: Parameter Difference

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

$$\mathcal{P}(\Delta\theta) = \int \mathcal{P}(\theta_A) \mathcal{P}(\theta_A - \Delta\theta) d\theta_A \quad (3.3)$$

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

$$\text{PTE} = \int_{\mathcal{P}(\Delta\theta) > \mathcal{P}(0)} \mathcal{P}(\Delta\theta) d\Delta\theta \quad (3.4)$$

To actually compute the integral we use the normalizing flow to learn the posterior and perform MCMC integration. The integration error is determined using the Clopper-Pearson interval on the binomial distribution.

3.1.2 Metric 2: Parameter Difference in Update Form

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

$$Q_{\text{UDM}} = (\mu_A - \mu_{A+B})^T (C_A - C_{A+B})^{-1} (\mu_A - \mu_{A+B}) \quad (3.5)$$

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

3.1.3 Metric 3: Goodness of Fit Degradation

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

$$Q_{\text{DMAP}} = 2\mathcal{L}_A(\hat{\theta}_A) + 2\mathcal{L}_B(\hat{\theta}_B) - 2\mathcal{L}_{A+B}(\hat{\theta}_{A+B}) \quad (3.6)$$

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

$$d = N - \text{tr}(C_{\Pi}^{-1} C_{\mathcal{P}}) \quad (3.7)$$

3.1.4 Metric 4: Eigentension

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

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

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

(here is a good place to compare with other metrics)

3.1.5 Evidence-Based Metrics

- Bayesian evidence ratio given by

$$R = \frac{\mathcal{E}_{AB}}{\mathcal{E}_A \mathcal{E}_B} \quad (3.8)$$

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

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

$$\log S = \log R - \log I \quad (3.9)$$

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

$$\log I = \mathcal{D}_A + \mathcal{D}_B - \mathcal{D}_{AB} \quad (3.10)$$

$$\mathcal{D} = \int \mathcal{P} \log \left(\frac{\mathcal{P}}{\Pi} \right) \quad (3.11)$$

3.1.6 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

$$n_\sigma = \sqrt{2}\text{Erf}^{-1}(P) \quad (3.12)$$

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

$$\begin{aligned} \mathcal{P}(\Delta\theta) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\theta^2/2} e^{-(\theta-\Delta\theta)^2/2} d\theta \\ &= \frac{1}{2\pi} \cdot \sqrt{\pi} e^{-(\Delta\theta)^2/4} \\ &= \frac{1}{\sqrt{4\pi}} e^{-(\Delta\theta)^2/4} \end{aligned} \quad (3.13)$$

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

$$\Delta = \int_{-a}^a e^{-(\Delta\theta)^2/4} d\Delta\theta \quad (3.14)$$

Thus we can compare the analytic result to the computational result for each of the metrics to see which ones are most accurate. (DO THIS FOR THE NEW METRICS, NOT JUST NF).

3.1.7 DES v Planck Results

A subset of these excersizes have been performed previously by the Dark Energy Survey (DES) against the planck results.

3.1.8 Building from previous results

3.2 Markov Chain Monte Carlo

Interestingly, MCMC algorithms have heavy analogies with statistical mechanics which are useful to demonstrate the concept. To examine this, lets first define what a Markov Chain is.

Definition 1. A sequence X_1, \dots, X_n of random elements is a Markov Chain if the conditional distribution X_{n+1} depends only on X_n . The set in which X_i take values is called the state space of the chain.

3.2.1 The Metropolis-Hastings Algorithm

Suppose we want to sample from a distribution $p(x)$. $p(x)$ can be high dimensional and is generally difficult to calculate (evidence is hard to compute since it requires integration over the entire parameter space). The goal is to use Markov Chains to sample from $p(x)$ without needing to compute the evidence. This will be represented as a path through state space until the chain reaches a stable point (stationary state).

We start with a proposal distribution $g(x_n)$. Sample from the proposal distribution to find the next state x_{n+1} with probability $g(x_{n+1}|x_n)$. This transition from state n to state $n+1$ must follow the *detailed balance condition*

$$p(x_n)g(x_{n+1}|x_n)A(x_n \rightarrow x_{n+1}) = p(x_{n+1})g(x_n|x_{n+1})A(x_{n+1} \rightarrow x_n) \quad (3.15)$$

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

$$\pi(x_n)\mathcal{L}(x_n)g(x_{n+1}|x_n)A(x_n \rightarrow x_{n+1}) = \pi(x_{n+1})\mathcal{L}(x_{n+1})g(x_n|x_{n+1})A(x_{n+1} \rightarrow x_n) \quad (3.16)$$

$$\Rightarrow \frac{A(x_n \rightarrow x_{n+1})}{A(x_{n+1} \rightarrow x_n)} = \frac{\pi(x_{n+1})\mathcal{L}(x_{n+1})g(x_n|x_{n+1})}{\pi(x_n)\mathcal{L}(x_n)g(x_{n+1}|x_n)} \equiv R_{n,n+1} \quad (3.17)$$

This allows us to define the acceptance probability as

$$A(x_n \rightarrow x_{n+1}) = \min(1, R_{n,n+1}) \quad (3.18)$$

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

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

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

3.3 Data Emulators

Traditional methods of computing likelihoods (e.g. COSMOSIS) require an immense number of CPU hours, and since we need tens of thousands of chains, we cannot rely on these traditional methods. To accelerate likelihood computation, and thus the MCMC process, we employ emulators which are neural networks that map the cosmological parameters to data vectors. The map is highly non-linear and thus a straightforward analytic mapping is not known. The neural network architecture used is similar for each experiment.

3.3.1 LSST Emulator

Unfortunately, we do not have access to the training samples for this emulator. We can however approximate the training region. First, we must define what tempered MCMC means. As discussed in the previous section, an MCMC decides whether to accept or reject a point based on the detailed balance condition. In our case, we want to ensure the training samples cover a wide range in the parameter space, so we can modify the posterior by raising it to a power T called the tempering factor. The detailed balance condition becomes

$$\frac{A(x_n \rightarrow x_{n+1})}{A(x_{n+1} \rightarrow x_n)} = \left[\frac{\pi(x_{n+1})\mathcal{L}(x_{n+1})}{\pi(x_n)\mathcal{L}(x_n)} \right]^T \frac{g(x_n|x_{n+1})}{g(x_{n+1}|x_n)} \quad (3.19)$$

It is believed the samples were generated via MCMC with a posterior tempering of 0.5. This gives us the following distribution

3.3.2 Planck C_ℓ Emulator

Fortunately the training samples for COSMOPOWER are available for download from google drive.

3.4 Normalizing Flows

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

The implementation in tensorflow uses *bijectors* which implements a local diffeomorphism between a manifold M and a target manifold N (which are our parameter spaces), i.e. $\phi : M \rightarrow N$ such that ϕ is differentiable and injective. In tensorflow it has three operations, Forward, Inverse, and `log_deg_jacobian`, which are exactly the three we want. By constructing a bijector for each masked input, the full normalizing map can be constructed.

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

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

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

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

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

$$\begin{aligned}\dot{\gamma}^i(t) &= v^i(\gamma(t)) \\ \gamma^i(0) &= c^i\end{aligned}\tag{3.20}$$

This can be solved by $\gamma^i(t) = \int_{U \subset X} v^i(\gamma(t)) \text{vol.}$. Since v is a tangent vector field, there exists a curve ϕ with $\phi' = v$, so the solution is $\gamma^i(t) = \phi^i(t) + C^i$. The integration constants are fixed by the initial condition $C^i = -c^i$, hence any two functions satisfying this system of ODE’s, including α and β , are equivalent for all $t \in I$. \square

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

Chapter 4

Approximating Tension Between LSST and Planck

4.1 Noise Realizations of the Data

4.1.1 Data Emulators

4.1.2 Results

4.2 Tension From Shifted Cosmologies

4.2.1 Approximated s8

4.2.2 Results

The isotropy condition means that the universe should have 0 net momentum, and by assumming the universe is smooth the energy-momentum tensor can be written

$$T_{\nu}^{\mu} = \begin{pmatrix} -\mathcal{E} & 0 & 0 & 0 \\ 0 & \mathcal{P} & 0 & 0 \\ 0 & 0 & \mathcal{P} & 0 \\ 0 & 0 & 0 & \mathcal{P} \end{pmatrix} \quad (4.1)$$

The usual conservation law holds

$$\nabla_{\mu} T_{\nu}^{\mu} = 0 \Rightarrow \partial_t \mathcal{E} + \frac{\dot{a}}{a} (3\mathcal{E} + 3\mathcal{P}) = 0 \quad (4.2)$$

We can use the geodesic equation to examine how the energy of the massless particles evolves through time.