Using Genetic Algorithms to Approximate Cosmological Quantities

A Thesis Presented

by

Amritpal Nijjar

to

The Graduate School

in Partial Fulfillment of the Requirements

for the Degree of

Master of Arts

in

Physics

Stony Brook University

May 2025

Abstract of the Thesis

Using Genetic Algorithms to Approximate Cosmological Quantities

by

Amritpal Nijjar

Master of Arts

in

Physics

Stony Brook University

2025

Currently, on the eve of a golden age in cosmology, it is more important than ever that our theoretical apparatuses are robust and ready to be applied to a range of proposed alternate cosmological models. In the course of testing alternate cosmological models, it is necessary to compute key cosmological quantities, such as the Cosmic Microwave Background (CMB) and the matter transfer function, fast and accurately in order to test theoretical predictions against observational data. Though analytic and semi-analytic treatments exist for such quantities, they are often lengthy and involved. This thesis explores the possibility of utilizing genetic algorithms, a subclass of machine learning, towards compressing analytic formulations of key cosmological quantities. This thesis is presented in three steps, first by developing the physics behind the CMB power spectrum, and detailing its well known semi-analytic formulation. Next machine learning techniques are introduced, namely genetic algorithms and symbolic regression, and their strengths are discussed. Genetic algorithms are then utilized

to develop a compressed physics-informed analytic approximation to the matter transfer function. Lastly, this work explores the possibility of modeling the CMB lensing tail using symbolic regression.

Contents

List of Figures **iv**

1 The Cosmic Microwave Background **1**
1.1 Part 1: Perturbing the FRW Universe 2
1.1.1 Preliminary Notes 3
1.1.2 General Relativity 4
1.1.3 The Boltzmann Equation 9
1.1.4 The CMB Power Spectrum 15
1.2 Part 2: Towards an Analytic Expression 17
1.2.1 Tight Coupling 18
1.2.2 Damping Effects 21
1.2.3 Free Streaming 23

2 Machine Learning **25**
2.1 Genetic Algorithms and Symbolic Regression 28
2.1.1 PySR 29

3 Compressing Analytic Expressions **33**
3.1 Part 1: The Matter Transfer Function 33
3.1.1 Computing $T(k)$ 37
3.1.2 Learning $T(k)$ 42
3.2 Part 2: The CMB Lensing Tail 48
3.2.1 The Lensing Tail 49
3.2.2 The Function Template 50
3.2.3 Symbolic Regression 51
3.2.4 Results 53

4 Conclusion **56**

Bibliography **58**

List of Figures

1.1 Temperature anisotropies in the CMB as measured by the Planck experiment. Figure from [1]. 2
1.2 The CMB power spectrum as measured by the Planck experiment as the red points, overlayed on top of the ΛCDM predicted spectrum in blue, for the Planck best fit parameters. Figure from [2]. 17

2.1 Two mathematical expressions represented as expression trees. Figure from [3]. 29
2.2 The crossover operation between two expressions. Figure from [4]. 31

3.1 Constraints on the matter power spectrum by various experiments. The solid black line shows the power spectrum as predicted by ΛCDM assuming linear perturbations. The impact of nonlinearities is depicted by the dotted line. Figure from [2]. . 36
3.2 CAMB generated matter transfer function for the Planck best fit cosmology, where $\Omega_b h^2 = 0.0224$ and $\Omega_c h^2 = 0.12$. 37
3.3 *Left*: BBKS approximation for the matter transfer function for the Planck best fit cosmology, where $\Omega_b h^2 = 0.0224$ and $\Omega_c h^2 = 0.12$. *Right*: The percent error between the BBKS approximation and the CAMB generated matter transfer functions for four cosmologies. 39
3.4 *Left*: EH approximation for the matter transfer function for the Planck best fit cosmology, where $\Omega_b h^2 = 0.0224$ and $\Omega_c h^2 = 0.12$. *Right*: The percent error between the EH approximated and CAMB generated matter transfer functions for four cosmologies. 42
3.5 *Left*: The GA learned approximation for the matter transfer function for the Planck best fit cosmology, where $\Omega_b h^2 = 0.0224$ and $\Omega_c h^2 = 0.12$. *Right*: The percent error between T_{GA} and T_{CAMB} for four sample cosmologies. 46

3.6 *Left*: The percent error of $T_{\rm GA}(k)$, $T_{\rm BBKS}(k)$ and $T_{\rm EH}(k)$ against $T_{\rm CAMB}(k)$ for the Planck best fit cosmology, where $\Omega_b h^2 = 0.0224$ and $\Omega_c h^2 = 0.12$. *Right*: The region of the left panel in the red dotted box enlarged. 47
3.7 *Left*: The average percent error of the EH transfer function for 100 cosmologies, with $\Omega_c h^2 \in [0.10, 0.14]$ and $\Omega_b h^2 \in [0.012, 0.03]$, and $H_0 = 67.4$. *Right*: The average percent error of the GA learned transfer function for the same cosmologies. 48
3.8 *Left*: $L_{\rm CAMB}(\ell)$ for the Planck best fit cosmology, where $\Omega_b h^2 = 0.0224$ and $\Omega_c h^2 = 0.12$. *Right*: $L_{\rm CAMB}(\ell)$ for four sample cosmologies for $\ell > 2000$. 50
3.9 $L_{\rm CAMB}(\ell)$ for the Planck best fit cosmology, where $\Omega_b h^2 = 0.0224$ and $\Omega_c h^2 = 0.12$ in yellow. The same curved smoothed according to eq. (3.37) is shown in blue. 52
3.10 *Right*: $L_{\rm CAMB}(\ell)$ for the Planck best fit cosmology, where $\Omega_b h^2 = 0.0224$ and $\Omega_c h^2 = 0.12$, in yellow, with the SR approximated lensing tail $L_{\rm SR}(\ell)$ in blue. The percent error between $L_{SR(\ell)}$ and $L_{\rm CAMB}(\ell)$ for three cosmologies. 54
3.11 The average percent error of $L_{\rm SR}(\ell)$ for 100 cosmologies, with $\Omega_c h^2 \in [0.10, 0.14]$ and $\Omega_b h^2 \in [0.012, 0.03]$, and $H_0 = 67.4$. . . 55

Chapter 1

The Cosmic Microwave Background

In this thesis, I am broadly interested in exploring the techniques used in constructing analytic and semi-analytic formulations of key cosmological quantities. To set the stage, and provide the cosmological context for the rest of the thesis, I will begin by discussing one of the hallmark cosmological observables, the Cosmic Microwave Background (CMB), and a powerful semi-analytic formulation of the CMB power spectrum, derived by Hu and Sugiyama in [5].

Since its first detection as a consistent noise signal in radio equipment permeating the sky [6], the CMB has gone on to become one of the most stringent tests of any cosmological model purporting to describe the age, evolution, and content of the universe. CMB photons, streaming nearly undisturbed for billions of years, give us the earliest snapshot possible of the universe. By studying how these photons are distributed on the sky, scientists can probe the fundamental physics governing the dynamics of the universe, and measure the viability of various cosmological models. Figure 1.1 shows the CMB as measured by the Planck experiment. The colored spots represent deviations from the global average temperature of CMB photons, fluctuations on the order of 10^{-5}. In this section, I will lay the ground work for understanding the nature of distribution of the temperature of these incoming photons, and describe the CMB power spectrum which collects the information of the fluctuations in temperature as a function of scale.

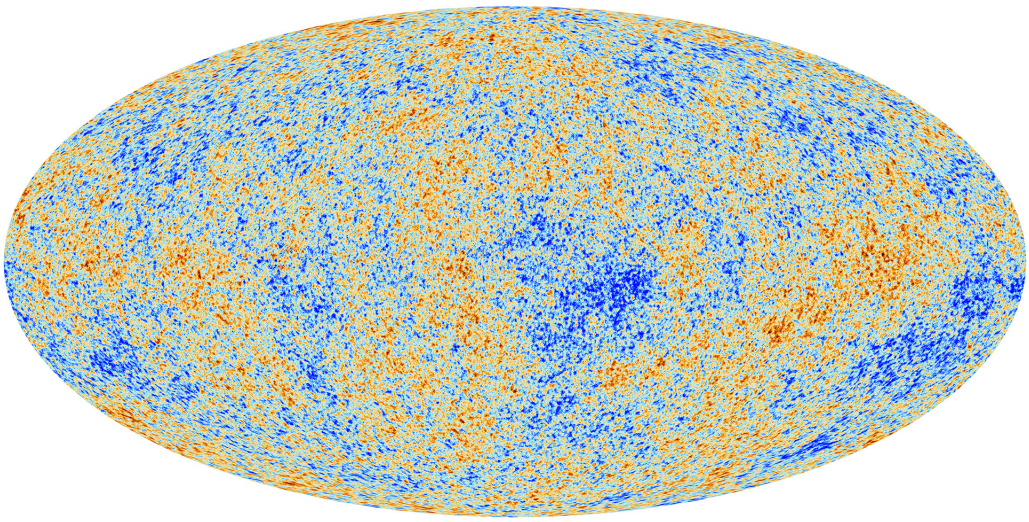


Figure 1.1: Temperature anisotropies in the CMB as measured by the Planck experiment. Figure from [1].

In Part 1 I introduce the relevant concepts required to understand the smooth FRW universe. I then explore the Boltzmann equation, and describe how the photon distribution function evolves in space and time, due to general relativity effects and Compton scattering off electrons. Following this, I describe how the photon distribution function of the previous section is related to the CMB Power Spectrum, the main observable we see today. In Part 2, I motivate a well-known semi-analytic formulation for the CMB power spectrum[5], treating the tight coupling, damping, and free-streaming of CMB photons as they travel from recombination to our sky today. [7].

1.1 Part 1: Perturbing the FRW Universe

The standard methodology adopted in cosmology is to consider features such as anisotropies in the CMB and inhomogeneities in the matter distribution (e.g. large scale structure) as perturbations against an otherwise smooth, homogeneous, and isotropic background. The assumptions of global homogeneity and isotropy are important, for example they form the backbone of the ΛCDM model of cosmology, the cosmological paradigm assumed herein. This assumption is reasonable, considering, for example, that the CMB temperature fluctuations are of the order 10^{-5}, whereas the average CMB temperature is $T = 2.726$ K [8]. In the following section, I describe the standard cosmological model, and use general relativity to provide a description of this smooth,

homogeneous, isotropic, background universe.

1.1.1 Preliminary Notes

Strong evidence suggests the universe is expanding, following a Big Bang event 14 billion years ago. Under this paradigm, the universe begins as a hot, dense, fluid comprised of baryons, photons, and a menagerie of other, potentially even exotic, particles. Initially the universe was so hot and dense, that stable nuclei could not form; if an atom were to a form a high energy photon would come along and break it apart. In this sense, the photons were coupled to the baryons[1], via Compton scattering, so that the two species evolved in an codependent manner. This coupled medium is commonly referred to as a relativistic photon-baryon fluid. The universe cools as it expands, however, and eventually the average temperature of photons drops low enough that neutral atoms are able to form without being broken apart by photons. This event, known as *recombination*, took place roughly when the universe was around 380,000 years old. During recombination, as the free electrons and protons coalesce into neutral hydrogen, the photons are no longer coupled to the baryons, and stream off into the direction they last scattered into. These photons are exactly what is observed in the CMB today. It is in this sense that the CMB provides one of the earliest snapshots of the universe. Though the CMB is largely uniform in temperature, subtle fluctuations exist on the order of 10^{-5}. The distribution of these fluctuations encodes a wealth of cosmological information.

In this thesis, I will focus on the ΛCDM model, sometimes referred to as the standard model of cosmology. I will list some relevant assumptions of the ΛCDM model[2] here at the outset, and reference and elucidate them at various points throughout the thesis. Under ΛCDM, it is taken that

1. Physics is the same throughout the observable Universe and General Relativity is an adequate description of gravity.

2. The Universe is homogeneous and isotropic on large scales.

3. The Universe was once hotter and denser, and has been expanding since early times.

[1] A slight misnomer, as electrons are often implicitly included in the term "baryon" in cosmology.

[2] Adapted from [2].

4. Dark energy behaves just like the cosmological constant Λ (which first appears in eq. (1.11)).

5. Dark matter is pressureless, and only interacts with normal matter through gravity.

6. The curvature of space is very small.

7. Neutrinos are almost massless, and stream like relativistic particles at the time of recombination.

8. Variations in density were laid down everywhere at early times, and are Gaussian, adiabatic, and nearly scale invariant.

Taking these into account, I first begin by outlining the machinery necessary from General Relativity, and its implications.

1.1.2 General Relativity[3]

In general relativity, the primary object for understanding the structure of a space-time is the *metric*. In an arbitrary vector space, the metric g_{ij} is a tensor which allows one to define the notion of a distance on the space. The distance, or *interval*, at a given point is given by

$$ds^2 = \sum_{i,j} g_{ij} dx^i dx^j, \tag{1.1}$$

where i and j enumerate the chosen coordinates of the space, and g_{ij} is the metric tensor, defined at every point in the vector space. For example, on the 2D plane using Cartesian coordinates, where the possible coordinates $i, j = x, y$, the metric g_{ij} takes a form similar to that of a diagonal matrix with 1's on the diagonal, so that the interval[4] is

$$ds^2 = \sum_{i,j} g_{ij} dx^i dx^j = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} dx \\ dy \end{pmatrix} \begin{pmatrix} dx \\ dy \end{pmatrix} = \begin{pmatrix} dx & dy \end{pmatrix} \begin{pmatrix} dx \\ dy \end{pmatrix}$$

$$ds^2 = dx^2 + dy^2. \tag{1.2}$$

[3]This section follows Chapter 2 in [9].

[4]While the metric is not exactly a matrix and dx^i and dx^j are not exactly vectors, index notation allows one to treat the interval expression as a matrix equation.

The use of indices to represent vectors and tensors is known as Einstein notation. Following the conventions of the notation, from here on out summations, such as the one appearing in eq. (1.1), will not be written out explicitly but implied wherever a set of upper and lower indices appear in the same expression. Following this notation, eq. (1.1) would be written as

$$ds^2 = g_{ij}dx^i dx^j. \tag{1.3}$$

The metric captures information relating to the curvature of a space. Minkowski space-time is given by

$$g_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{1.4}$$

where $dx^0 = dt$ is the time coordinate and $dx^1, dx^2, dx^3 = dx, dy, dz$ are the spatial coordinates. Note that in this thesis, as is common in the literature, I work in a system of units where the speed of light $c = 1$. Thus the interval, known as the proper time in relativity, is given by inserting the metric from eq. (1.4) into eq. (1.3), resulting in

$$ds^2 = -dt^2 + dx^2 + dy^2 + dz^2. \tag{1.5}$$

If the universe is expanding, we expect the spatial component of the proper time to be scaled by some scale factor which changes, in this case expands, with time, but is otherwise uniform throughout space. The interval in an expanding space-time should look like

$$ds^2 = -dt^2 + a^2(t)\left(dx^2 + dy^2 + dz^2\right). \tag{1.6}$$

Note that this is the metric for a universe absent any curvature, which is the assumption I take for the rest of this work. This sort of expanding space-time describes the Friedmann-Robertson-Walker (FRW) space-time, whose metric is

$$\begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & a^2(t) & 0 & 0 \\ 0 & 0 & a^2(t) & 0 \\ 0 & 0 & 0 & a^2(t) \end{pmatrix}. \tag{1.7}$$

The scale factor a is often set so that its value today a_0 is 1. It's time evolution describes the expansion history of the universe.

To describe a particle in a given space-time, we consider its position and momentum four-vectors as

$$x_\mu = (t, \vec{x}),$$
$$P_\mu = \left(E, \vec{P}\right), \tag{1.8}$$

where t is time, E is the particle's energy, and $\vec{x}$ and $\vec{P}$ are the particle's spatial position and momentum, respectively. In an arbitrarily curved space-time, a particle will not follow a straight line in general, but a *geodesic* path, which can be understood as the shortest line between two paths on a curved surface. The geodesic equation governs a particle's motion in a curved space-time, and is given as

$$\frac{d^2x^\mu}{d\lambda^2} = -\Gamma^\mu_{\alpha\beta}\frac{dx^\alpha}{d\lambda}\frac{dx^\beta}{d\lambda}, \tag{1.9}$$

where the $\Gamma^\mu_{\alpha\beta}$ is known as the *Christoffel symbol*, constructed via derivatives of the metric as

$$\Gamma^\mu_{\alpha\beta} = \frac{g^{\mu\nu}}{2}\left[\frac{\partial g_{\alpha\nu}}{\partial x^\beta} + \frac{\partial g_{\beta\nu}}{\partial x^\alpha} - \frac{\partial g_{\alpha\beta}}{\partial x^\nu}\right]. \tag{1.10}$$

The Christoffel symbol contains the necessary information for understanding geodesics in curved space-time. In general relativity, the metric is related to the matter and energy content of the universe. The Einstein equations describe this complex relationship, given by

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}\mathcal{R} + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu}. \tag{1.11}$$

$T_{\mu\nu}$ is the stress-energy tensor, G is Newton's constant, and R and $\mathcal{R}$ are the Ricci tensor and Ricci scalar, respectively, given by

$$R_{\mu\nu} = \frac{\partial\Gamma^\alpha_{\mu\nu}}{\partial x^\alpha} - \frac{\partial\Gamma^\alpha_{\mu\alpha}}{\partial x^\nu} + \Gamma^\alpha_{\beta\alpha}\Gamma^\beta_{\mu\nu} - \Gamma^\alpha_{\beta\nu}\Gamma^\beta_{\mu\alpha}$$
$$\mathcal{R} = g^{\mu\nu}R_{\mu\nu}. \tag{1.12}$$

The Λ term is known as the *cosmological constant*, and in ΛCDM this represents the elusive dark energy component. Though not much is yet known of its precise physical origin, its mathematical description is quite simple, a constant Λ multiplying the metric appearing in Einstein's equations. In practice, eq. (1.11) is extremely complicated to deal with, as it contains 10 highly nonlinear coupled differential equations. However, we can arrive at simple and useful equations by approximating the universe as a perfect fluid. This entails

taking a stress-energy tensor of the form,

$$T^{\mu}_{\nu} = \begin{pmatrix} -\rho & 0 & 0 & 0 \\ 0 & \mathcal{P} & 0 & 0 \\ 0 & 0 & \mathcal{P} & 0 \\ 0 & 0 & 0 & \mathcal{P} \end{pmatrix}, \tag{1.13}$$

where ρ is the energy density of the fluid, and $\mathcal{P}$ is the pressure. Taking this in conjunction with the metric of eq. (1.7), we can arrive at the Friedmann equations, the first of which is

$$H^2 = \frac{8\pi G}{3}\rho, \tag{1.14}$$

where H is Hubble's constant, $H \equiv (da/dt)/a$. This is often rewritten as

$$\frac{H^2}{H_0^2} = \frac{\rho}{\rho_{cr}}, \tag{1.15}$$

where H_0 is the value of Hubble's constant today, and the critical density ρ_{cr} is

$$\rho_{cr} \equiv \frac{3H_0^2}{8\pi G}. \tag{1.16}$$

When considering the entire universe, the energy density ρ in eq. (1.15) accounts for all particle species. In order to understand how each of the separate components of the universe (matter, radiation, and dark energy) contributes to the total energy density, we can apply the continuity equation, modified from its traditional form to account for the curvature of space-time, to the stress-energy tensor of eq. (1.13),

$$\frac{\partial T^{\mu}_{\nu}}{\partial x^{\mu}} + \Gamma^{\mu}_{\alpha\mu} T^{\alpha}_{\nu} - \Gamma^{\alpha}_{\nu\mu} T^{\mu}_{\alpha} = 0. \tag{1.17}$$

This results in the equation

$$\frac{\partial \rho}{\partial t} + \frac{\dot{a}}{a}\left[3\rho + 3\mathcal{P}\right] = 0. \tag{1.18}$$

If we consider the energy density of pressureless matter (ordinary and dark) ρ_m, the ordinary and dark matter in our universe, we find that

$$\rho_m \propto a^{-3}. \tag{1.19}$$

Intuitively this makes sense, ρ_m is a density and scales inversely with volume, which is expanding in our universe as a^3. Radiation satisfies the relation $\mathcal{P}_r = \rho_r/3$, so eq. (1.18) gives

$$\rho_r \propto a^{-4} \tag{1.20}$$

as the scaling for the energy density of radiation. Again, the expansion of the universe supplies a factor of a^{-3}, and in the case of radiation, redshifting adds another factor of a^{-1}. Finally, in ΛCDM, dark energy satisfies the relation $\rho_\Lambda = -\mathcal{P}_\Lambda$, so eq. (1.18) gives

$$\frac{\partial \rho_\Lambda}{\partial t} = 0, \tag{1.21}$$

i.e. the energy density of dark energy remains constant. Recalling that we set the scale factor today $a_0 = 1$, we can use the scaling relationships above to write

$$\rho_m = \rho_{m,0}\ a^{-3}, \tag{1.22a}$$
$$\rho_r = \rho_{r,0}\ a^{-4}, \tag{1.22b}$$
$$\rho_\Lambda = \rho_{\Lambda,0}, \tag{1.22c}$$

where the 0 subscript denotes the energy density value today. Taking these expressions, we can write eq. (1.15) as

$$\frac{H^2}{H_0^2} = \frac{\rho_{m,0}}{\rho_{cr}} a^{-3} + \frac{\rho_{r,0}}{\rho_{cr}} a^{-4} + \frac{\rho_{\Lambda,0}}{\rho_{cr}}. \tag{1.23}$$

It is common to define the dimensionless density parameter as

$$\Omega_i = \frac{\rho_{i,0}}{\rho_{cr}}, \tag{1.24}$$

so that eq. (1.25) reads as

$$\frac{H^2}{H_0^2} = \Omega_m a^{-3} + \Omega_r a^{-4} + \Omega_\Lambda. \tag{1.25}$$

Two of the parameters in ΛCDM that constraints are sought for are Ω_c and Ω_b, the density parameters for dark matter and baryonic matter, respectively. From eq. (1.25), we can identify three distinct phases of the universe: for small values of the scale factor, early times, the universe was dominated by radiation, for intermediate times it was dominated by matter, and at late times dark energy dominates.

The above work outlines the necessary details required of general relativity towards an understanding of the CMB. In the following section, I turn towards the Boltzmann equation, which will be wielded to describe the evolution of a photon distribution in a perturbed FRW space-time, taking into account Compton scattering with electrons.

1.1.3 The Boltzmann Equation[5]

In its simplest form, the Boltzmann equation is written as

$$\frac{df}{dt} = C\left[f\right]. \tag{1.26}$$

The left-hand side is the time derivative of a distribution function f, and the right-hand side contains all the collision terms affecting the evolution of f. Before beginning work on this equation, we need to take stock of the background metric and the form of the photon distribution.

The photon distribution is taken as a perturbed Bose-Einstein distribution,

$$f\left(\vec{x}, p, \hat{p}, t\right) = \left[\exp\left\{\frac{p}{T(t)\left[1+\Theta\left(\vec{x},\hat{p},t\right)\right]}\right\}\right]^{-1}. \tag{1.27}$$

Perturbations in the photon distribution, as inhomogeneities and anisoptropies in the temperature, are captured in the Θ term. Taylor expanding about Θ gives

$$f \simeq f^{(0)} - p\frac{\partial f^{(0)}}{\partial p}\Theta, \tag{1.28}$$

where $f^{(0)}$ is the zero-order distribution, i.e. eq. (1.27) without any perturbation term.

The next ingredient to consider is the metric. We take our metric to have small perturbations[6] about the smooth FRW metric as described in eq. (1.7). The components of the perturbed FRW metric are taken to be

$$\begin{aligned} g_{00} &= -1 - 2\Psi\left(\vec{x},t\right), \\ g_{0i} &= 0, \\ g_{ij} &= a^2\delta_{ij}\left(1+2\Phi\left(\vec{x},t\right)\right), \end{aligned} \tag{1.29}$$

[5]This section follows Chapter 4 in [9].

[6]In this derivation, I only consider scalar perturbations to the metric, and I work in the Newtonian gauge.

where Ψ and Φ are known as gravitational potentials, the first quantity being akin to a Newtonian potential and the second quantity a perturbation to spatial curvature. These perturbations will be treated as small quantities in the following analysis. Expanding the total derivative of eq. (1.26) over the appropriates dependencies of the photon distribution function gives

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial x^i}\frac{dx^i}{dt} + \frac{\partial f}{\partial p}\frac{dp}{dt} + \frac{\partial f}{\partial \hat{p}^i}\frac{d\hat{p}^i}{dt}. \tag{1.30}$$

The photon distribution depends on t, the spatial coordinate x^i, and, notably, the magnitude and direction unit-vector of the spatial component of the four-momentum, here represented by p and $\hat{p}^i$ respectively. Firstly, the last term in eq. (1.30) disappears if we only consider first order terms. This is because, any nonzero contribution to $\partial f/\partial \hat{p}^i$ is first-order, because the unperturbed f is isotropic. Similarly, any nonzero contribution to $d\hat{p}^i/dt$ must also be first-order, because the photon propagation direction only changes due to the presence of the first order potentials Ψ and Φ.

We can express dx^i/dt as

$$\frac{dx^i}{dt} = \frac{dx^i}{d\lambda}\frac{d\lambda}{dt} = \frac{P^i}{P^0}, \tag{1.31}$$

using the fact that momentum is defined as $P^\mu = dx^\mu/d\lambda$, where λ is some affine parameter characterizing the path of the photon. In order to re-express this in terms of p and $\hat{p}^i$, and noting from general relativity that masslessness of a particle implies

$$g_{\mu\nu}P^\mu P^\nu = 0, , \tag{1.32}$$

we find

$$\begin{aligned} g_{\mu\nu}P^\mu P^\nu &= g_{00}(P^0)^2 + g_{ij}P^i P^j \\ 0 &= -\left(1+2\Psi\right)(P^0)^2 + p^2 \\ P^0 &= p\left(1-\Psi\right), \end{aligned} \tag{1.33}$$

where the last line follows by Taylor expanding in Ψ, and p is again the magnitude of the spatial component of the momentum. Using the fact that the momentum P^i is proportional to the unit-vector $\hat{p}^i$, we find

$$\begin{aligned} p^2 = g_{ij}P^i P^j &= g_{ij}(C\hat{p}^i)(C\hat{p}^j) \\ p^2 &= C^2 a^2\left(1+2\Phi\right)\delta_{ij}\hat{p}^i\hat{p}^j \end{aligned}$$

$$\Rightarrow C = \frac{p}{a}(1-\Phi)$$
$$P^i = C\hat{p}^i = p\hat{p}^i\frac{1-\Phi}{a}. \tag{1.34}$$

Thus, the term in eq. (1.31) can be written as

$$\frac{dx^i}{dt} = \frac{\hat{p}^i}{a}(1+\Psi-\Phi), \tag{1.35}$$

after performing the appropriate Taylor expansion in Ψ. However, because this term appears in eq. (1.30) multiplying $\partial f/\partial x^i$, which is first-order (as the unperturbed f is homogeneous) we can neglect the first-order potential terms when importing eq. (1.35) into eq. (1.30). With this exposition thus far, eq. (1.30) becomes

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{\hat{p}^i}{a}\frac{\partial f}{\partial x^i} + \frac{\partial f}{\partial \hat{p}^i}\frac{d\hat{p}^i}{dt}. \tag{1.36}$$

Next, we consider the $d\hat{p}^i/dt$ term. We start from eq. (1.9), using the fact that $P^\mu = dx^\mu/dt$ to transform it into an equation of four-momenta,

$$\frac{dP^\mu}{d\lambda} = -\Gamma^\mu_{\alpha\beta}P^\alpha P^\beta. \tag{1.37}$$

If we consider only the time component of the tensor, we have

$$\frac{dP^0}{d\lambda} = \frac{dt}{d\lambda}\frac{dP^0}{dt} = -\Gamma^0_{\alpha\beta}P^\alpha P^\beta$$
$$\Rightarrow \frac{d}{dt}\left[p(1-\Psi)\right] = -\Gamma^0_{\alpha\beta}\frac{P^\alpha P^\beta}{p}(1+\Psi), \tag{1.38}$$

where the second line follows by noticing that $dt/d\lambda = P^0$, and replacing P^0 terms using eq. (1.33), followed by the appropriate Taylor expansion on Ψ. By expanding the time derivative and rearranging, eq. (1.38) becomes

$$\frac{dt}{dt} = p\frac{d\Psi}{dt} - \Gamma^0_{\alpha\beta}\frac{P^\alpha P^\beta}{p}(1+\Psi)$$
$$= p\left\{\frac{\partial\Psi}{\partial t} + \frac{dx^i}{dt}\frac{\Psi}{x^i}\right\} - \Gamma^0_{\alpha\beta}\frac{P^\alpha P^\beta}{p}(1+\Psi)$$
$$\frac{dp}{dt} = p\left\{\frac{\partial\Psi}{\partial t} + \frac{\hat{p}^i}{a}\frac{\Psi}{x^i}\right\} - \Gamma^0_{\alpha\beta}\frac{P^\alpha P^\beta}{p}(1+\Psi). \tag{1.39}$$

The Christoffel symbol can be calculated using eq. (1.10) and the perturbed FRW metric as given in eq. (1.29), and imported into eq. (1.39) to first order to give

$$\begin{aligned} \frac{dp}{dt} &= -pH - p\frac{\partial\Phi}{\partial t} - p\frac{\hat{p}^i}{a}\frac{\partial\Psi}{\partial x^i} \\ \frac{1}{p}\frac{dp}{dt} &= -H - \frac{\partial\Phi}{\partial t} - \frac{\hat{p}^i}{a}\frac{\partial\Psi}{\partial x^i}. \end{aligned} \tag{1.40}$$

Collecting the results thus far, we can write eq. (1.30) as

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{\hat{p}^i}{a}\frac{\partial f}{\partial x^i} - p\frac{\partial f}{\partial p}\left[H + \frac{\partial\Phi}{\partial t} + \frac{\hat{p}^i}{a}\frac{\partial\Psi}{\partial x^i}\right] \tag{1.41}$$

Because we are interested in understanding inhomogeneities and anisotropies in the photon distribution, we import the first order term of eq. (1.28) into eq. (1.41) in place of f, leading to the following expression for the left-hand side of eq. (1.26),

$$\left.\frac{df}{dt}\right|_{\text{first order}} = -p\frac{\partial f^{(0)}}{\partial p}\left[\frac{\partial\Theta}{\partial t} + \frac{\hat{p}^i}{a}\frac{\partial\Theta}{\partial x^i} + \frac{\partial\Phi}{\partial t} + \frac{\hat{p}^i}{a}\frac{\partial\Psi}{\partial x^i}\right]. \tag{1.42}$$

In effect, this describes the evolution of Θ in the absence of collision terms.

Next, we need to consider Compton scattering in the collision term on the right-hand side of eq. (1.26). Schematically, Compton scattering is represented as

$$e^-(\vec{q}) + \gamma(\vec{p}) \longleftrightarrow e^-(\vec{q}') + \gamma(\vec{p}'), \tag{1.43}$$

where an incoming electron with momentum $\vec{q}$ and incoming photon with momentum $\vec{p}$ are scattered into an outgoing electron with momentum $\vec{q}'$ and photon with momentum $\vec{p}'$. The collision term for the scattering event of a photon with incoming momentum $\vec{p}$ takes the form

$$C[f] = \sum_{\vec{q},\vec{q}',\vec{p}'} |\mathcal{M}|^2 \left\{f_e(\vec{q}')f(\vec{p}') - f_e(\vec{q})f(\vec{p})\right\}. \tag{1.44}$$

The sum is over all possible incoming and outgoing momenta of the collision particles, and $\mathcal{M}$ is the amplitude for Compton scattering, given by

$$|\mathcal{M}|^2 = 8\pi\sigma_T m_e^2, \tag{1.45}$$

where σ_T is the Thomson cross-section and m_e is the electron mass. In order to account for all possible collisions of a photon with incoming momentum $\vec{p}$, we need to expand the sum in eq. (1.44) to integrals over all possible incoming and outgoing momenta for the photons and electrons in the collision, and enforce energy and momentum conservation using delta functions. Eq. (1.44) thus expanded becomes

$$\begin{aligned} C[f] =& \frac{1}{p}\int \frac{d^3q}{(2\pi)^3 2E_e(q)} \int \frac{d^3q'}{(2\pi)^3 2E_e(q')} \int \frac{d^3p'}{(2\pi)^3 2E(p')} |\mathcal{M}|^2 (2\pi)^4 \\ & \times \delta^3[\vec{p}+\vec{q}-\vec{p}'-\vec{q}']\delta[E(p)+E_e(q)-E(p')-E_e(q')] \\ & \times \{f_e(\vec{q}')f(\vec{p}') - f_e(\vec{q})f(\vec{p})\}\,. \end{aligned} \tag{1.46}$$

The various extra factors accompany the integrals over phase space. The photon energies are simply $E(p) = p$, and in the epochs of interest the electrons have negligible kinetic energy, so $E_e(q) = m_e$. Thus,

$$\begin{aligned} C[f] =& \frac{\pi}{4m_e^2 p}\int \frac{d^3q}{(2\pi)^3} \int \frac{d^3p'}{(2\pi)^3 p'} \int d^3q' |\mathcal{M}|^2 \\ & \times \delta^3[\vec{p}+\vec{q}-\vec{p}'-\vec{q}']\delta\left[p+\frac{q^2}{2m_e}-p'-\frac{q'^2}{2m_e}\right] \\ & \times \{f_e(\vec{q}')f(\vec{p}') - f_e(\vec{q})f(\vec{p})\}\,. \end{aligned}$$

$$\begin{aligned} C[f] =& \frac{\pi}{4m_e^2 p}\int \frac{d^3q}{(2\pi)^3} \int \frac{d^3p'}{(2\pi)^3 p'} |\mathcal{M}|^2 \\ & \times \delta\left[p+\frac{q^2}{2m_e}-p'-\frac{(\vec{q}+\vec{p}-\vec{p}')^2}{2m_e}\right] \\ & \times \{f_e(\vec{q}+\vec{p}-\vec{p}')f(\vec{p}') - f_e(\vec{q})f(\vec{p})\}\,, \end{aligned} \tag{1.47}$$

where in the last expression the delta function had been used to evaluate the q' integral. Using the fact that q is much larger than p and p', and that $p \simeq p'$,

eq. (1.47) can be rewritten as

$$C[f] = \frac{\pi}{4m_e^2 p}\int \frac{d^3q f_e(q)}{(2\pi)^3}\int \frac{d^3p'}{(2\pi)^3 p'}|\mathcal{M}|^2\delta\left[p - p' - \frac{(\vec{p}-\vec{p}')\cdot\vec{q}}{m_e}\right]\{f(\vec{p}') - f(\vec{p})\}\,. \tag{1.48}$$

Because $p \simeq p'$, we can expand the delta function about $\vec{p}-\vec{p}'$, and insert eq. (1.45) for the amplitude, to get

$$C[f] = \frac{2\pi^2\sigma_T}{p}\int \frac{d^3q f_e(q)}{(2\pi)^3}\int \frac{d^3p'}{(2\pi)^3 p'}$$

$$\times\left\{\delta\left(p-p'\right) + \frac{(\vec{p}-\vec{p}')\cdot\vec{q}}{m_e}\frac{\partial\delta\left(p-p'\right)}{\partial p'}\right\}\{f(\vec{p}') - f(\vec{p})\}\,. \tag{1.49}$$

Now we are in a position to insert eq. (1.28) for f and split the triple integral into separate integrals over magnitude and solid angle, giving

$$C[f] = \frac{n_e\sigma_T}{4\pi p}\int dp' p'\int d\Omega'\left[\delta\left(p-p'\right)\left(-p'\frac{\partial f^{(0)}}{\partial p'}\Theta(\hat{p}') + p\frac{\partial f^{(0)}}{\partial p}\Theta(\hat{p})\right)\right.$$

$$\left. + (\vec{p}-\vec{p}')\cdot\vec{v_b}\frac{\partial\delta\left(p-p'\right)}{\partial p'}\left(f^{(0)}(p') - f^{(0)}(p)\right)\right], \tag{1.50}$$

where respective second order terms have been neglected, and the integral over q is evaluated as the electron number density n_e for the term with no q dependence, and $n_e v_b$ for the $\vec{q}/m_e$ term. The integral evaluates to

$$C[f] = -p\frac{\partial f^{(0)}}{\partial p}n_e\sigma_T\left[\Theta_0 - \Theta(\hat{p}) + \hat{p}\cdot\vec{v}_b\right], \tag{1.51}$$

where Θ_0 is the perturbation monopole, defined as

$$\Theta_0(\vec{x},t) = \frac{1}{4\pi}\int d\Omega'\Theta\left(\hat{p}',\vec{x},t\right). \tag{1.52}$$

With this expression for the collision term, we can equate eqs. (1.51) and (1.42), giving

$$a\frac{\partial\Theta}{\partial t} + \hat{p}^i\frac{\partial\Theta}{\partial x^i} + a\frac{\partial\Phi}{\partial t} + \hat{p}^i\frac{\partial\Psi}{\partial x^i} = n_e\sigma_T a\left[\Theta_0 - \Theta(\hat{p}) + \hat{p}\cdot\vec{v}_b\right]. \tag{1.53}$$

It is more convenient to work with this equation in Fourier space, so that the coupled system above may reduce to ordinary differential equations. It is also convenient to work with conformal time[7], defined as

$$\eta = \int \frac{dt}{a(t)}. \tag{1.54}$$

Thus, eq. (1.53) becomes

$$\dot{\tilde{\Theta}} + i(\vec{k}\cdot\hat{p})\tilde{\Theta} + \dot{\tilde{\Phi}} + i(\vec{k}\cdot\hat{p})\tilde{\Psi} = n_e\sigma_T a\left[\tilde{\Theta}_0 - \tilde{\Theta} + \frac{(\hat{p}\cdot\vec{k})}{k}\tilde{v_b}\right]. \tag{1.55}$$

We define the cosine of the angle between the wavenumber $\vec{k}$ and the photon direction $\hat{p}$ as

$$\mu \equiv \frac{\vec{k}\cdot\hat{p}}{k}, \tag{1.56}$$

and the optical depth as

$$\tau(\eta) = \int_{\eta}^{\eta_0} d\eta' n_e\sigma_T a \tag{1.57}$$

(η_0 being the conformal time today), so that we may write a final equation for the evolution of the photon distribution, as

$$\dot{\tilde{\Theta}} + ik\mu\tilde{\Theta} + \dot{\tilde{\Phi}} + ik\mu\tilde{\Psi} = -\dot{\tau}\left[\tilde{\Theta}_0 - \tilde{\Theta} + \mu\tilde{v_b}\right]. \tag{1.58}$$

Thus we have a linear equation for the perturbation to the photon distribution. The next step is to translate the somewhat abstract photon perturbation Θ to something that we can observe on sky.

1.1.4 The CMB Power Spectrum[8]

In order to describe the photon perturbation as it would appear on the sky, we expand the field in terms of spherical harmonics. The expansion is

$$\Theta(\vec{x},\hat{p},\eta) = \sum_{l=1}^{\infty}\sum_{m=-l}^{l} a_{lm}(\vec{x},\eta)\, Y_{lm}(\hat{p}). \tag{1.59}$$

[7]Another way to think of conformal time is as the maximum comoving distance traveled by light since the beginning of the universe, or the distance traveled by light in the expanding frame of the universe.

[8]This section follows Chapter 8 in [9].

The a_{lm} amplitudes contain the temperature field information. If we invert eq. (1.59) by exploiting the orthogonality of the spherical harmonics, we get the expression

$$a_{lm}\left(\vec{x},\eta\right)=\int\frac{d^3k}{(2\pi)^3}e^{i\vec{k}\cdot\vec{x}}\int d\Omega Y^*_{lm}(\hat{p})\Theta(\vec{k},\hat{p},\eta), \tag{1.60}$$

where the photon distribution is Fourier transformed, to be as it appears in eq. (1.58). It is useful to perform a multipole expansion on Θ, so that

$$\Theta(\vec{k},\hat{p},\eta)=\sum_l(2l+1)(-i)^l\Theta_l(\vec{k},\eta)\mathcal{P}_l(\hat{p}\cdot\hat{k}), \tag{1.61}$$

where $\mathcal{P}_l$ is the lth Legendre polynomial. We have already seen the Θ_0, the monopole of the photon distribution, in eq. (1.52). The variance of the a_{lm}s is called C_l,

$$\langle a_{lm}a^*_{l'm'}\rangle=\delta_{ll'}\delta_{mm'}C_l, \tag{1.62}$$

and this is the quantity we can get a handle on via observations. We cannot make any predictions about the particular a_{lm}s, just the distribution they are drawn from. Note that from here on, I suppress the η in my notation to reduce clutter. We can use eqs. (1.60) and (1.61) to compute the variance.

$$a_{lm}a^*_{l'm'}=\sum_{l_1,l_2}(2l_1+1)(2l_2+1)(-i)^{l_1-l_2}\int\frac{d^3kd^3k'}{(2\pi)^6}e^{i(\vec{k}-\vec{k}')\cdot\vec{x}}$$

$$\times\int d\Omega_{\hat{p}}d\Omega_{\hat{p}'}Y^*_{lm}(\hat{p})Y_{l'm'}(\hat{p}')\mathcal{P}_{l_1}(\hat{p}\cdot\hat{k})\Theta_{l_1}(\vec{k})\mathcal{P}_{l_2}(\hat{p}'\cdot\hat{k}')\Theta^*_{l_1}(\vec{k}') \tag{1.63}$$

$$a_{lm}a^*_{l'm'}=\sum_{l_1,l_2}(2l_1+1)(2l_2+1)(-i)^{l_1-l_2}\int\frac{d^3kd^3k'}{(2\pi)^6}e^{i(\vec{k}-\vec{k}')\cdot\vec{x}}\Theta_{l_1}(\vec{k})\Theta^*_{l_1}(\vec{k}')$$

$$\times\int d\Omega_{\hat{p}}Y^*_{lm}(\hat{p})\mathcal{P}_{l_1}(\hat{p}\cdot\hat{k})\int d\Omega_{\hat{p}'}Y_{l'm'}(\hat{p}')\mathcal{P}_{l_2}(\hat{p}'\cdot\hat{k}'). \tag{1.64}$$

To get the C_l,

$$C_l=\frac{2}{\pi}\int\frac{d^3kd^3k'}{(2\pi)^3}e^{i(\vec{k}-\vec{k}')\cdot\vec{x}}\langle\Theta_{l_1}(\vec{k})\Theta^*_{l_1}(\vec{k}')\rangle Y_{lm}(\hat{k})Y^*_{lm}(\hat{k}'). \tag{1.65}$$

where the delta function in (1.62) enforces $l'=l$, and the summation collapses

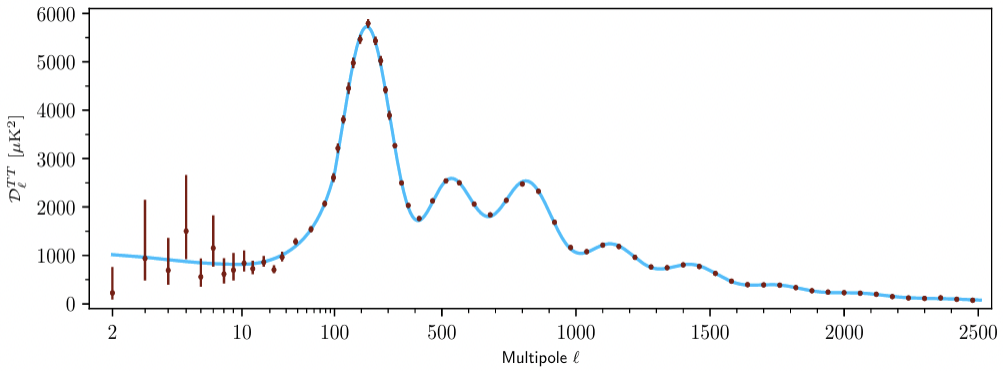


Figure 1.2: The CMB power spectrum as measured by the Planck experiment as the red points, overlayed on top of the ΛCDM predicted spectrum in blue, for the Planck best fit parameters. Figure from [2].

in calculating the angular integrals, which are only nonzero for the terms where $l_1 = l_2 = l$. The photon perturbation can be decomposed into the product between two components, an initial perturbation component not dependent on direction, and the subsequent evolution in the perturbation (via the Boltzmann equation) which generates a directional dependence. The final expression for C_l is thus

$$C_l = \frac{2}{\pi} \int_0^\infty P(k) |\Theta_l(k)|^2 k^2 dk, \tag{1.66}$$

where $P(k)$ is the matter power spectrum, a quantity which I discuss in detail later in this work. Thus we obtain the information required to plot the CMB power spectrum, the C_l's, one of the most powerful cosmological probes. In effect, it describes the size of the perturbations in the photon distribution as a function of angular scale, l. Figure 1.2 shows the CMB power spectrum as measured by the Planck experiment as the red points, overlaid on top of the ΛCDM predicted curve in blue for the Planck best fit parameters. The plot shows $\mathcal{D}_l$, where

$$\mathcal{D}_l = \frac{l(l+1)C_l}{2\pi}. \tag{1.67}$$

1.2 Part 2: Towards an Analytic Expression

In this part, I follow [9] and [5] in deriving a semi-analytic formulation for the photon perturbation moments, the Θ_l, which makes possible a semi-analytic computation of the CMB power spectrum. First, the tight coupling approximation is applied to solve for the monopole and dipole moments of

the perturbation at recombination. These are subsequently damped due to photon diffusion damping at recombination, giving an effective perturbation at last scattering. Finally, the effective perturbation at last scattering is *free-streamed* to ($\eta = \eta_0$), resulting in the perturbation we observe today.

1.2.1 Tight Coupling[9]

Under the tight coupling approximation, we assume that the electron-proton fluid is tightly coupled to the photons via Compton scattering. This implies that the scattering rate is much larger than the expansion rate, or consequently the optical depth in eq. (1.57) $\tau \gg 1$. In this limit, only the first two moments of the perturbation are nonnegligible. The evolution equations for Θ_l can be obtained by multiplying eq. (1.58) by $\mathcal{P}_l(\mu)$ and integrating over μ, exploiting the orthogonality of the Legendre polynomials. This results in the following two equations for the first two moments of the perturbation:

$$\dot{\Theta}_0 + k\Theta_1 = -\dot{\Phi} \tag{1.68a}$$

$$\dot{\Theta}_1 - \frac{k\Theta_0}{3} = \frac{k\Psi}{3} + \dot{\tau}\left[\Theta_1 - \frac{iv_b}{3}\right]. \tag{1.68b}$$

Note that in this section, all relevant quantities are in Fourier space, unless otherwise specified. η dependence is also implied. An analysis similar to the Boltzmann analysis above but for baryons gives an equation for the electron-baryon fluid velocity:

$$v_b = -3i\Theta_1 + \frac{R}{\dot{\tau}}\left[\dot{v}_b + \frac{\dot{a}}{a}v_b + ik\Psi\right] \tag{1.69}$$

$$v_b \simeq -3i\Theta_1 + \frac{R}{\dot{\tau}}\left[-3i\dot{\Theta}_1 + \frac{\dot{a}}{a}\left(-3i\Theta_1\right) + ik\Psi\right], \tag{1.70}$$

where the second line follows by replacing v_b with its approximation to first order in τ, $-3i\Theta_1$. R is the ratio of baryon to photon density, $3\rho_b/4\rho_\gamma$. Inserting this expression into eq. (1.68b) gives

$$\dot{\Theta}_1 + \frac{\dot{a}}{a}\frac{R}{1+R}\Theta_1 - \frac{k}{3(1+R)}\Theta_0 = \frac{k\Psi}{3}. \tag{1.71}$$

Along with eq. (1.68a), this is a system of first-order coupled equations, and can be combined into one second order equation by taking the derivative of

[9]This section follows Chapter 8 in [9].

eq. (1.68a),

$$\ddot{\Theta}_0 + k\dot{\Theta}_1 = -\ddot{\Phi}, \tag{1.72}$$

and inserting this into eq. (1.71) leaves

$$\ddot{\Theta}_0 + \frac{\dot{a}}{a}\frac{R}{1+R}\dot{\Theta}_0 + k^2 c_s^2 \Theta_0 = F(\eta), \tag{1.73}$$

where $F(\eta)$ is

$$F(\eta) = -\ddot{\Phi} - \frac{\dot{a}}{a}\frac{R}{1+R}\dot{\Phi} - \frac{k^2\Psi}{3}. \tag{1.74}$$

This is of the form of a damped harmonic oscillator, driven by a forcing function $F(\eta)$. c_s is the sound speed of the fluid, defined as

$$c_s = \sqrt{\frac{1}{3(1+R)}}. \tag{1.75}$$

The homogenous solutions of eq. (1.73), via the WKB approximation, are

$$\theta_a = (1+R)^{-1/4}\cos(kr_s) \tag{1.76a}$$

$$\theta_b = (1+R)^{-1/4}\sin(kr_s), \tag{1.76b}$$

where

$$r_s = \int_0^\eta c_s d\eta' \tag{1.77}$$

is the sound horizon, the comoving distance traveled by a sound wave by time η. These homogenous solutions can then be used to construct a particular solution via Green's method, giving

$$\hat{\Theta}_0(\eta) = C_1\theta_a(\eta) + C_2\theta_b(\eta) + \int_0^\eta \frac{\theta_a(\eta')\theta_b(\eta) - \theta_a(\eta)\theta_b(\eta')}{\theta_a(\eta')\dot{\theta}_b(\eta') - \dot{\theta}_a(\eta')\theta_b(\eta')}F(\eta')d\eta', \tag{1.78}$$

giving, with the help of simplifying trigonometric identities,

$$\hat{\Theta}_0(\eta) = C_1\theta_a(\eta) + C_2\theta_b(\eta) + \frac{\sqrt{3}}{k}\int_0^\eta [1+R(\eta')]^{3/4}\sin[kr_s(\eta) - kr_s(\eta')]F(\eta')d\eta'. \tag{1.79}$$

The constants are fixed by initial conditions, giving

$$[1+R(\eta)]^{1/4}\,\hat{\Theta}_0(\eta) = \Theta_0(0)\cos(kr_s(\eta)) + \frac{\sqrt{3}}{k}\left[\dot{\Theta}_0(0) + \frac{1}{4}\dot{R}(0)\Theta_0(0)\right]\sin(kr_s(\eta))$$

$$+ \frac{\sqrt{3}}{k} \int_0^\eta [1 + R(\eta')]^{3/4} \sin[kr_s(\eta) - kr_s(\eta')] F(\eta') d\eta'. \tag{1.80}$$

In order to mold this into a more workable form, we can integrate by parts twice on eq. (1.80), in order to get rid of the double derivative terms in eq. (1.74) appearing in the integral, giving

$$(1+R)^{1/4} [\hat{\Theta}_0(\eta) + \Phi(\eta)] = [\cos kr_s(\eta) + J(0) \sin kr_s(\eta)][\Theta_0(0) + \Phi(0)] + I(\eta), \tag{1.81}$$

where the expressions

$$J(\eta) \equiv \frac{\sqrt{3}}{4k} \frac{\dot{R}}{\sqrt{1+R}}, \tag{1.82a}$$

$$I(\eta) \equiv \frac{k}{\sqrt{3}} \int_0^\eta d\eta' \Phi(\eta') G(\eta') \sin[kr_s(\eta) - kr_s(\eta')], \tag{1.82b}$$

$$G(\eta) \equiv (1+R)^{-1/4} \left[1 - (1+R)\frac{\Psi}{\Phi} + \frac{3}{4k^2} \ddot{R} - J^2 \right]. \tag{1.82c}$$

Using eq. (1.68a), we obtain the following expression for $\hat{\Theta}_0$,

$$\begin{aligned}(1+R)^{3/4} \frac{\hat{\Theta}_1(\eta)}{\sqrt{3}} &= [1 + J(\eta)J(0)][\Theta_0(0) + \Phi(0)] \sin(kr_s(\eta)) \\ &+ [J(\eta) - J(0)][\Theta_0(0) + \Phi(0)] \cos(kr_s(\eta)) + J(\eta) I(\eta) \\ &- \frac{k}{\sqrt{3}} \int_0^\eta d\eta' \Phi(\eta') G(\eta') cos[kr_s(\eta) - kr_s(\eta')].\end{aligned} \tag{1.83}$$

Large scale modes, not well approximated by the WKB approximation, are instead found by setting R = 0 in the above equations[5], giving

$$\begin{aligned}\hat{\Theta}_0(\eta) + \Phi(\eta) &= [\Theta_0(0) + \Phi(0)] \cos kr_s(\eta) \\ &+ \frac{k}{\sqrt{3}} \int_0^\eta d\eta' [\Phi(\eta') - \Psi(\eta')] \sin[kr_s(\eta) - kr_s(\eta')]\end{aligned} \tag{1.84a}$$

$$(1+R)^{1/2} \frac{\hat{\Theta}_1(\eta)}{3\sqrt{3}} = [\Theta_0(0) + \Phi(0)] \sin kr_s(\eta)$$

$$- \frac{k}{\sqrt{3}} \int_0^\eta d\eta' [\Phi(\eta') - \Psi(\eta')] \cos[kr_s(\eta) - kr_s(\eta')]. \quad (1.84b)$$

Thus, the above eqs. (1.81)-(1.84b), matched up appropriately [10], describe the monopole and dipole of the perturbation at last scattering[11], having not yet accounted for damping.

1.2.2 Damping Effects

To arrive at the effective monopole and dipole at last scattering, we need to account for photon diffusion damping. Damping occurs on very small scales, as a result of the photons scattering and random walking between each scatter. Thus, modes on the order of the mean free path between scatters, $\lambda_{\rm MFP}$, are expected to be washed out. Furthermore, as recombination occurs over a period of time, $\lambda_{\rm MFP}$ increases and approaches infinity, so some measure of damping will be observed across a range of modes.

Diffusion is represented as a non-negligible quadrupole moment in the perturbation, so eqs. (1.68a)-(1.68b) need to be augmented with an expression for Θ_2. Following a similar method, but not neglecting Θ_2 this time, we get

$$\dot{\Theta}_0 + k\Theta_1 = 0 \quad (1.85a)$$

$$\dot{\Theta}_1 + k\left(\frac{2}{3}\Theta_2 - \frac{1}{3}\Theta_0\right) = \dot{\tau}\left(\Theta_1 - \frac{iv_b}{3}\right) \quad (1.85b)$$

$$\dot{\Theta}_2 - \frac{2k}{5}\Theta_1 = \frac{9}{10}\dot{\tau}\Theta_2, \quad (1.85c)$$

where the potentials Φ and Ψ have been dropped, as their effect is negligible on small scales. This system of equations is supplemented with eq. (1.69), again dropping the potential, and rearranging,

$$3i\Theta_1 + v_b = \frac{R}{\dot{\tau}}\left[\dot{v}_b + \frac{\dot{a}}{a}v_b\right]. \quad (1.86)$$

In order to account for damping in the oscillatory solution to this system, we write the solution for the variables in the form of

$$\Theta_0, \Theta_1, v_b \sim e^{i\int \omega d\eta}, \quad (1.87)$$

[10] [5] suggests joining them at $k = k_S = 0.08h^3$.

[11] Supplied with the appropriate gravitational potentials and initial conditions, as predicted by whatever theory this calculation is taking place within.

where damping appears as an imaginary component of ω. Damping occurs on small scales, corresponding to large k, and in this limit we can drop terms $\propto \eta^{-1}$. Thus, eqs. (1.85a), (1.85c), and (1.86), give

$$i\omega\Theta_0 = -k\Theta_1, \tag{1.88}$$

$$\Theta_2 = -\frac{4k}{9\dot{\tau}}\Theta_1, \tag{1.89}$$

$$v_b \simeq -3i\Theta_1\left[1 + \frac{i\omega R}{\dot{\tau}} - \left(\frac{\omega R}{\dot{\tau}}\right)^2\right], \tag{1.90}$$

respectively. Inserting these into eq. (1.85b) gives the dispersion relation

$$i\omega - \frac{8k^2}{27\dot{\tau}} + \frac{k^2}{3i\omega} = \dot{\tau}\left(1 - \left[1 + \frac{i\omega R}{\dot{\tau}} - \left(\frac{\omega R}{\dot{\tau}}\right)^2\right]\right), \tag{1.91}$$

from which we find

$$\omega = kc_s - \frac{ik^2}{2(1+R)\dot{\tau}}\left[c_s^2 R^2 + \frac{8}{27}\right]. \tag{1.92}$$

Inserting this into eq. (1.87), the solutions must be of the form

$$\Theta_0, \Theta_1 \sim \exp\left\{ik\int d\eta c_s\right\}\exp\left\{-\frac{k^2}{k_D^2}\right\}, \tag{1.93}$$

where the damping length k_D is defined as

$$\frac{1}{k_D^2(\eta)} \equiv \int_0^\eta \frac{d\eta'}{6(1+R)n_e\sigma_T a(\eta')}\left[\frac{R^2}{(1+R)} + \frac{8}{9}\right]. \tag{1.94}$$

Thus, the expected perturbation monopole at recombination, $\eta = \eta_*$, in terms of the tight coupling solution $\hat{\Theta}_0$ becomes

$$(\Theta_0 + \Psi)(\eta_*) = (\hat{\Theta}_0 + \Psi)(\eta_*)e^{-[k/k_D(\eta_*)]^2}. \tag{1.95}$$

This however, supposes that recombination occurs instantaneously at $\eta = \eta_*$, which is not true. In order to take into account the finite duration of recombination, we instead define the average damping factor

$$\mathcal{D}(k) = \int_0^{\eta_0} \dot{\tau}e^{-\tau}e^{-[k/k_D(\eta)]^2}d\eta, \tag{1.96}$$

so that the expected monopole at recombination is

$$(\Theta_0 + \Psi)(\eta_*) = (\hat{\Theta}_0 + \Psi)(\eta_*)\mathcal{D}(k). \tag{1.97}$$

Similarly, the dipole is given by

$$\Theta_1(\eta_*) = \hat{\Theta}_1(\eta_*)\mathcal{D}(k). \tag{1.98}$$

With the perturbations thus evolved through the end of recombination, what remains to be explored is how these perturbations evolve past recombination, through to today.

1.2.3 Free Streaming

After recombination, the universe had cooled enough to allow the previously free electrons to combine with the free protons and form neutral hydrogen atoms, without being broken apart by high energy photons. What this amounts to is the decoupling of the photon-baryon fluid. The perturbations no longer evolve as they did in the fluid, and stream through to the present almost unchanged. In order to physically bridge the perturbations from recombination to the present, we return to eq. (1.58), this time without assuming the tight-coupling approximation. Eq. (1.58) can be rewritten as

$$\dot{\Theta} + \dot{\Psi} + (ik\mu - \dot{\tau})(\Theta + \Psi) = \dot{\Psi} - \dot{\Phi} - \dot{\tau}\left[\dot{\Theta}_0 + \Psi + \mu v_b\right]. \tag{1.99}$$

This can be expressed in the following form

$$e^{-(ik\mu\eta - \tau)}\frac{d}{d\eta}\left[e^{ik\mu\eta-\tau}(\Theta + \Psi)\right] = \dot{\Psi} - \dot{\Phi} - \dot{\tau}\left[\dot{\Theta}_0 + \Psi + \mu v_b\right] \tag{1.100}$$

where we can solve for $\Theta + \Psi$ today as

$$(\Theta + \Psi)(\eta_0) = \int_0^{\eta_0}\left[\dot{\Psi} - \dot{\Phi} - \dot{\tau}\left(\Theta_0 + \Psi + \mu v_b\right)\right]e^{-\tau(\eta)}e^{ik\mu(\eta-\eta_0)}d\eta. \tag{1.101}$$

In order to recover an expression for the moments, the Θ_l in eq. (1.61), we multiply by $\mathcal{P}_l(\mu)$ and integrate over μ. It is helpful to make use of the identity

$$\int_{-1}^{1}\frac{d\mu}{2}\mathcal{P}_l(\mu)e^{ik\mu(\eta-\eta_0)} = \frac{1}{(-i)^l}j_l\left[k(\eta - \eta_0)\right], \tag{1.102}$$

where j_l are the spherical Bessel functions of the first kind. Furthermore, we can take advantage of the fact that the function $-\dot{\tau}e^{-\tau}$ is sharply peaked at

recombination, η_*, and appears multiplied against slowly varying functions, so it can be treated as a delta function to collapse the respective integrals over η. With the above in mind with respect to eq. (1.101), the approximation for the perturbation given by [5] can be arrived at,

$$\begin{aligned}\Theta_l(\eta_0) \approx\ & (\Theta_0 + \Psi)(\eta_*)(2l+1)j_l[k(\eta_0 - \eta_*)] \\ & + \Theta_1(\eta_*)\left\{l j_{l-1}[k(\eta_0 - \eta_*)] - (l+1)j_{l+1}[k(\eta_0 - \eta_*)]\right\} \\ & + (2l+1)\int_{\eta_*}^{\eta_0}(\dot{\Psi} - \dot{\Phi})j_l[k(\eta_0 - \eta)]d\eta. \end{aligned} \tag{1.103}$$

With this last piece, the steps for reconstructing the CMB Power Spectrum semi-analytically have been laid out. The necessary ingredients are the initial conditions of the perturbation and the potentials. With these, one can construct eqs. (1.81)-(1.84b), damp them appropriately via eqs. (1.97)-(1.98), free stream these quantities via eq. (1.103), and integrate these to calculate the CMB power spectrum C_l's via eq. (1.66).

In this chapter I've described outlined of the assumptions of the ΛCDM model, and described a key cosmological observable, the CMB power spectrum. I then describe a popular semi-analytic formulation for the CMB power spectrum, in order both to serve as a CMB primer but also to convey the general workflow associated with constructing an accurate, rigorous, and physics-informed analytic approximation.

Chapter 2

Machine Learning

Generally speaking, the term Machine Learning encompasses a broad range of algorithmic procedures, or models, designed to detect patterns in datasets [10]. Supervised machine learning, which is the focus of the study herein, is a subclass of machine learning in which the model attempts to learn the relationship between inputs and their corresponding outputs in a given dataset, ultimately being able to generalize to unseen data.

Supervised machine learning models require a training dataset to learn from, which is composed of model inputs and their corresponding outputs, or targets. For our purposes we can imagine the training dataset to be the set

$$D_{\text{train}} = \{(\vec{x}_1, \vec{y}_1),\ \ldots,\ (\vec{x}_n, \vec{y}_n)\}\,, \tag{2.1}$$

though this need not always be the case. The dataset D contains n data points, pairs of input vectors $\vec{x}_i$ and the corresponding target vectors $\vec{y}_i$. One assumes that there exists a relationship between pairs of inputs and outputs, and constructs an supervised machine learning model to learn that relationship, or approximate it to an acceptable degree. A well trained model will then be able to take in unseen inputs and return its modeled targets. For a cosmologically motivated example, one may want to construct a model which takes as its input a set of parameters defining a cosmology and outputs the expected CMB power spectrum, the set of C_l's as defined in eq. (1.66), of a universe with the input cosmology.

The process of learning takes place as follows. The supervised machine learning model begins as a randomly or pseudo-randomly constructed model. A measure of how well the model accurately describes the data is calculated, and model is altered in a way so that the new model better describes the data

than its previous iteration. This is one *epoch*, or generation, of training. This process repeats until the model is deemed good enough, or a fixed number of epochs has passed. To facilitate this learning, supervised machine learning models consist of two primary components: the "goodness" of fit metric, or the *loss function*, and the iterative learning step.

The loss function of a model is a quantitative measure of how good the model is, i.e. how well the model has learned the data. The loss function takes as its input a true expected target value and the model's own predicted target value, for a given set of data points, and returns a measure of how "close" the two are. One example of a loss function is the mean absolute error (MAE). For a dataset D, where the output y is a real number, the MAE is

$$\text{MAE} = \frac{\sum_{i=1}^{n} |\vec{y}_i^{\,\text{pred}} - \vec{y}_i^{\,\text{target}}|}{n}, \tag{2.2}$$

where $\vec{y}_i^{\,\text{target}}$ is the actual target, while $\vec{y}_i^{\,\text{pred}}$ is the model's prediction for the i^{th} data point. The loss function gives one the ability to measure the model's performance, and allows for a criteria to measure a model's improvement. If the chosen loss function was MAE, a good model should have a smaller MAE than a poor model.

There is one important detail to note when calculating a model's accuracy as previously described. If a model's loss is calculated using the data it was trained on, the loss only describes how well the model predicts to the *training data*, and gives no indication on how well the model may generalize to unseen data. Let's consider a brief example. If we wish to model a physical process where an input data point 1 maps to the target 0, and 3 maps to the target 4, we can very easily construct a quadratic expression to describe the relationship between the inputs x_i and the targets y_i as

$$y_{\text{pred}} = 4 - (x_{\text{input}} - 3)^2\,, \tag{2.3}$$

which results in a MAE of 0 for the given points. However, if the data points $(5, 10)$ and $(6, 14)$ are generated by the same physical process, the model does not accurately predict these values as well as the training points. This common issue is known as *over-fitting*. Machine learning models may learn the training set perfectly, but are useless at predicting unseen data if they cannot generalize. One way to keep track of potential over-fitting is to reserve a separate validation dataset, $D_{\text{validation}}$. The model continues learning and iterating on D_{train}, but the reported loss is calculated using the data points

in $D_{\text{validation}}$. Furthermore, a general rule of thumb to combat over-fitting is to have more data points than number of free model parameters. There are infinitely many cubic polynomials that can fit any two points, but only one unique cubic polynomial that can fit any three points. For four or more points, there may not exist any cubic polynomial that can fit them all perfectly.

The second component of supervised machine learning models is the iterative step. This is an algorithmic procedure for altering the model, so that the new model is better than the old model, according to the chosen loss function. This is required for the model to improve in each subsequent epoch, and eventually "learn" the data.

One popular algorithm for iterating machine learning models is known as gradient descent. Gradient descent outlines a procedure for evolving machine learning models that are constructed out of differentiable mathematical expressions. In general, the loss value of a predicted target is the function $L(\vec{y}_{\text{target}}, \vec{y}_{\text{pred}})$, defined so that a smaller value of L is better. The constructed model for predicting $\vec{y}_{\text{pred}}$ is a differentiable mathematical expression, $\vec{y}_{\text{pred}} = f(\vec{x}; \alpha_i)$, where α_i are tunable parameters, such as the coefficients of an n^{th} degree polynomial. In order to improve the model's prediction, the parameters α_i need to be updated at each step so that the loss L is less for each subsequent epoch of evolution. This amounts to searching for minima of the function $L(\alpha_i)$. Mathematically, the parameters are updated as

$$\alpha_i \to \alpha_i - \gamma \sum_{j=1}^{n} \frac{\partial L}{\partial \alpha_i} \left(\vec{y}_j,\ \alpha_i \right), \tag{2.4}$$

where γ is the learning rate, which controls how much the parameters are updated at each step. This γ is known as a *hyperparameter*, a parameter set by the user which controls various aspects of the behavior of the machine learning model.

The above paragraphs outline the general principles that underlie machine learning models. Next, I will discuss genetic algorithms, an alternative approach to gradient descent models.

2.1 Genetic Algorithms and Symbolic Regression

Genetic Algorithms (GAs) are a class of machine learning technique designed to mimic evolutionary biology [11]. As opposed to gradient descent-based machine learning models, GAs consider a population of candidate models which are not necessarily differentiable mathematical expressions. The user simply needs to define the template of the candidate model. The best few among these models, the most *fit*, are chosen to construct the next population of models. Two operations facilitate the learning, random mutation so that the candidate model space is being explored reasonably well, and crossover so that the GA converges to a model most fit to describe the dataset.

Random mutation randomly alters components, or genes, of the candidate models, in order to explore the model space in a stochastic manner. If the candidate model were a mathematical expression with variable parameters, a random mutation may multiply one of the parameters by a random value. If the form of mathematical expression itself can be altered, as is the case in symbolic regression, random mutation may change operations at random, for example changing $\alpha_1 + x \rightarrow \alpha_1 \times x$. Random mutation thus introduces slight changes in the population of functions, changes which may increase the model's fitness and propagate through to the next epoch.

Given a loss function, as defined above, the most fit few models in a population survive, mimicking the "survival of the fittest" paradigm of evolutionary biology. Fit models have the chance to crossover, i.e. have their genes mixed in a pseudo-random manner, producing children candidate models. The motivating idea is that combining various aspects of two distinct strong models may produce an offspring candidate model which is even better than its parents.

One of GAs principal strengths lies in this flexibility. Whereas gradient descent-based models require their candidate model to be a differentiable mathematical expressions, GAs allow for a more abstract definition of the candidate models' forms. One simply needs to define the template of the candidate model, and crossover and random mutation operations that generate new models. I will illustrate these ideas now by describing Symbolic Regression (SR), a type of GA where the candidate models are mathematical expressions without a fixed form.

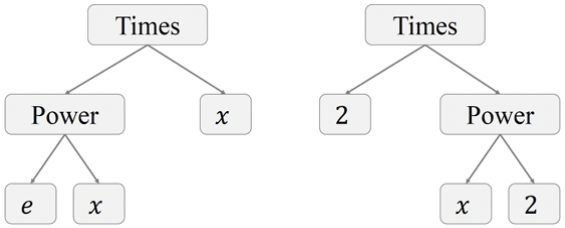


Figure 2.1: Two mathematical expressions represented as expression trees. Figure from [3].

2.1.1 PySR

In this work, I use the Python package PySR [4] to implement symbolic regression. I will describe briefly how SR in PySR works.

Firstly, the candidate models are different analytical expressions. For example one candidate expression may be an exponential fit, whereas another may be a polynomial fit. Candidate expressions are stored as expression trees, as shown in fig. 2.1. Expression trees are a data structure consisting of parent nodes, in this case operators, and left and right child nodes. Child nodes may be sub-expression trees or leaf nodes. Leaf nodes are terminal, as they have no further child nodes, and are either constant values or variables. In fig. 2.1, the left expression tree depicts the expression $e^x x$, while the right expression tree is $2x^2$.

As mentioned above the three components of an expression tree are binary operators, variables, and constants. Binary operators are elementary operations such as addition and multiplication, which take two input values. The binary operator is evaluated with it's two child nodes as inputs. For example, in the left expression in fig. 2.1 the multiplication operator acts between its child nodes to render the expression $(e^x) \times (x)$. It is important to note that while some binary operators may be commutative, this is not always true. Swapping the left and right child nodes of the multiplication operator does not change the overall expression, while swapping inputs for exponentiation will change the resulting expression, e.g. $(e) \wedge (x) \neq (x) \wedge (e)$. The operators available to the SR algorithm in its search for expressions are specified in a user-defined *grammar*.

Variables are inputs into the expression, that correspond to the input variables in the training dataset, e.g. the components x_i of the input vectors $\vec{x}$.

The variables available to the SR algorithm are specified by the user, and may or may not be all the components of the input vector. Constants are simply real numbers, as not only the form but also the constant values in a mathematical expression are important to tune.

Given the possible expressions, SR needs well-defined random mutation and crossover mechanisms. In order to randomly mutate an expression, one of the following mutations is randomly selected:

- Randomize Constant: A constant in the expression is selected at random and is changed to a random value.
- Randomize Operator: An operator in the expression is selected at random and is changed to a different operator from the specified grammar.
- Append/Prepend Node: A random expression is generated, and either a leaf node in the original expression is swapped with the new expression, or the root node (the node without a parent node) of the original expression is made the child node of the new expression.
- Insert Node: A random expression is generated and inserted in place of a randomly selected node anywhere in the expression tree.
- Delete Subtree: An interior node of the expression tree (a node with both a parent and children nodes) is replace either by a variable or constant.
- New Tree: The entire expression tree is replaced at by another randomly generated expression tree.
- Do Nothing: No mutation is performed.

In PySR, there are many tunable hyperparameters associated with the mutation scheme, such as the probability of any given one of the mutations above being selected.

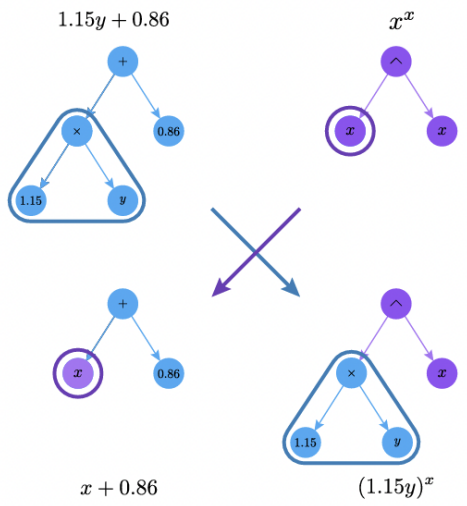


Figure 2.2: The crossover operation between two expressions. Figure from [4].

The crossover mechanism between two candidate expressions in PySR is illustrated in fig. 2.2. Copies are made of both parent expressions, and a random interior or leaf node is selected in both of them. These nodes are swapped across expressions, and the resulting two new expressions are the output of the crossover mechanism.

PySR selects expressions for random mutation and crossover using a tournament method. The tournament method involves selecting a random subset of the population of models and choosing the fittest member, according to some loss function. In PySR, the default metric is the mean-squared error, given by

$$\text{MSE} = \frac{\sum_{i=1}^{n}(\vec{y}_i^{\,\text{pred}} - \vec{y}_i^{\,\text{target}})^2}{n}. \tag{2.5}$$

Once two fit expressions are selected and crossed over, two other expressions must be selected to be replaced. In PySR, the oldest members of the population are replaced by these new crossed over expressions.

PySR initializes n_p independent populations of expressions, and each population evolves mostly on its own, in parallel. After each epoch, the best members in each population are selected have a chance to migrate to other populations, to facilitate some mixing between populations. This is so that a wider extent of the model space is explored than one single isolated population might be able to.

In brief, the PySR algorithm is as follows:

1. The user gives PySR a training dataset, a grammar, and the option to tune various hyperparameters.
2. n_p populations are initialized at random, each containing n_e expressions.
3. A tournament is performed on a randomly selected subset of expressions in the population, and either a mutation or crossover is performed. If crossover, the new children models replace the two oldest expressions in the population.
4. Step 3 is performed M times.
5. Steps 3-4 are performed in parallel for each of the n_p populations.
6. The best expressions seen across each population are stored in a hall of fame. These hall of fame expressions then have a chance of being introduced into each population at random.
7. Steps 3-6 comprise one generation. n_g generations of evolution are carried out
8. The final hall of fame, comprised of the best expressions seen, is returned to the user.

PySR's unique ability to produce analytic fits to data makes it a powerful tool. As a test, PySR has been utilized to successfully reproduce famous physics formulae, such as Newton's law of gravitation and Leavitt's law [4]. In the next chapter, I use PySR towards constructing an analytic approximation to the CMB lensing tail.

Chapter 3

Compressing Analytic Expressions

3.1 Part 1: The Matter Transfer Function

The matter transfer function, like the CMB power spectrum, is an excellent probe of cosmological models. Just as we can extract valuable cosmological information from the CMB power spectrum and its inherent temperature anisotropies, so too can we study the matter distribution and its inhomogeneities. Matter is not uniformly distributed in the universe, this much is clear from observing the fact that there is plenty of matter here on Earth, while nearly none in the vacuum of space a mere 200 miles above the surface. While cosmology cannot quite probe on scales this small, the analogy still holds well on larger scales. There are pockets of space rife with matter, such as galaxy clusters. On the other hand, there are also cosmological voids, vast spaces that contain little to no galaxies. The information on how matter is distributed at different scales in the universe is contained in the *matter power spectrum*, $P(k)$, in a somewhat analogous manner to the CMB power spectrum. In this section, I motivate the linear matter power spectrum, and in particular focus on the *matter transfer function*, a key component of the linear matter power spectrum.

In order to discuss how matter is distributed in the universe, it is useful to know what the gravitational potential looks like. This is because gravitational instability sources inhomogeneity in the matter distribution. One can arrive at this fact in simpler terms by noting that matter is pulled towards gravitational potentials. Thus, matter is pulled to over-dense regions under the influence of gravity, and the excess matter in turn increases the gravitational

potential. From this, we can extrapolate the fact that matter accumulates in initially over-dense regions, regardless of how small the initial over-density was. We take again the perturbed FRW metric of eq. (1.29), and consider the gravitational potential $\Phi(\vec{k}, a)$. As earlier, working in Fourier space is more convenient.

In order to understand the distribution of matter today, it is important to consider how the evolution of the universe affects the gravitational potential. As Fourier modes, large k correspond to smaller distance scales, and vice versa. Why should modes of different modes evolve differently? In order to understand this, we need to recall from Chapter 1 that η, the conformal time, can also be interpreted as the distance traveled by light in the expanding frame of the universe. Any two points separated by a distance larger than η are not causally connected, since a photon could not have traveled from one to the other. In this sense, η is known as the *particle horizon*, or simply the horizon. Thus we should focus on k modes corresponding to a wavelength shorter than η, those larger are not subject to a causal evolution beyond their initial values. As the horizon grows over time, wavelength modes fall into the horizon, and their evolution is then governed by dominating component of the universe at that time.

At early times, the horizon was smaller than it is today, and nearly all modes of interest would have been outside the horizon, $k\eta << 1$, close to their initial values. At intermediate times, the horizon expands and wavelengths fall into the horizon. Also during this time, the universe passes from radiation domination to matter domination. Modes that enter the horizon during radiation domination will evolve differently than modes that enter during matter domination. At matter-radiation equality the two components' densities are equal. At late times, all modes of interest are in the horizon and evolve identically as matter-domination dictates. Thus, we expect the gravitational potential to have some non-trivial scale dependence.

Following the line of thought above, we can schematically represent the the full gravitational potential as the composition of three distinct components:

$$\Phi(\vec{k}, a) = \Phi_p(\vec{k}) \times T(k) \times D_+(a). \tag{3.1}$$

The Φ_p term is the initial over-density mentioned above, the primordial perturbation[1]. All scale-dependence is captured in the transfer function, $T(k)$, and the growth of the modes over time is captured in the growth function $D(a)$. $T(k)$ is normalized so that

$$\lim_{k\to 0} T(k) = 1. \tag{3.2}$$

The matter overdensity is defined as

$$\delta_m(k,\ a) = \frac{\rho_m(k,\ a) - \bar{\rho}_m(a)}{\bar{\rho}_m(a)}, \tag{3.3}$$

where $\bar{\rho}_m(a)$ is the mean, or background, density of matter at a given time, and $\rho_m(k,\ a)$ is the actual density at a given k-mode and time. δ_m is then a measure of how matter across different Fourier modes is distributed relative to the mean density, $\delta > 0$ corresponds to an overdense region, while $\delta < 0$ corresponds to an underdense region. The matter overdensity can be related to the gravitational potential via a form of Poisson's Equation, namely

$$\Phi(k,\ a) = \frac{4\pi G a^2 \rho_m \delta_m}{k^2}. \tag{3.4}$$

Collecting the k-dependence, we see that the overdensity is related to the gravitational potential as

$$\begin{aligned} \delta_m(k) &\propto k^2 \Phi(k) \\ &\propto k^2 \Phi_p(k) T(k). \end{aligned} \tag{3.5}$$

In ΛCDM, the initial perturbation takes the form

$$k^3 \Phi_p^2(k) = A_s k^{n_s - 1} \tag{3.6}$$

where A_s, the primordial perturbation amplitude, and n_s, known as the *spectral index*, are both in ΛCDM to be fit to observed data. With this, the k-dependence of the matter overdensity is

$$\delta_m(k) \propto k^{n_s/2} T(k). \tag{3.7}$$

[1]The inflation mechanism[12] is one proposed explanation for how the primordial perturbation might have been generated.

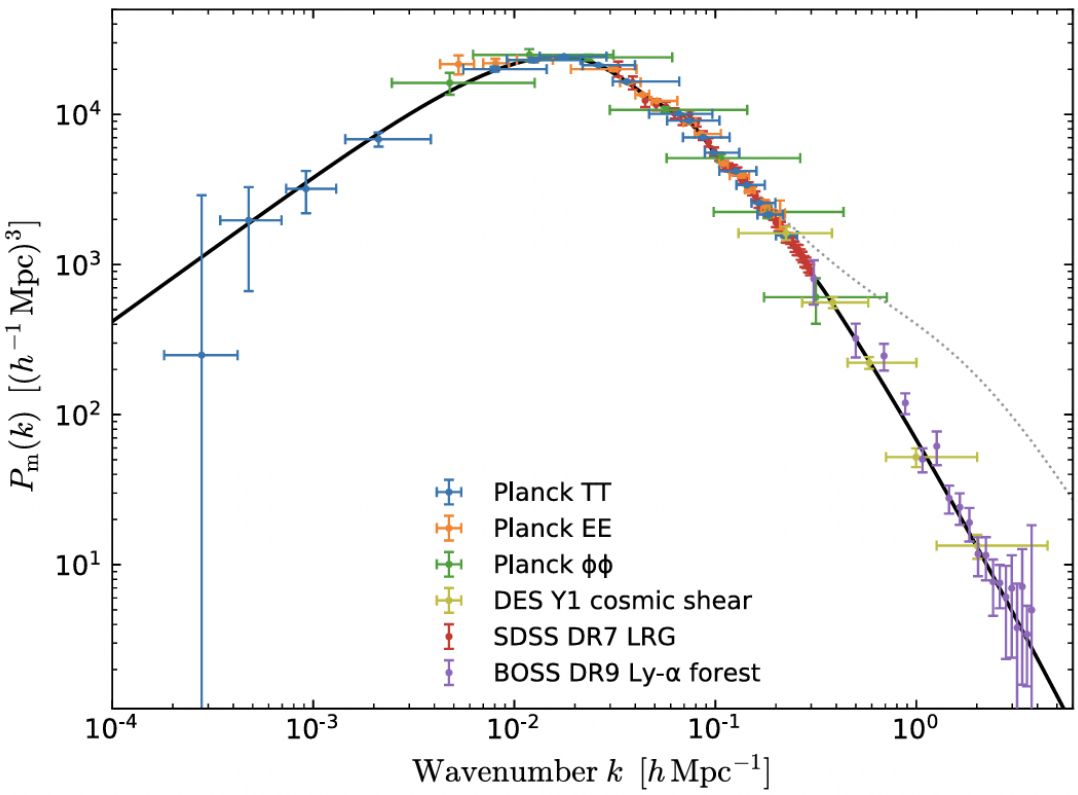


Figure 3.1: Constraints on the matter power spectrum by various experiments. The solid black line shows the power spectrum as predicted by ΛCDM assuming linear perturbations. The impact of nonlinearities is depicted by the dotted line. Figure from [2].

The *linear* regime[2] is when $\delta_m << 1$, and can be effectively treated as a perturbation. The linear matter spectrum is the two-point correlation function of the matter overdensity in Fourier space, so that

$$(2\pi)^3\delta^3(k-k')P_m(k) = \langle \delta_m(k)\ \delta_m(k')\rangle, \tag{3.8}$$

so that

$$P_m(k) \propto \delta_m^2(k). \tag{3.9}$$

Collecting eq. (3.9) and eq. (3.7), we find that

$$P_m(k) \propto k^{n_s}T^2(k). \tag{3.10}$$

[2]While radiation remains linear throughout all epochs and at all scales of interest, the same is not true for matter.

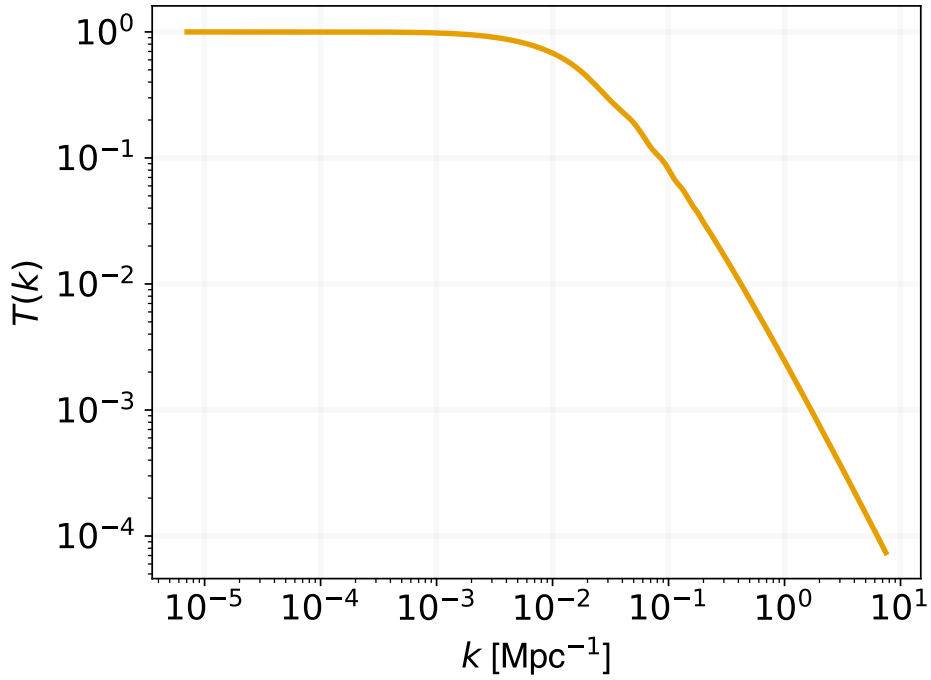


Figure 3.2: CAMB generated matter transfer function for the Planck best fit cosmology, where $\Omega_b h^2 = 0.0224$ and $\Omega_c h^2 = 0.12$.

Thus, the form of the power spectrum is entirely determined by the transfer function. Fig. 3.1 depicts constraints on $P_m(k)$ by various cosmological probes, against the ΛCDM predicted curve plotted as the solid black line. It's clear that the linear matter power spectrum is an excellent tool for verifying cosmological theories.

In the above section I lay out the importance of the matter transfer function. In the next section, I explore the various ways of computing the matter transfer function.

3.1.1 Computing $T(k)$

CAMB

Figure 3.2 shows the matter transfer function $T(k)$ for the Planck best fit cosmology, the best fit of the ΛCDM parameters as determined by the Planck experiment[2]. In particular the parameter values are $\Omega_b h^2 = 0.0224$, $\Omega_c h^2 = 0.12$ and $H_0 = 67.4$.

Figure 3.2 was generated by CAMB, or the Code for Anisotropies in the Microwave Background[13]. CAMB is a powerful computational package which solves the intertwined set of coupled differential equations resulting from the Boltzmann equation eq. (1.26) for the various components of the universe. CAMB can be used to compute quantities such as the CMB power spectrum and the matter power spectrum to a high precision.

In this thesis, I utilize CAMB to construct the matter transfer function, as above, and in a later section the CMB lensing tail. I use CAMB constructed quantities to use in machine learning model verification, so that the best fitting formula, say for the transfer function, is the one which minimizes its error with respect to the CAMB transfer function. While CAMB is a powerful tool, it is always nice to have analytic or semi-analytic formulations of cosmological quantities. There are two popular formulas for the matter transfer function, one constructed by Bardeen, Bond, Kaiser, and Szalay (BBKS) [14], and the other by Eisenstein and Hu (EH) [15], which I shall now discuss in turn.

BBKS

As mentioned in the previous section, a nontrivial scale dependence appears in the matter power spectrum because perturbations evolve differently depending on which component dominates the universe when they enter into the horizon. Modes entering the horizon during radiation-domination will have a qualitatively different evolution than modes which enter during matter-domination. Furthermore, when exactly the universe transitions from radiation-domination to matter-domination depends, in large part, on the matter density parameter, $\Omega_m h^2$. Because ordinary matter and dark matter evolve differently[3], it is important to consider both the dark matter and ordinary matter densities, $\Omega_c h^2$ and $\Omega_b h^2$ respectively.

With these considerations in mind, the BBKS formula is a simple approximation for the transfer function when dark matter dominates the total matter density, or when $\Omega_b << \Omega_m$. It is given by

$$T_{\text{BBKS}}(q) = \frac{\ln(1+2.34q)}{2.34q}\left(1+3.89q+(16.1q)^2+(5.46q)^3+(6.71q)^4\right)^{-1/4}, \tag{3.11}$$

[3]Dark matter is collisionless, while ordinary matter is not, as in the photon-baryon coupling discussed earlier.

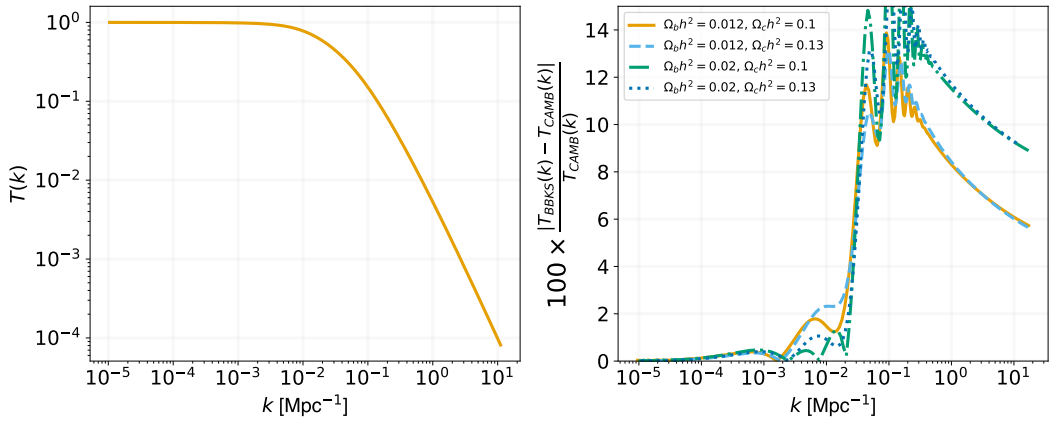


Figure 3.3: *Left*: BBKS approximation for the matter transfer function for the Planck best fit cosmology, where $\Omega_b h^2 = 0.0224$ and $\Omega_c h^2 = 0.12$. *Right*: The percent error between the BBKS approximation and the CAMB generated matter transfer functions for four cosmologies.

where q is defined as

$$q \equiv \sqrt{\frac{\rho_r}{1.68\rho_\gamma}}\frac{k}{\Omega_c h^2}, \tag{3.12}$$

and ρ_r is the background density of relativistic particles (photons and neutrinos) and ρ_γ is the background density of photons. Using the relation between these densities given by [9] in eq. (2.77), namely,

$$\rho_\nu = 3\frac{7}{8}\left(\frac{4}{11}\right)^{4/3}\rho_\gamma, \tag{3.13}$$

we can rewrite q as

$$q \equiv \sqrt{\left(1 + 3\frac{7}{8}\left(\frac{4}{11}\right)^{4/3}\right)\frac{1}{1.68}}\frac{k}{\Omega_c h^2}. \tag{3.14}$$

The left plot in fig. 3.3 shows the BBKS approximation of the Planck best fit cosmology, i.e. the BBKS approximation of the curve in fig. 3.2. The $T_{\text{BBKS}}(k)$ approximates $T(k)$ with an average error within 15%, and works exceptionally well up to around $k = 10^{-2}\text{Mpc}^{-1}$.

EH

A more accurate fitting formula was described by Eisenstein and Hu in [15]. While the $T_{\mathrm{BBKS}}(k)$ is simple and matches the large scale behavior relatively well, it performs poorer at smaller scales, particularly so in the neighborhood of k_{eq}. This is the scale which enters the horizon at matter-radiation equality, when the universe switches from radiation-domination to matter-domination. [9] gives an approximate expression for k_{eq} as

$$k_{\mathrm{eq}} = 0.073\mathrm{Mpc}^{-1}\Omega_m h^2, \tag{3.15}$$

according to which $k_{\mathrm{eq}} = 0.007\ \mathrm{Mpc}^{-1}$ in the Planck best fit cosmology of fig. 3.2. The EH approximation recognizes that the transition from radiation-domination to matter-domination is a complex process. Qualitatively this is visible in fig. 3.2 near k_{eq}, where there are small wiggles in the transfer function. Inability to account for this behavior leads to the oscillations in the BBKS error as seen in the right panel of fig. 3.3.

The EH fitting formula is motivated by considering the effects of phenomena such as acoustic oscillations, Compton drag, velocity overshoot, baryon infall, adiabatic damping, Silk damping, CDM suppression. In order to accurately account for all these terms, the EH approximation is comprised of around 30 expressions. The EH transfer function takes the form

$$T_{\mathrm{EH}}(k) = \frac{\Omega_b}{\Omega_m} T_b(k) + \frac{\Omega_c}{\Omega_m} T_c(k). \tag{3.16}$$

The terms involved in this formula are the following:

$$\begin{aligned}
T_b &= \left[\frac{\tilde{T}_0(k;1,1)}{1+(ks/5.2)^2} + \frac{\alpha_b}{1+(\beta_b/ks)^3} e^{-\left(\frac{k}{k_{\mathrm{Silk}}}\right)^{1.4}} \right] j_0(k\tilde{s}), \\
T_c &= f\tilde{T}_0(k,1,\beta_c) + (1-f)\tilde{T}_0(k,\alpha_c,\beta_c), \\
T_0(k;\alpha_c,\beta_c) &= \frac{\ln(e+1.8\beta_c q)}{\ln(e+1.8\beta_c q) + Cq^2}, \\
f &= \frac{1}{1+(ks/5.4)^4}, \\
C &= \frac{14.2}{\alpha_c} + \frac{386}{1+69.9q^{1.08}}, \\
q &= \frac{k}{13.41 k_{\mathrm{eq}}},
\end{aligned}$$

$$
\begin{aligned}
R &= 31.5\Omega_b h^2 \Theta_{2.7}^{-2} \mathrm{Mpc}^{-1}, \\
k_{\mathrm{Silk}} &= 1.6(\Omega_b h^2)^{0.52}(\Omega_m h^2)^{0.73}\left[1+(10.4\Omega_m h^2)^{-0.95}\right] \mathrm{Mpc}^{-1}, \\
k_{\mathrm{eq}} &= 7.46\times 10^{-2}(\Omega_m h^2)\Theta_{2.7}^{-2}\mathrm{Mpc}^{-1}, \\
\alpha_b &= 2.07 k_{\mathrm{eq}} s (1+R_d)^{-3/4} G\left(\frac{1+z_{\mathrm{eq}}}{1+z_d}\right), \\
\beta_b &= 0.5 + \frac{\Omega_b h^2}{\Omega_m h^2} + \left(3 - 2\frac{\Omega_b}{\Omega_m}\right)\sqrt{(17.2\Omega_m h^2)^2+1}, \\
\beta_{\mathrm{node}} &= 8.41(\Omega_m h^2)^{0.435}, \\
s &= \frac{2}{3k_{\mathrm{eq}}}\sqrt{\frac{6}{R_{\mathrm{eq}}}}\ln\frac{\sqrt{1+R_d}+\sqrt{R_d+R_{\mathrm{eq}}}}{1+\sqrt{R_{\mathrm{eq}}}}, \\
\tilde{s} &= s\left[1+\left(\frac{\beta_{\mathrm{node}}}{ks}\right)^3\right]^{-1/3} \\
G(y) &= -6y\sqrt{1+y} + y(2+3y)\ln\left(\frac{\sqrt{1+y}+1}{\sqrt{1+y}-1}\right), \\
y &\equiv \frac{1+z_{\mathrm{eq}}}{1+z}, \\
\alpha_c &= a_1^{-\Omega_b/\Omega_m} a_2^{-(\Omega_b/\Omega_0)^3}, \\
a_1 &= (46.9\Omega_m h^2)^{0.670}\left[1+(32.1\Omega_m h^2)^{-0.532}\right], \\
a_2 &= (12.0\Omega_m h^2)^{0.424}\left[1+(45.0\Omega_m h^2)^{-0.582}\right], \\
\beta_c^{-1} &= 1 + b_1\left[(\Omega_c/\Omega_m)^{b_2}-1\right], \\
b_1 &= 0.944\left[1+(458\Omega_m h^2)^{-0.780}\right]^{-1}, \\
b_2 &= (0.395\Omega_m h^2)^{-0.0266}, \\
z_{\mathrm{eq}} &= 2.50\times 10^4 \Omega_m h^2 \Theta_{2.7}^{-4}, \\
z_d &= 1291\frac{(\Omega_m h^2)^{0.251}}{1.+0.659(\Omega_m h^2)^{0.828}}\left[1+b_{1,\mathrm{z}}(\Omega_b h^2)^{b_{2,\mathrm{z}}}\right], \\
b_{1,\mathrm{z}} &= 0.313(\Omega_m h^2)^{-0.419}\left[1+0.607(\Omega_m h^2)^{0.674}\right], \\
b_{2,\mathrm{z}} &= 0.238(\Omega_m h^2)^{0.223},
\end{aligned}
$$

where $T_{\mathrm{CMB}} \equiv 2.7\Theta_{2.7}\mathrm{K}$, $R_d \equiv R(z_d)$ and $R_{\mathrm{eq}} \equiv R(z_e q)$, and z_d and z_{eq} are the redshifts of the drag epoch and of matter-radiation equality, respectively.

$T_{\mathrm{EH}}(k)$ is a powerful approximation of $T(k)$, with an average error within 5%. The left plot in fig. 3.4 shows the EH approximation of the Planck best fit cosmology, i.e. the EH approximation of the curve in fig. 3.2. The right

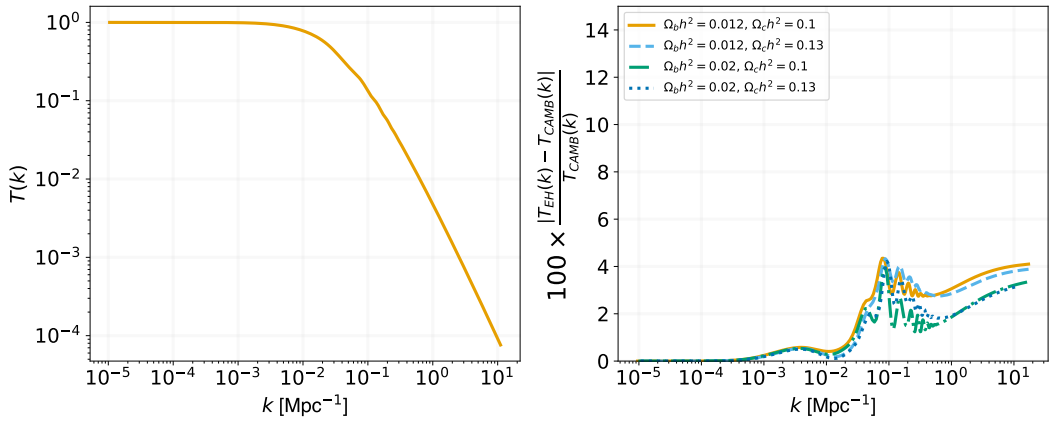


Figure 3.4: *Left*: EH approximation for the matter transfer function for the Planck best fit cosmology, where $\Omega_b h^2 = 0.0224$ and $\Omega_c h^2 = 0.12$. *Right*: The percent error between the EH approximated and CAMB generated matter transfer functions for four cosmologies.

plot shows the point-wise error of $T_{\text{EH}}(k)$ when measured against $T_{\text{CAMB}}(k)$, for four sample cosmologies. The careful consideration of the various physical phenomena in $T_{\text{EH}}(k)$ results in an order of magnitude improvement over the BBKS approximation.

While powerful, $T_{\text{EH}}(k)$ is fairly complex. An interesting question posed by the authors in [3] is whether it's possible to find a formula for the matter transfer function that is as accurate as $T_{\text{EH}}(k)$, but with a simpler form. Following their work, I implement a genetic algorithm to learn a simpler form of the matter transfer function.

3.1.2 Learning $T(k)$

In this section, I follow [3], using their simplified function template for the matter transfer function along with my own GA implementation in order to learn a strong fit. I will discuss the template function they use, describe my GA implementation, and finally I compare the results of my function against $T_{\text{EH}}(k)$.

The Function Template

In [3], the authors construct a function template for the matter transfer function of the form

$$T_{\mathrm{GA}}(k) = (1 + \alpha_1 x^{\alpha_2} + \alpha_3 x^{\alpha_4} + \alpha_5 x^{\alpha_6} + \alpha_7 x^{\alpha_8})^{-1/4}\,, \tag{3.17}$$

where

$$x = \frac{k}{\Omega_c h^2 \mathrm{Mpc}}. \tag{3.18}$$

The physical form is similar to that of $T_{\mathrm{GA}}(k)$ with the outer exponent of $-1/4$, but differs in that it allows the inner exponents to vary. As an initial test of GA, I take this function template and use GA to learn values of the α_i.

The Genetic Algorithm

In the last chapter, I discussed the key ingredients in GA. These are, the candidate model, the training dataset, the loss function, and the mutation and crossover operations. I will now describe my implementation of these pieces.

Each candidate model in my GA should be a function of the form eq. (3.17), with differing parameter values. In code, each candidate model is an object represented by the array

$$\alpha = [\alpha_1,\ \ldots,\ \alpha_8]\,, \tag{3.19}$$

and is equipped with a function which evaluates eq. (3.17) with the candidate's respective parameter array on a given input of k values. In the terminology of GA, these parameters are known as genes.

I used CAMB to generate the training dataset. I generated 300 data points of the form

$$d_i = \{\Omega_b^i h^2,\ \Omega_c^i h^2,\ \vec{k^i},\ \vec{T}^i_{\mathrm{CAMB}}\}, \tag{3.20}$$

where

$$\vec{k^i} = \left(k_1^i,\ \ldots,\ k_n^i\right) \tag{3.21}$$

are the values of k given by CAMB, and the respective values of the CAMB generated transfer function,

$$\vec{T}^i_{\mathrm{CAMB}} = \left(T^i_{\mathrm{CAMB,1}},\ \ldots,\ T^i_{\mathrm{CAMB,n}}\right), \tag{3.22}$$

for a cosmology with baryonic and cold dark matter density parameters $\Omega_b^i h^2$ and $\Omega_c^i h^2$, and where $T^i_{\mathrm{CAMB,j}} \equiv T^i_{\mathrm{CAMB}}(k^i_j)$. For all data points, I chose

the Planck best fit value for the Hubble parameter, $H_0 = 67.4$, and selected values of the density parameters at random from within the ranges $\Omega_b h^2 \in [0.0214, 0.0234]$ and $\Omega_c h^2 \in [0.12, 0.14]$, the same ranges as in [3].

The loss function I chose was the mean relative error over all data points d_i. The error E_i of a candidate model $T_{\rm GA}$ evaluated on a data point d_i is given by

$$E_i = \frac{1}{n} \sum_j \frac{\left| T^i_{\rm GA}(k_j) - T^i_{\rm CAMB,j} \right|}{T^i_{\rm CAMB,j}}. \tag{3.23}$$

Functions with a lower error over the entire dataset,

$$E_{\rm tot} = \sum_i E_i, \tag{3.24}$$

are considered more fit.

If a candidate function is selected for mutation, one of the following actions is selected:

1. A single, randomly selected, gene is mutated.
2. A random subset of the genes is mutated.
3. The candidate model is replaced by an entirely new, randomly generated, candidate model.
4. No mutation occurs.

with probabilities $p_1 = 0.65$, $p_2 = 0.20$, $p_3 = 0.05$, $p_4 = 0.10$, respectively. When a gene is selected for mutation, it is multiplied by a factor

$$\alpha_{\rm i,mutated} = q \times \alpha_{\rm i,old}, \tag{3.25}$$

where

$$q = s \times (1.1 + T)^r, \tag{3.26}$$

where $s = -1(+1)$ with probability $p = 0.5(0.5)$, $r \in [-1, 1]$ is uniformly randomly sampled, and T is an annealing temperature. T is given by

$$T = 1 - \frac{i}{M}, \tag{3.27}$$

for the i^{th} mutation performed out of the M total mutation operations per generation. This term is included so that mutations earlier on in a generation have a larger magnitude than those later on in a generation.

The crossover operation between two selected candidate models with genes $\alpha_a = [\alpha_1^a,\ \ldots,\ \alpha_n^a]$ and $\alpha_b = \left[\alpha_1^b,\ \ldots,\ \alpha_n^b\right]$ produces children models according to the rule

$$\alpha_a,\ \alpha_b \to \left[\alpha_1^a,\ \ldots,\ \alpha_{n/2}^a,\ \alpha_{n/2+1}^b,\ \ldots,\ \alpha_n^b\right],\ \left[\alpha_1^b,\ \ldots,\ \alpha_{n/2}^b,\ \alpha_{n/2+1}^a,\ \ldots,\ \alpha_n^a\right] \tag{3.28}$$

so that each child model derives half its genes from one parent, and half from the other.

In implementing my GA, I take inspiration from the PySR algorithm. My algorithm runs as follows

1. The user gives the GA a training dataset with data points as in eq. (3.20), and a template function format, in this case eq. (3.17).
2. A population of n_p random expressions is created.
3. A tournament is performed on a randomly selected subset of the population of size s_t, and either a mutation or crossover with probability p_c is performed. If crossover, the new children models replace the two oldest expressions in the population.
4. Step 3 is performed M times in each generation.
5. n_g generations of evolution are carried out.
6. By the end of n_g generations, the most fit function found by GA is returned.

Results

The table below gives the values of the hyperparameters I set when running the genetic algorithm.

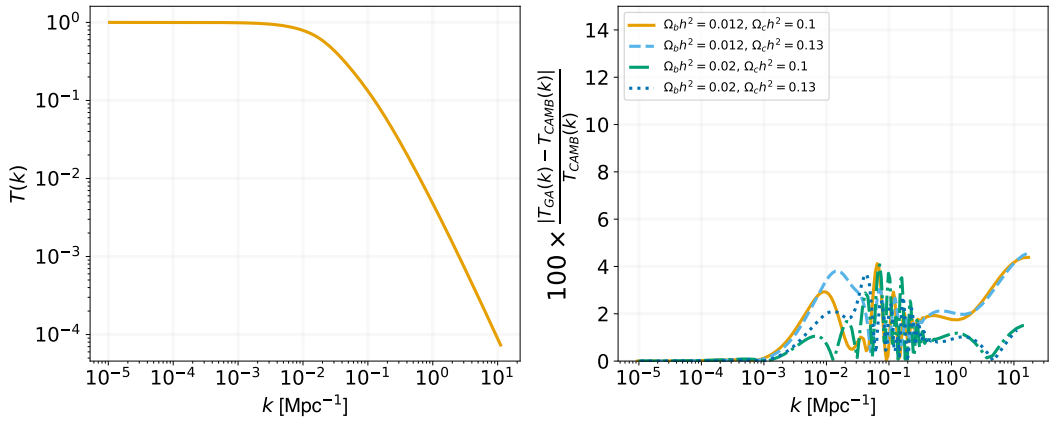


Figure 3.5: *Left*: The GA learned approximation for the matter transfer function for the Planck best fit cosmology, where $\Omega_b h^2 = 0.0224$ and $\Omega_c h^2 = 0.12$. *Right*: The percent error between T_{GA} and T_{CAMB} for four sample cosmologies.

Hyperparameter	Symbol	Value
Number of Generations	n_g	200
Mutations per Generation	M	250
Population Size	n_p	100
Probability of Crossover	p_c	0.01
Tournament Subset Size	s_t	15

Running this algorithm on the function format in eq. (3.17), GA learned the following transfer function

$$T_{\text{GA}}\left(k;\ \Omega_b h^2,\ \Omega_c h^2\right) = \left(1 + 58.2x^{1.52} + 4380x^{3.96} + 3160x^{6.03} + 143x^{7.30}\right)^{-1/4}, \tag{3.29}$$

where again x is again a defined in eq. (3.18).

Following the previous figures, fig. 3.5 shows $T_{\text{GA}}(k)$ for the Planck best fit cosmology, as well as it's error against $T_{\text{CAMB}}(k)$ for some sample cosmologies. Remarkably, the error in $T_{\text{GA}}(k)$ is on the same order as $T_{\text{EH}}(k)$. Furthermore, $T_{\text{GA}}(k)$ achieves this accuracy while only depending on the cold dark matter density parameter $\Omega_c h^2$.

Figure 3.6 summarizes the accuracy of the three fitting formulas described in this section, by showing their errors against $T_{\text{CAMB}}(k)$ on the Planck best fit cosmology. It is useful in summarizing the order of magnitude accuracy of the

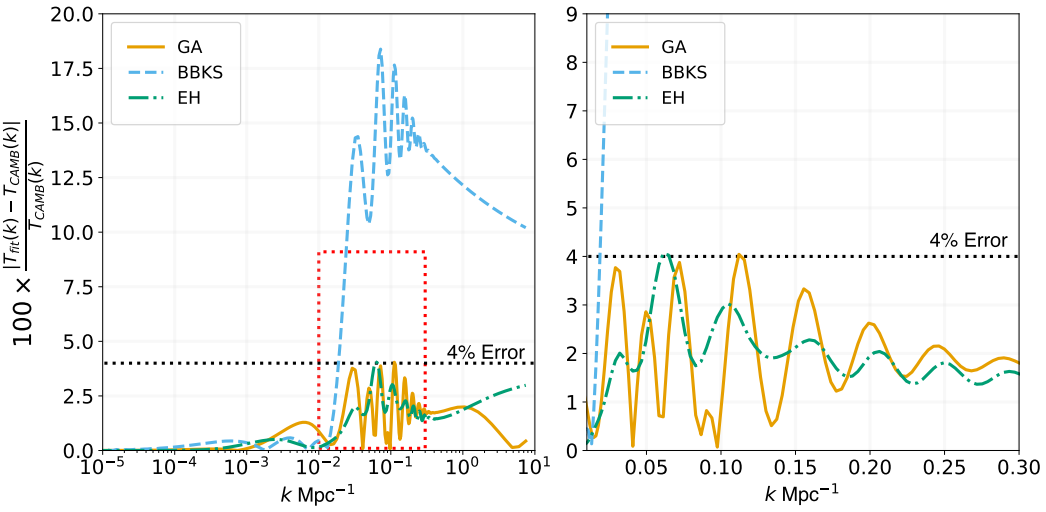


Figure 3.6: *Left*: The percent error of $T_{\mathrm{GA}}(k)$, $T_{\mathrm{BBKS}}(k)$ and $T_{\mathrm{EH}}(k)$ against $T_{\mathrm{CAMB}}(k)$ for the Planck best fit cosmology, where $\Omega_b h^2 = 0.0224$ and $\Omega_c h^2 = 0.12$. *Right*: The region of the left panel in the red dotted box enlarged.

three fitting formulas. The right panel zooms in on $T_{\mathrm{GA}}(k)$ and $T_{\mathrm{EH}}(k)$ in the transition regime, emphasizing the fact that $T_{\mathrm{GA}}(k)$ and $T_{\mathrm{EH}}(k)$ are accurate to a similar order of magnitude, both mostly within 4% error.

To get a full picture of the performance of the two approximations, however, it is important to check their average errors over a range of possible cosmologies. I define the average error as

$$E_{\mathrm{avg}} = \frac{1}{n_k} \sum_{i=1}^{n_k} \frac{|T_{\mathrm{fit}}(k_i) - T_{\mathrm{CAMB}}(k_i)|}{T_{\mathrm{CAMB}}(k_i)}, \tag{3.30}$$

and compute this value for 100 cosmologies with $\Omega_c h^2 \in [0.10, 0.14]$ and $\Omega_b h^2 \in [0.012, 0.03]$. Figure 3.7 shows the average percent error against $T_{\mathrm{CAMB}}(k)$ for the $T_{\mathrm{EH}}(k)$ and $T_{\mathrm{GA}}(k)$, on the left and right respectively. Firstly, in the selected range both approximations have an average percent error below 3%, making them both an order of magnitude better than $T_{\mathrm{BBKS}}(k)$. The sharp band structure for the GA approximation results from the fact that the $T_{\mathrm{GA}}(k)$ can better account for $\Omega_c h^2$, which dominates the matter density and is most directly responsible for the general shape of the transfer function, whereas baryons induce secondary effects (e.g. the baryon acoustic oscillations in the transition regime) which the simple formula cannot account for. While $T_{\mathrm{GA}}(k)$ is better than $T_{\mathrm{EH}}(k)$ in some ranges, its approximating power is not as robust

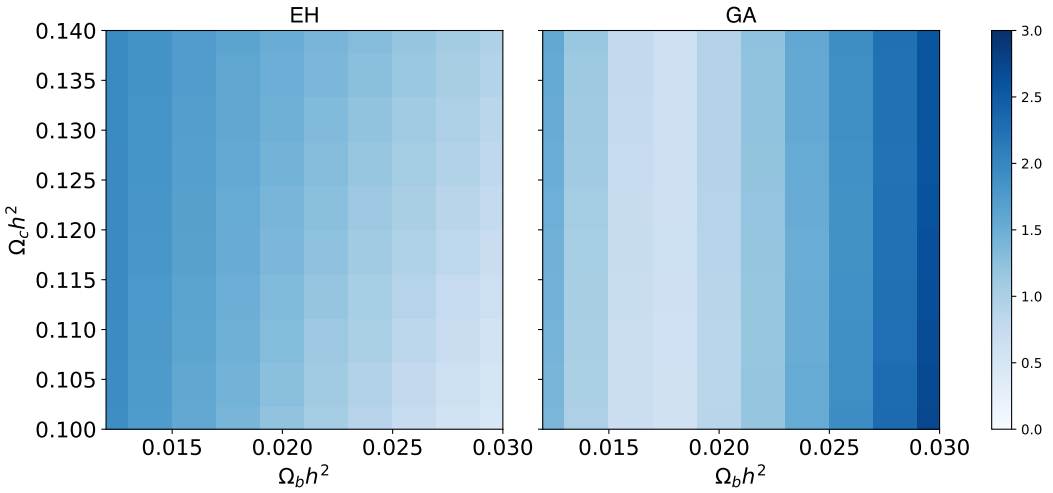


Figure 3.7: *Left*: The average percent error of the EH transfer function for 100 cosmologies, with $\Omega_c h^2 \in [0.10, 0.14]$ and $\Omega_b h^2 \in [0.012, 0.03]$, and $H_0 = 67.4$. *Right*: The average percent error of the GA learned transfer function for the same cosmologies.

as $T_{\rm EH}(k)$, as evident by the smoothly varying average error. However, it is important to recall the simplicity of $T_{\rm GA}(k)$ compared to $T_{\rm EH}(k)$. Using GA to reduce the complexity of the transfer function only marginally affects its accuracy.

In the above section I discusses the matter transfer function, it's importance as a cosmological probe, and described the BBKS and EH analytic formulas for the transfer function. Then, I utilized genetic algorithms as described in the previous section to reproduce the results of [3]. I created a GA-learned fitting formula for the transfer function that is both as simple as the BBKS formula, yet as accurate as the EH formula, demonstrating the power of GA. In the next section, I will use Symbolic Regression to learn a fitting formula for the lensing tail of the CMB power spectrum.

3.2 Part 2: The CMB Lensing Tail

Next generation experiments will allow us to measure the CMB power spectrum to unprecedented levels of precision and accuracy. Increased resolution at small scales, or high ℓs, will allow researchers to probe the effects of weak gravitational lensing on the CMB photons to higher precision than previously

possible, allowing for more stringent constraints on cosmological models. Similar to how the effects of diffusion damping and reionization on the photon distribution as it streams from last scattering to today can be modeled as damping tails on CMB power spectrum [7], I explore the possibility of modeling the so called CMB lensing tail using Symbolic Regression. I qualitatively describe the effects of weak gravitational lensing on the CMB. I then construct a simple fitting function format to guess the shape of the lensing tail, and utilize Symbolic Regression via PySR to learn a fit to the lensing tail. This preliminary investigation will serve to show the strength of Symbolic Regression in creating simple, yet powerful analytic formulas.

3.2.1 The Lensing Tail

Weak gravitational lensing is the phenomenon whereby the path of distant photons traveling to us is curved in the presence of an intervening mass distribution. That general relativity implies the curving of photon trajectories in the presence of mass distributions is evident by the work done on the Boltzmann equation in Chapter 1. Weak gravitational lensing is the particular focus on how photon trajectories from distant sources are curved due to the presence of overdensities in the line of sight between us and the source. Cosmic shear experiments[16] obtain cosmological constraints by measuring the distortion of the shapes of distant galaxies as they appear to us due to weak lensing. Measuring lensing via the CMB provides information at larger scales and higher redshifts than other experiments, given that the CMB is the oldest signal we can directly observe. Measuring lensing in the CMB then offers the potential to break parameter degeneracies and improve constraints on quantities that affect the geometry or density perturbations at late times [17].

The effect of weak lensing on the CMB is an ever so subtle distortion on larger scales. It serves to smooth out the CMB power spectrum. On small scales weak lensing increases the power of the CMB signal, by upto 10% past $\ell > 2000$ [17]. Let the lensing tail be described by the function $L(\ell)$ so that

$$C_\ell^{\text{lensed}} = L(\ell) \times C_\ell^{\text{unlensed}}, \tag{3.31}$$

where C_ℓ^{lensed} and C_ℓ^{unlensed} are the lensed and unlensed CMB power spectra, respectively. As in the previous section, I used CAMB to generate both the lensed and unlensed CMB power spectra, and divide them to obtain $L_{\text{CAMB}}(\ell)$, so that

$$L_{\text{CAMB}}(\ell) = \frac{C_\ell^{\text{CAMB,lensed}}}{C_\ell^{\text{CAMB,unlensed}}}. \tag{3.32}$$

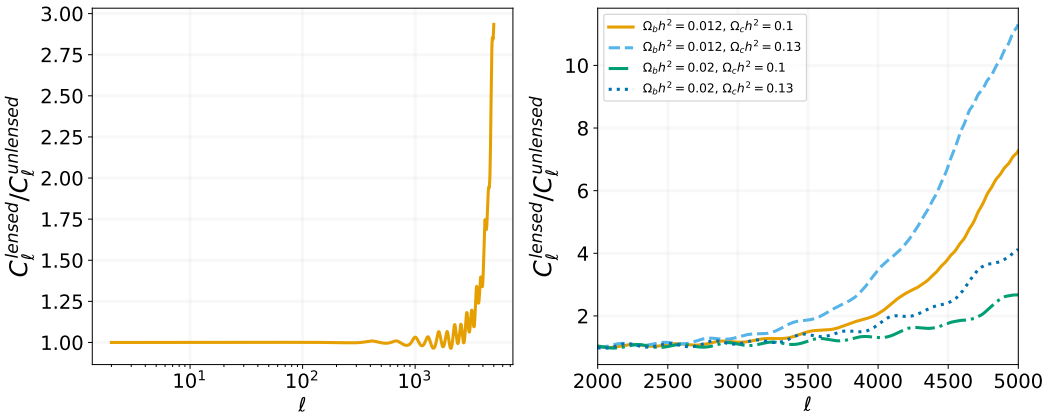


Figure 3.8: *Left*: $L_{\text{CAMB}}(\ell)$ for the Planck best fit cosmology, where $\Omega_b h^2 = 0.0224$ and $\Omega_c h^2 = 0.12$. *Right*: $L_{\text{CAMB}}(\ell)$ for four sample cosmologies for $\ell > 2000$.

The left panel in Figure 3.8 shows $L_{\text{CAMB}}(\ell)$ for the Planck best fit cosmology, and the right panel shows the lensing tail for some sample cosmologies. As mentioned above, the lensing tail increases the power of the unlensed spectrum for $\ell > 2000$.

3.2.2 The Function Template

In this preliminary investigation, I focus primarily on modeling the general shape of the lensing tail using Symbolic Regression. I take the template function to take the form of a scaled polynomial which is "activated" by a sigmoid function after a specific value of ℓ. The template function is of the form

$$L_{\text{SR}}(\ell) = 1 + \omega(\ell)\left[\beta_1 \left(\frac{\ell}{\beta_2}\right)^{\beta_{\text{SR}}\left(\Omega_b h^2,\ \Omega_m h^2\right)} - 1\right], \tag{3.33}$$

where

$$\omega(\ell) = \left[1 + e^{-(\ell - \beta_3 \times 2000)/100}\right]^{-1} \tag{3.34}$$

is the sigmoid function, centered at $\beta_3 \times 2000$ with a width of 100. Inclusion of the sigmoid is motivated by the fact that the lensing tail is expected to start deviating significantly from 1 for $\ell > 2000$. The β_i are free parameters which PySR will optimize in its search. The function $\beta_{\text{SR}}\left(\Omega_b h^2,\ \Omega_m h^2\right)$ is to be learned by PySR, as a function of $\Omega_b h^2$ and $\Omega_m h^2$.

The parameters β_1 and β_2 are included to allow for tuning of the polynomial scaling. The parameter β_3 is so that PySR may learn a potentially better "activation" point for the sigmoid function.

3.2.3 Symbolic Regression

Dataset

In PySR, it is possible to fix a template function, as in eq. (3.33), so that PySR can optimize the parameters β_i as well as learn the form of the function β_{SR}. I use CAMB to generate $L_{\text{CAMB}}(\ell)$ for 100 cosmologies with parameters selected uniformly randomly from the ranges

$$\begin{aligned} H_0 &\in [45,\ 95] \\ \Omega_b h^2 &\in [0.015,\ 0.033] \\ \Omega_c h^2 &\in [0.03,\ 0.18] \end{aligned}$$

for $\ell = [1000,\ \ldots,\ 5000]$. I disregard the large scale values so that the model may better focus on learning the region of interest, namely small scales or high ℓ. Due to how PySR is implemented, the input data points must be transformed from the form

$$d_i = \left\{\Omega_b^i h^2,\ \Omega_c^i h^2,\ H_0^i,\ \vec{\ell},\ \vec{L}^i_{\text{CAMB}}\right\}$$

to

$$d_j^i = \left\{\Omega_b^i h^2,\ \Omega_m^i h^2,\ H_0^i,\ \ell_j,\ L^i_{\text{CAMB,j}}\right\}, \tag{3.35}$$

where $L^i_{\text{CAMB,j}} \equiv L^i_{\text{CAMB}}(\ell_j)$ is the lensing tail value of the ℓ_j for the i^{th} cosmology. Each ℓ_j and corresponding $L^i_{\text{CAMB}}(\ell_j)$ for each cosmology is its own data point, as opposed to the transfer function dataset in the last section where each data point contained the entire $T(k)$ for each cosmology. The total size of the input dataset now becomes $n_{\text{cosmologies}} \times n_\ell = 400000$. PySR is not well equipped to calculate the error for that many data points at once, so it calculates the error on subsets $B(D)$ of the dataset D at a time, a procedure known as batching in machine learning. For each fitness computation, it calculates the error for a randomly sampled subset of the dataset of size n_b, known as a batch.

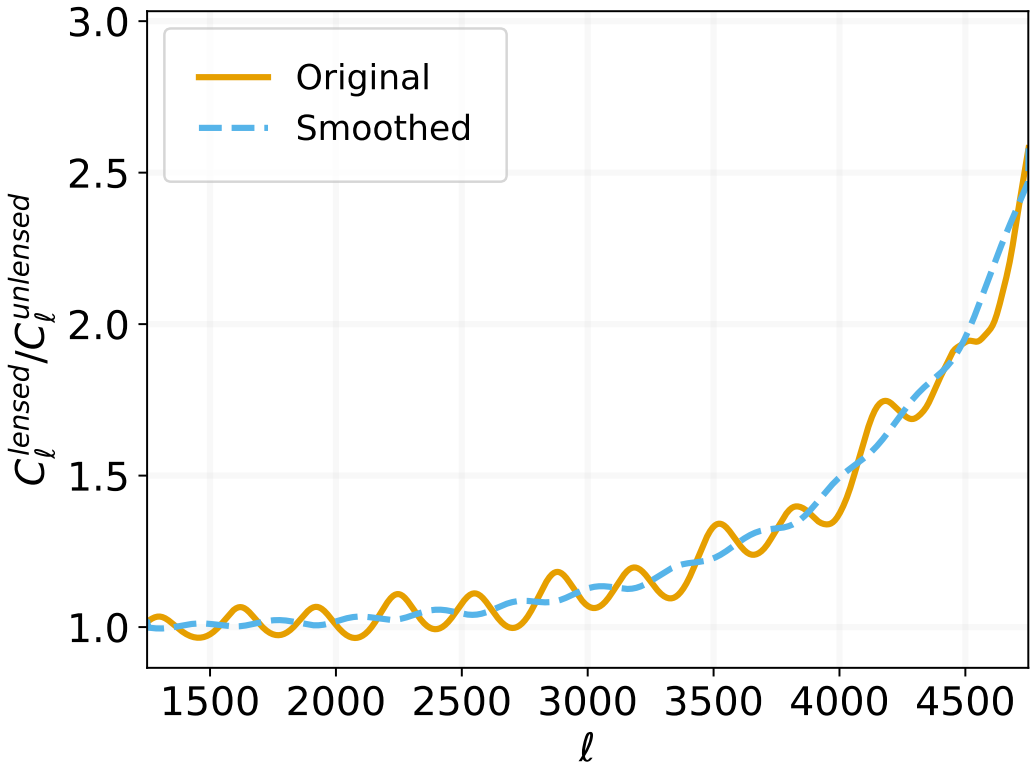


Figure 3.9: $L_{\text{CAMB}}(\ell)$ for the Planck best fit cosmology, where $\Omega_b h^2 = 0.0224$ and $\Omega_c h^2 = 0.12$ in yellow. The same curved smoothed according to eq. (3.37) is shown in blue.

I leave the error function as the PySR default, which is the mean-squared error in eq. (2.5). Namely, the error is computed as

$$E = \frac{1}{n_b} \sum_{i,j \in B(D)} \left(L_{\text{SR}}(\ell_j;\ \Omega_b^i h^2,\ \Omega_m^i h^2) - L^i_{\text{CAMB,j}} \right)^2 \tag{3.36}$$

for each randomly selected batch $B(D)$. In principal, $B(D)$ for each computation may contain randomly sampled ℓ values from any or all cosmologies in the dataset.

Smoothing the Data

As is evident in fig. 3.8, the lensing tail has oscillatory behavior. My purpose here is to learn a fit for the general shape of the lensing tail, so the oscillatory behavior may slow down the convergence of the model. In order that the

model may better learn the general shape, I apply a smoothing filter to the dataset for each cosmology. For a cosmology i, I first use CAMB to generate

$$\vec{L}^i_{\text{CAMB}} = [L^i_{\text{CAMB},1000},\ \ldots,\ L^i_{\text{CAMB},5000}]$$
$$\text{for } \vec{\ell} = [1000,\ \ldots,\ 5000].$$

The $L^i_{\text{CAMB,j}}$ data points are transformed according to

$$L^i_{\text{CAMB,j}} \to \frac{1}{500} \sum_{k=-250}^{249} L^i_{\text{CAMB},j+k}, \tag{3.37}$$

for $j = 1250,\ \ldots,\ 4750$. This has the effect of averaging the points in the oscillatory region, so that each point is replaced by the average of $L^i(\ell_j)$ in a window of 500 points around it. After this smoothing is performed, the data points are expanded to the form in eq. (3.35). Figure 3.9 shows the effect of the smoothing filter.

PySR Settings

In PySR, there is the option to tune a large number of hyperparameters in order to aid in model convergence. This is the set of configurations I found to work best. Firstly, I limit the grammar to the operations addition, multiplication, subtraction, and exponentiation, so that the final expression is not overly complex.

Hyperparameter	Symbol	Value
Number of Independent Populations	n_p	20
Number of Generations	n_g	200
Expressions per Population	n_e	1000
Batch Size	n_b	10000

All other hyperparameters I leave at the PySR default values.

3.2.4 Results

After running the symbolic regression, PySR converged to the following expression:

$$L_{\text{SR}}(\ell) = 1 + \omega(\ell) \left[0.83 \times \left(\frac{\ell}{3089} \right)^{\beta_{\text{SR}}\left(\Omega_b h^2,\ \Omega_m h^2\right)} - 1 \right], \tag{3.38}$$

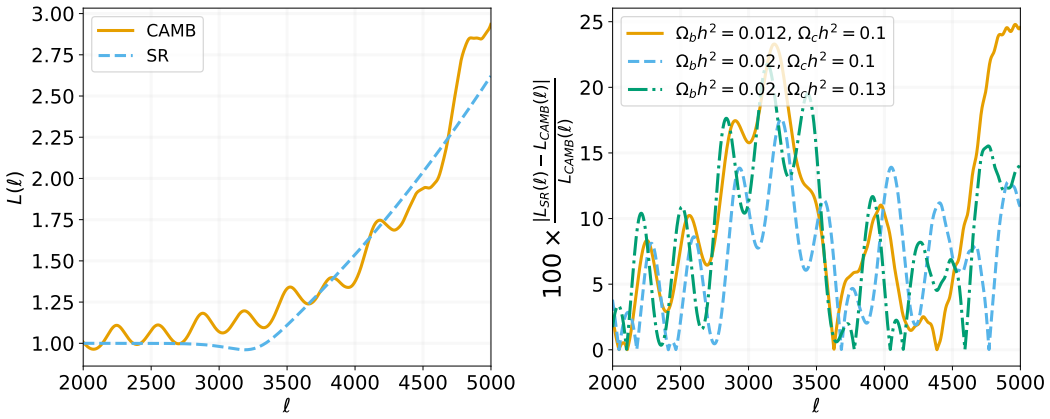


Figure 3.10: *Right*: $L_{\text{CAMB}}(\ell)$ for the Planck best fit cosmology, where $\Omega_b h^2 = 0.0224$ and $\Omega_c h^2 = 0.12$, in yellow, with the SR approximated lensing tail $L_{\text{SR}}(\ell)$ in blue. The percent error between $L_{SR(\ell)}$ and $L_{\text{CAMB}}(\ell)$ for three cosmologies.

where

$$\omega(\ell) = \left[1 + e^{-(\ell - 3240)/100}\right]^{-1} \tag{3.39}$$

and

$$\beta_{\text{SR}}\left(\Omega_b h^2,\ \Omega_m h^2\right) = \Omega_m h^2 (\Omega_b h^2)^{-0.877} - 3.342 \Omega_m h^2 - 1.118. \tag{3.40}$$

The left panel in figure 3.10 shows $L_{\text{CAMB}}(\ell)$ for the Planck best fit cosmology, as in fig. 3.8, along with the $L_{\text{SR}}(\ell)$ approximation. The right panel shows the percent error between $L_{\text{SR}}(\ell)$ and $L_{\text{CAMB}}(\ell)$ for some sample cosmologies. Calculating the average error for the $\ell > 2000$ region, as

$$E_{\text{avg}} = \frac{1}{3000} \sum_{\ell=2000}^{5000} \frac{|L_{\text{SR}}(\ell) - L_{\text{CAMB}}(\ell)|}{L_{\text{CAMB}}(\ell)}, \tag{3.41}$$

$L_{\text{SR}}(\ell)$ has an average error under 15%, as shown in fig. 3.11, with an error as low as 5% for some cosmologies. This initial lensing tail fit is about as accurate, on average, as the BBKS fit is for the transfer function. From fig. 3.10, it's clear that a large chunk of the error in $L_{\text{SR}}(\ell)$ comes around the transition point of the sigmoid, around $\ell = 3240$. A better modeling of this transition may be able to improve the accuracy of the fit. Furthermore, a logical next step would be to add a factor to account for the oscillatory behavior[4].

[4]The authors of [18] do something similar to build on their initial GA transfer function in [3], the work I follow in the last section.

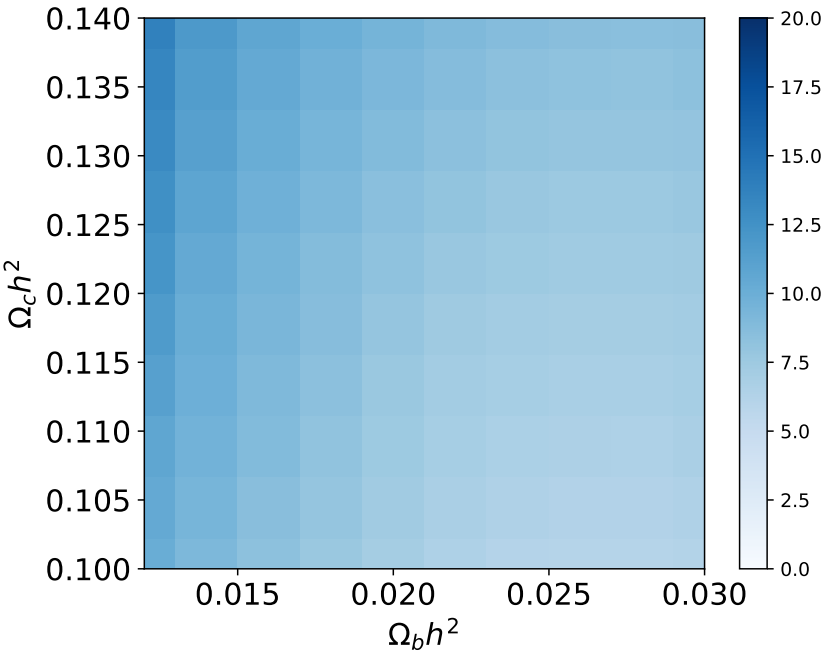


Figure 3.11: The average percent error of $L_{\mathrm{SR}}(\ell)$ for 100 cosmologies, with $\Omega_c h^2 \in [0.10, 0.14]$ and $\Omega_b h^2 \in [0.012, 0.03]$, and $H_0 = 67.4$.

Chapter 4

Conclusion

In this work, I've explored various analytic methodologies involved in modeling key cosmological quantities. In Chapter One, I explore the semi-analytic formulation of the Cosmic Microwave Background (CMB) angular power spectrum, a key observable, as presented by Hu and Sugiyama in [5]. Their work is an example of the strong insight one can gain into cosmological phenomena when they are approached analytically. Though not strictly a replacement to quick and accurate numerical codes, analytic formulations offer to enrich our understanding of the underlying physics, and are important tools to have.

In Chapter Two, I introduce machine learning and in particular Genetic Algorithms (GAs) and Symbolic Regression (SR). I describe the algorithm behind a popular SR package, PySR[4]. Packages like PySR, when used in conjunction with physical motivation, can be utilized to discover simple and powerful approximating expressions.

In Chapter Three, I explore how the tools described in Chapter Two can be leveraged to create simple, yet powerful, approximate analytic expressions for cosmological quantities of interest. I begin by briefly describing the matter transfer function and its strength as an observable. I then mention two existing analytic formulae, one by Bardeen, Bond, Kaiser, and Szalay from [14], and another by Eisenstein and Hu in [15]. Following [3], I implement a GA to create an approximate matter transfer function, which is similar to the BBKS formula in its simplicity, yet as accurate as the EH formula.

Finally in the last section of Chapter Three, I explore the use of SR towards approximating the nontrivial CMB lensing tail. In my preliminary investigation, I find SR to be quite successful, given how few details I supply the model

with. SR can be further utilized to create a more accurate fit, for example to account for the oscillatory behavior in the lensing tail. It remains to be shown how incorporating more physically motivated terms in the template function may lead to a stronger lensing tail fit.

Bibliography

[1] ESA and the Planck Collaboration. Planck cmb, 2013. URL `https://www.esa.int/ESA_Multimedia/Images/2013/03/Planck_CMB`. [Online; accessed April 03, 2025].

[2] N. Aghanim, Y. Akrami, F. Arroja, 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, B. Casaponsa, A. Challinor, H. C. Chiang, L. P. L. Colombo, C. Combet, D. Contreras, B. P. Crill, F. Cuttaia, P. de Bernardis, G. de Zotti, J. Delabrouille, J.-M. Delouis, F.-X. Désert, E. Di Valentino, C. Dickinson, J. M. Diego, S. Donzelli, O. Doré, M. Douspis, A. Ducout, X. Dupac, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, E. Falgarone, Y. Fantaye, J. Fergusson, R. Fernandez-Cobos, F. Finelli, F. Forastieri, M. Frailis, 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, G. Helou, 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, M. Langer, A. Lasenby, M. Lattanzi, C. R. Lawrence, M. Le Jeune, J. P. Leahy, 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. D. Meerburg, P. R. Meinhold, A. Melchiorri, A. Mennella, M. Migliaccio, M. Millea, S. Mitra, M.-A. Miville-Deschênes, D. Molinari, A. Moneti, L. Montier, G. Morgante, A. Moss, S. Mottet, M. Münchmeyer, P. Natoli, H. U. Nørgaard-Nielsen, C. A. Oxborrow, L. Pagano, D. Paoletti, B. Partridge, G. Patan-

chon, T. J. Pearson, M. Peel, H. V. Peiris, F. Perrotta, V. Pettorino, F. Piacentini, L. Polastri, G. Polenta, J.-L. Puget, J. P. Rachen, M. Reinecke, M. Remazeilles, C. Renault, 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, M. Shiraishi, C. Sirignano, G. Sirri, L. D. Spencer, R. Sunyaev, A.-S. Suur-Uski, J. A. Tauber, D. Tavagnacco, M. Tenti, L. Terenzi, L. Toffolatti, M. Tomasi, T. Trombetti, 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. Planck2018 results: I. overview and the cosmological legacy ofplanck. *Astronomy amp; Astrophysics*, 641:A1, September 2020. ISSN 1432-0746. doi: 10.1051/0004-6361/201833880. URL http://dx.doi.org/10.1051/0004-6361/201833880.

[3] J. Bayron Orjuela-Quintana, Savvas Nesseris, and Wilmar Cardona. Using machine learning to compress the matter transfer function. *Physical Review D*, 107(8), April 2023. ISSN 2470-0029. doi: 10.1103/physrevd.107.083520. URL http://dx.doi.org/10.1103/PhysRevD.107.083520.

[4] Miles Cranmer. Interpretable Machine Learning for Science with PySR and SymbolicRegression.jl, May 2023. URL http://arxiv.org/abs/2305.01582. arXiv:2305.01582 [astro-ph, physics:physics].

[5] Wayne Hu and Naoshi Sugiyama. Anisotropies in the cosmic microwave background: an analytic approach. *The Astrophysical Journal*, 444:489, May 1995. ISSN 1538-4357. doi: 10.1086/175624. URL http://dx.doi.org/10.1086/175624.

[6] A A Penzias and R W Wilson. A measurement of excess antenna temperature at 4080 MC/s. *Astrophys. J.*, 142:419, July 1965.

[7] Wayne Hu and Martin White. The damping tail of cosmic microwave background anisotropies. *The Astrophysical Journal*, 479(2):568–579, April 1997. ISSN 1538-4357. doi: 10.1086/303928. URL http://dx.doi.org/10.1086/303928.

[8] D. J. Fixsen. The temperature of the cosmic microwave background. *The Astrophysical Journal*, 707(2):916–920, November 2009. ISSN 1538-4357. doi: 10.1088/0004-637x/707/2/916. URL http://dx.doi.org/10.1088/0004-637X/707/2/916.

[9] Scott Dodelson. *Modern Cosmology*. 2003.

[10] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, 1 edition, 2007. ISBN 0387310738.

[11] John R. Koza. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. Bradford Books, 1992.

[12] Alan H. Guth. Inflationary universe: A possible solution to the horizon and flatness problems. , 23(2):347–356, January 1981. doi: 10.1103/PhysRevD.23.347.

[13] Antony Lewis and Anthony Challinor. CAMB: Code for Anisotropies in the Microwave Background. Astrophysics Source Code Library, record ascl:1102.026, February 2011.

[14] J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay. The Statistics of Peaks of Gaussian Random Fields. , 304:15, May 1986. doi: 10.1086/164143.

[15] Daniel J. Eisenstein and Wayne Hu. Baryonic features in the matter transfer function. *The Astrophysical Journal*, 496(2):605–614, April 1998. ISSN 1538-4357. doi: 10.1086/305424. URL `http://dx.doi.org/10.1086/305424`.

[16] Marika Asgari, Chieh-An Lin, Benjamin Joachimi, Benjamin Giblin, Catherine Heymans, Hendrik Hildebrandt, Arun Kannawadi, Benjamin Stölzner, Tilman Tröster, Jan Luca van den Busch, Angus H. Wright, Maciej Bilicki, Chris Blake, Jelte de Jong, Andrej Dvornik, Thomas Erben, Fedor Getman, Henk Hoekstra, Fabian Köhlinger, Konrad Kuijken, Lance Miller, Mario Radovich, Peter Schneider, HuanYuan Shan, and Edwin Valentijn. Kids-1000 cosmology: Cosmic shear constraints and comparison between two point statistics. *Astronomy amp; Astrophysics*, 645:A104, January 2021. ISSN 1432-0746. doi: 10.1051/0004-6361/202039070. URL `http://dx.doi.org/10.1051/0004-6361/202039070`.

[17] Duncan Hanson, Anthony Challinor, and Antony Lewis. Weak lensing of the cmb. *General Relativity and Gravitation*, 42(9):2197–2218, June 2010. ISSN 1572-9532. doi: 10.1007/s10714-010-1036-y. URL `http://dx.doi.org/10.1007/s10714-010-1036-y`.

[18] J. Bayron Orjuela-Quintana, Domenico Sapone, and Savvas Nesseris. Analytical emulator for the linear matter power spectrum from physics-informed machine learning, 2024. URL `https://arxiv.org/abs/2407.16640`.