

Constraining primordial perturbations on small scales

A Thesis

submitted to

Indian Institute of Science Education and Research Pune

in partial fulfillment of the requirements for the

BS-MS Dual Degree Programme

by

Neel Shah



Indian Institute of Science Education and Research Pune

Dr. Homi Bhabha Road,
Pashan, Pune 411008, INDIA.

April, 2023

Supervisor: L Sriramkumar

© Neel Shah 2023

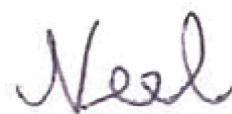
All rights reserved

Certificate

This is to certify that this dissertation entitled 'Constraining primordial perturbations on small scales' towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study/work carried out by Neel Shah at the Indian Institute of Technology Madras under the supervision of L Sriramkumar, Professor, Department of Physics, during the academic year 2022-2023.



L Sriramkumar



Neel Shah

Committee:

Prof. L Sriramkumar

Dr. Arka Banerjee

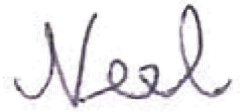
This thesis is dedicated to my family.

Declaration

I hereby declare that the matter embodied in the report entitled Constraining primordial perturbations on small scales are the results of the work carried out by me at the Department of Physics, Indian Institute of Science Education and Research, Pune, under the supervision of L Sriramkumar and the same has not been submitted elsewhere for any other degree.



L Sriramkumar



Neel Shah

Acknowledgments

I would first like to convey my gratitude to my supervisor Prof. L. Sriramkumar for leading by example and thereby helping me learn many things in physics and beyond physics. I am also grateful to him for strongly believing in my capabilities, often more than I myself did. I am thankful to my other collaborator, Prof. Shiv Sethi for his illuminating lectures and our many stimulating discussions on cosmology. I would also like to thank all members of Prof. Sriramkumar's group for providing me with a very friendly and relaxing work environment.

I would like to thank my previous mentors: Dr. Diptimoy Ghosh for refining my sense of scientific rigor and exposing me to many interesting areas in cosmology, Dr. Torsten Ensslin and Dr. Julia Stadler for giving me a very enriching and memorable experience with my first computational project, and Prof. T.S. Mahesh for giving me my first experience in being a part of a research group. I would like to thank all my instructors and other faculty at IISER Pune for exposing me to a wide variety of ways of thinking about science, thereby helping me to determine the kind of scientist I want to be.

I am grateful to all my wonderful friends at IISER Pune. I am not sure if I will ever find such a community again, where my peers are like-minded enough to make me feel like I belong, but different enough that I have learnt something from each and every one of them. I am indebted to my friends who stayed in touch with me during the days of the pandemic and now during my thesis project away from campus, as these people have prevented me from being alone in my joys and sufferings in my academic journey.

Last but not the least, I am grateful to my family for their unconditional love for me, for allowing me to define my own goals, for not leaving my side at any stage of my life, and for reminding me that I matter irrespective of my successes and failures.

Abstract

The power spectrum of primordial curvature perturbations of our universe is very well measured on the largest observable length scales by cosmic microwave background (CMB) anisotropies. But on smaller scales, observational constraints on it are much weaker. In this thesis, we study various probes of primordial perturbations on small scales, namely CMB spectral distortions, primordial black holes, and secondary gravitational waves. We use these probes to constrain parametric forms of the primordial power spectrum. We investigate the dependence of these constraints on the shape of the spectrum. We then show how the same methodologies can be used to constrain an underlying model of inflation.

Contents

Abstract	xi
1 Preliminaries	1
1.1 Basics of FLRW cosmology	2
1.2 Implementing inflation with a single scalar field	4
1.3 The evolution of primordial perturbations	8
1.4 Quantizing primordial perturbations	11
1.5 Primordial power spectra	14
1.6 Ultra slow roll	19
1.7 Aims of this thesis	20
2 Probes of primordial perturbations on small scales	23
2.1 Introduction	23
2.2 CMB spectral distortions	24
2.3 Primordial black holes from inflation	28
2.4 Secondary gravitational waves	32
3 Deriving small scale constraints on the power spectrum	39
3.1 Constraining parametric power spectra	39
3.2 Constraining an inflationary potential	46
4 Results and Discussion	49
4.1 Constraining a lognormal power spectrum	49
4.2 Constraining a broken power law power spectrum	51
4.3 Constraining the USR-1 potential	53
5 Conclusions and outlook	55
A A simple proof of the decomposition theorem	57

CONTENTS

A.1	Defining the metric and matter perturbations	57
A.2	Gauge transformations and gauge invariant quantities	60
A.3	Proving the decomposition theorem	65
A.4	Generalizing to curved FLRW universes	73
	Bibliography	79

List of Figures

1.1	The USR-1 potential	20
1.2	The USR-1 power spectrum	21
2.1	The scales probed by various cosmological observables	24
2.2	Window function for the μ distortion	27
2.3	The integration kernel for SGWs	35
3.1	Constraints from f_{PBH} bounds for different δ_c	44
4.1	Constraints on the lognormal power spectrum	50
4.2	Constraints on the broken power law spectrum. The values 0.5 and 4 for α and β are chosen to conform with the values chosen in [22].	52
4.3	Constraints on the USR-1 model	53
4.4	Constraints on the USR-1 model	54

LIST OF FIGURES

Chapter 1

Preliminaries

Cosmology as we know it today started after the discovery of general relativity by Einstein, and the development of the simplest cosmological model by Friedmann, Lemaitre, Robertson, and Walker. After several decades of observations of galaxies, type IA supernovae, and the cosmic microwave background (CMB), we have a very good, albeit incomplete, picture of what the universe looks like at present, and what it looked like for the majority of the past 13.8 billion years.

Some of the most profound statements about our universe are statistical in nature. The best example of this is the Cosmological Principle, which states that if we look at our universe upon coarse-graining it over smaller scales, we find it to be homogeneous and isotropic on the largest observable length scales, meaning that there are no special locations, and there are no preferred directions from any vantage point. The validity of the cosmological principle hugely simplifies theoretical models of our universe, as it allows us to assume the simplest kind of background cosmology which is the Friedmann-Lemaitre-Robertson-Walker (FLRW) universe.

Another statistical piece of information with significant implications for our universe is the primordial power spectrum. This is a function of the length scale which quantifies the degree of curvature perturbations at each length scale that our universe started with. The power spectrum has been measured accurately at the largest observable length scales from the CMB observations by the Planck satellite [1]. The most popular mechanism for the physical origin of the power spectrum is inflation: an era when the very early universe expanded very quickly. The focus of this thesis

is on observational imprints of the power spectrum on smaller scales and on current and future observational limits on it.

In the following sections, we lay a basic theoretical foundation for describing the evolution of the FLRW universe. We will introduce the mechanism of inflation and how it determines the primordial power spectrum. We will then review the observational status of the primordial power spectrum.

1.1 Basics of FLRW cosmology

We have collected below some fundamental concepts and equations in FLRW cosmology that will be used throughout this article.

1. The line element for the FLRW metric in polar coordinates is

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu = dt^2 - a^2(t) \left(\frac{dr^2}{1 - kr^2} + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2 \right). \quad (1.1)$$

2. We often specialize to the spatially flat FLRW metric, which has $k = 0$. When we do linear perturbation theory around this metric it is convenient to use Cartesian coordinates. The line element for the flat FLRW metric in Cartesian coordinates is

$$ds^2 = dt^2 - a^2(t)(dx^2 + dy^2 + dz^2). \quad (1.2)$$

3. The Einstein tensor for the flat FLRW metric is given by

$$G_0^0 = 3H^2, \quad (1.3a)$$

$$G_i^0 = 0, \quad (1.3b)$$

$$G_0^i = 0, \quad (1.3c)$$

$$G_j^i = (2\dot{H} + 3H^2)\delta_j^i. \quad (1.3d)$$

4. The stress energy tensor corresponding to the FLRW metric can be written as that of a perfect

fluid, and is given by

$$T_{\mu\nu} = (\rho + p)u_\mu u_\nu - pg_{\mu\nu} , \quad (1.4)$$

where ρ is the energy density, p is the pressure, u_α is the 4-velocity.

5. Since we are using the (+, -, -, -) metric signature, the normalization of the 4-velocity is

$$u^\mu u_\mu = 1. \quad (1.5)$$

This follows from the fact that the 4-velocity of an observer is the timelike basis vector in their comoving frame.

6. Since the metric and stress energy tensor are both isotropic, the perfect fluid is at rest in the comoving coordinates, thus

$$u_i = 0 = u^i . \quad (1.6)$$

From the normalization of the 4-velocity eq. (1.5) we get

$$u_0 = \frac{1}{\sqrt{g^{00}}} . \quad (1.7)$$

Thus, for the FLRW metric eq. (1.1), we get

$$u^\mu = (1, 0, 0, 0) = u_\mu . \quad (1.8)$$

7. The Einstein equations for the FLRW metric reduce to two equations- the first Friedmann equation which is the 00 component, and its time derivative, which is called the second Friedmann equation.

$$H^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} , \quad (1.9a)$$

$$\dot{H} + H^2 = -\frac{4\pi G}{3}(\rho + 3p) . \quad (1.9b)$$

1.2 Implementing inflation with a single scalar field

The two biggest motivations for inflation are that it can generate primordial perturbations consistent with what we see in the CMB, and that it can expand the universe by such an amount that the largest observable scale today went from being smaller than the Hubble radius H^{-1} to being larger than it at the time of the CMB. The latter motivation is in order to explain why the entire CMB is observed to have nearly the same temperature; it indicates that even the most distant parts of the CMB must have been in contact in the past. Thus, what we need from a model of inflation is that it:

1. Ensures that the largest observable scale today is inside the Hubble radius at some time in the past. This is often measured by the number of e-folds by which the scale factor expands between the Hubble radius exit of a reference scale k_{pivot} and the end of inflation, given by

$$N_{\text{pivot}} = \ln \left(\frac{a_f}{a_*} \right) . \quad (1.10)$$

2. Generates a spectrum of primordial perturbations consistent with what is observed from the CMB.

The simplest mechanism of implementing this is having a single canonical scalar field be the dominant constituent of the universe. The scalar field can drive a sufficient amount of inflation for suitable forms of its potential, and suitable values of the potential parameters and initial field. In this case it is called the inflaton. The action of a canonical scalar field ϕ is

$$S[\phi(x), \partial_\mu \phi(x)] = \int d^4x \sqrt{-g} \left[\frac{1}{2} \partial^\mu \phi \partial_\mu \phi - V(\phi) \right] . \quad (1.11)$$

For the FLRW metric this leads to the equation of motion

$$\ddot{\phi} + 3H\dot{\phi} + V'(\phi) = 0 , \quad (1.12)$$

where $V'(\phi)$ denotes $dV/d\phi$. Note that we have neglected the spatial gradient term in the action and equation of motion as we are working in a homogeneous universe.

The stress energy tensor of the field is given by

$$T^{\mu\nu} = \frac{2}{\sqrt{-g}} \frac{\delta S}{\delta g_{\mu\nu}} . \quad (1.13)$$

The energy density (kinetic + potential) of the scalar field is

$$\rho = \frac{1}{2} \dot{\phi}^2 + V(\phi) , \quad (1.14)$$

and the pressure is

$$p = \frac{1}{2} \dot{\phi}^2 - V(\phi) . \quad (1.15)$$

Further, when $\partial^\mu \phi \partial_\mu \phi > 0$, if we define the 4-velocity of the field as

$$u^\alpha = \frac{\partial^\alpha \phi}{\sqrt{\partial^\mu \phi \partial_\mu \phi}} , \quad (1.16)$$

then we can write the stress energy tensor of a scalar field as that of a perfect fluid eq. (1.4), with the fluid variables given by eq. (1.14)-eq. (1.16). This tells us that when $\partial^\mu \phi \partial_\mu \phi > 0$, a scalar field behaves exactly as a perfect fluid.

With the energy density coming from the scalar field, the first Friedmann equation is

$$H^2 = \frac{1}{3M_{\text{Pl}}^2} \left[\frac{1}{2} \dot{\phi}^2 + V(\phi) \right] . \quad (1.17)$$

1.2.1 The slow roll approximation and the slow roll parameters

Given just the potential $V(\phi)$ and the initial conditions ϕ_i and $\dot{\phi}_i$, it is in principle possible but very tedious to see whether they lead to sufficient e-folds of inflation. A better alternative is to characterise inflationary dynamics by introducing a 'hierarchy' of parameters called slow roll parameters, defined

by

$$\epsilon_0 = \frac{H_i}{H}, \quad (1.18)$$

$$\epsilon_{n+1} = \frac{1}{\epsilon_n} \frac{d\epsilon_n}{dn}, \quad n \geq 0, \quad (1.19)$$

where H_i is the initial value of the Hubble parameter, and N is the number of e-folds of inflation after $t = t_i$.

When a canonical scalar field is the dominant component, we have

$$\epsilon_1 = \frac{3\dot{\phi}^2}{2\rho}, \quad (1.20)$$

$$\epsilon_2 = \frac{1}{H} \left(2\frac{\ddot{\phi}}{\dot{\phi}} - \frac{\dot{\rho}}{\rho} \right). \quad (1.21)$$

The slow roll parameters do not contain any new information that the field initial conditions and equation of motion do not have. They are just a different parametrization of the field's dynamics. We can show that an accelerating expansion is equivalent to the condition $\epsilon_1 < 1$:

$$\dot{H} = \frac{d}{dt} \left(\frac{\dot{a}}{a} \right) = \frac{\ddot{a}}{a} - H^2. \quad (1.22)$$

Thus, $\ddot{a} > 0$ if and only if

$$\epsilon_1 = -\frac{\dot{H}}{H^2} < 1. \quad (1.23)$$

Also, $\epsilon_1 \ll 1$ implies that the kinetic energy of the field is negligible in comparison to the potential energy, coming from eq. (1.20). When both $\epsilon_1, \epsilon_2 \ll 1$, we have $\ddot{\phi} \ll H\dot{\phi}$ cf. eq. (1.21). Then, the first term in the scalar field equation of motion eq. (1.12) can be neglected and we can reduce a second-order problem to a first-order problem. These approximations result in the dynamical

equations of 'slow roll inflation':

$$3H\dot{\phi} \simeq -V'(\phi), \quad (1.24)$$

$$H^2 \simeq \frac{V(\phi)}{3M_{\text{Pl}}^2}. \quad (1.25)$$

When these equations are valid, H is entirely a function of V . Further, when ϕ monotonically increases or decreases with time, as is usually the case during inflation, we can use ϕ instead of t as the time variable. This enables us to write the slow roll parameters in terms of the shape of the potential, ie. V and its derivatives with respect to ϕ :

$$\epsilon_1 \simeq \frac{M_{\text{Pl}}^2}{2} \left(\frac{V'(\phi)}{V} \right)^2, \quad (1.26)$$

$$\epsilon_2 \simeq 2M_{\text{Pl}}^2 \left[\left(\frac{V'(\phi)}{V} \right)^2 - \frac{V_{,\phi\phi}}{V} \right]. \quad (1.27)$$

These approximate expressions only hold when the slow roll parameters are small. In numerical calculations, we give initial conditions many e-folds before the end of inflation, and at such early times we use the slow roll approximation to give initial conditions—given the initial field value, the field velocity does not need to be specified and is determined by eq. (1.24). The fact that the slow roll equation of motion eq. (1.24) is first order instead of the original equation of motion eq. (1.12) allows us to find analytical solutions for $\phi(t)$ in the slow roll approximation for a lot of potentials, while it is more difficult to find an analytical solution to eq. (1.12).

For building more complicated models, one can consider non-canonical fields, for which the kinetic term in the Lagrangian is different from $\partial^\mu\phi\partial_\mu\phi/2$, or fields that are nonminimally coupled to gravity, which have a term explicitly dependent on the spacetime geometry in their action. Also, one does not have to limit themselves to a single field by any means. Models beyond canonical single field inflation are out of the scope of this thesis.

1.3 The evolution of primordial perturbations

We can now turn to the focus of this thesis: primordial perturbations. We can make a lot of predictions with linear perturbation theory in cosmology. Throughout this thesis, we will deal with linear perturbations in an FLRW background, except when we deal with secondary gravitational waves. The reader can refer to [2], [3], and [4] for extensive reviews on cosmological perturbation theory.

In the FLRW background metric, the metric perturbations can be decomposed according to their behaviour under local spatial rotations on an equal time hypersurface. This property leads to the classification of perturbations as scalars, vectors, and tensors under the group of rotations in 3 spatial dimensions. Scalar perturbations are invariant under rotations. Vector and tensor perturbations, as their names indicate, transform as vectors and tensors respectively.

A cornerstone result of linear cosmological perturbation theory is the *decomposition theorem*. It states two things:

1. Scalar, vector, and tensor perturbations of the metric give rise to scalar, vector, and tensor perturbations respectively in the stress energy tensor.
2. Scalar, vector, and tensor perturbations evolve completely independently of each other.

Note that both these statements are only being made for linear perturbation theory. We will see that the decomposition theorem does not apply at higher orders in perturbation theory when we discuss secondary gravitational waves.

We have developed a simple proof for the decomposition theorem for flat and curved FLRW backgrounds. The reader can refer to the Appendix for this proof and for the issue of gauge fixing in cosmological perturbation theory. In the subsequent sections, we are going to make use of the decomposition theorem and work with gauge invariant quantities. We will also restrict ourselves to a flat FLRW background, as there is strong observational evidence from the CMB that our universe is close to being spatially flat [1].

1.3.1 Scalar perturbations

We shall work in the so-called longitudinal gauge. There are two degrees of freedom in the scalar perturbations, which are the gauge invariant quantities Φ and Ψ , called the Bardeen potentials. The metric in terms of Φ and Ψ is

$$ds^2 = (1 + 2\Phi)dt^2 - a^2(1 - 2\Psi)\delta_{ij}dx^i dx^j . \quad (1.28)$$

From this perturbed metric it is tedious but straightforward to get the perturbed Einstein tensor at linear order. To write down the Einstein equations, we now need to model the perturbations in the stress energy tensor.

We consider the case of a scalar field, which does not have anisotropic stresses at linear order in perturbations because we can write it as a perfect fluid as shown by eq. (1.16). The scalar perturbations generated by a scalar field are

$$T_0^0 = \delta\rho = \dot{\phi}\delta\dot{\phi} - \dot{\phi}^2\Phi + V'(\phi)\delta\phi , \quad (1.29)$$

$$T_i^0 = \partial_i\delta\sigma = \dot{\phi}\partial_i\delta\phi , \quad (1.30)$$

$$T_i^i = \delta p\delta_j^i = \delta_j^i(\dot{\phi}\delta\dot{\phi} - \dot{\phi}^2\Phi - V'(\phi)\delta\phi) . \quad (1.31)$$

A useful quantity to define is the so called comoving curvature perturbation \mathcal{R} .

$$\mathcal{R} = \Phi + \frac{2\rho}{3\mathcal{H}} \frac{\Phi' + \mathcal{H}\Phi}{\rho + p} . \quad (1.32)$$

Note that a prime denotes the derivative with respect to conformal time $d\eta = dt/a$. \mathcal{R} is gauge invariant. We will now derive the equation of motion of \mathcal{R} in the presence of a scalar field.

On substituting the scalar field perturbations eq. (1.29)-eq. (1.31) in the Einstein equations in the appendix eq. (A.40), after some further algebra, gives the equation of motion of \mathcal{R} . We have stated the below equation in Fourier space, as it becomes much more convenient to analyze linear perturbations

in Fourier space since different Fourier modes are not coupled.

$$\mathcal{R}_k'' + 2\frac{z'}{z}\mathcal{R}_k' + k^2\mathcal{R}_k = 0, \quad (1.33)$$

$$\text{where } z = \frac{a\phi'}{\mathcal{H}}. \quad (1.34)$$

We also define the Mukhanov Sasaki variable v_k as $v_k \equiv \mathcal{R}_k z$. The equation of motion of v is

$$v_k'' + \left(k^2 - \frac{z''}{z}\right)v_k = 0. \quad (1.35)$$

1.3.2 Vector perturbations

The metric with vector perturbations is given by

$$ds^2 = dt^2 - 2a(t)(\partial_i B + S_i). \quad (1.36)$$

Note that we are only considering scalar fields which are perfect fluids, whose stress energy tensor perturbations are given by eq. (1.29)-eq. (1.31). There are no vector perturbations in the stress energy tensor. Thus, there are no sources for B_i and S_i . We will not discuss them further in this thesis.

1.3.3 Tensor perturbations

Tensor perturbations are described by the transverse and traceless tensor h_{ij} . Tensor perturbations do not require gauge fixing as h_{ij} is already gauge invariant. The metric for tensor perturbations around an FLRW background is

$$ds^2 = dt^2 - a^2(\delta_{ij} + h_{ij})dx^i dx^j. \quad (1.37)$$

In the absence of anisotropic stresses, the Fourier space Einstein equation for h_{ij} is given by

$$h_k'' + 2\mathcal{H}h_k' + k^2 h_k = 0. \quad (1.38)$$

Note that the notation in this equation hides the fact that h_{ij} has two degrees of freedom. This corresponds to two polarizations called $+$ and \times . However, in the absence of anisotropic stress the evolution of the two polarizations is identical, hence the indices ij are usually suppressed.

In analogy with scalar perturbations, for the variable $u_k \equiv ah_k$,

$$u_k'' + \left(k^2 - \frac{a''}{a} \right) u_k = 0. \quad (1.39)$$

1.4 Quantizing primordial perturbations

After obtaining how metric perturbations evolve, the question we now address is what are the initial conditions to evolve them from, and at what time should the initial conditions be imposed.

In an inflating universe, successively smaller scales are exiting the Hubble radius as time passes. If we go far back enough in time, even the largest observable scales today were much smaller than the Hubble radius. In Fourier space this means $k \gg \mathcal{H}$, where $\mathcal{H} = aH = a'/a$ is called the comoving Hubble radius. Small wavelength modes don't feel the expansion of the universe and evolve as they would in Minkowski spacetime. As a result, we can apply standard quantum field theory and canonically quantize them.

At linear order in the slow roll parameters, the quantities z''/z and a''/a appearing in the evolution equations of v_k , eq. (1.35), and u_k , eq. (1.39), are given by

$$\frac{z''}{z} = \frac{4 + 6\epsilon_1 + 3\epsilon_2}{2\eta^2}, \quad (1.40)$$

$$\frac{a''}{a} = \frac{2 + 6\epsilon_1}{\eta^2}. \quad (1.41)$$

When the slow roll approximation holds well, ϵ_1 and ϵ_2 can be neglected in the numerators. Then

$$z''/z = a''/a = 2/\eta^2, \quad (1.42)$$

so v_k and u_k have the same equation of motion. We work with v_k here, and the analysis for u_k is

identical. The equation of motion for v_k becomes

$$v_k'' + \left(k^2 - \frac{2}{\eta^2}\right)v_k = 0. \quad (1.43)$$

The solution to this is

$$v_k(\eta) = c_1 \left(1 - \frac{i}{k\eta}\right) e^{-ik\eta} + c_2 \left(1 + \frac{i}{k\eta}\right) e^{ik\eta}. \quad (1.44)$$

Now, we quantize v_k canonically. Note that we now have to include its dependence on the full vector \mathbf{k} instead of its magnitude k . Parts of the below discussion on quantization follow the discussion in [5].

$$\hat{v}_{\mathbf{k}}(\eta) = v_{\mathbf{k}}(\eta)\hat{a}_{\mathbf{k}} + v_{\mathbf{k}}^*(\eta)\hat{a}_{-\mathbf{k}}^\dagger, \quad (1.45)$$

with the canonical commutation relation

$$[\hat{a}_{\mathbf{k}}, \hat{a}_{\mathbf{k}'}^\dagger] = \delta^3(\mathbf{k} + \mathbf{k}'). \quad (1.46)$$

The canonical commutation relation can also be written as

$$[\hat{v}_{\mathbf{k}}, \hat{v}'_{\mathbf{k}'}] = i\delta^3(\mathbf{k} + \mathbf{k}'). \quad (1.47)$$

On calculating $[\hat{v}_{\mathbf{k}}, \hat{v}'_{\mathbf{k}'}]$ from eq. (1.45) and eq. (1.46), and equating it to the RHS of eq. (1.47), we get

$$v_{\mathbf{k}}(\eta)v_{\mathbf{k}'}^{*\prime}(\eta) - v_{\mathbf{k}}^*(\eta)v'_{\mathbf{k}'}(\eta) = i. \quad (1.48)$$

On substituting $v_{\mathbf{k}}(\eta)$ from eq. (1.44), we get

$$|c_1|^2 - |c_2|^2 = \frac{1}{2k}. \quad (1.49)$$

The vacuum state $|0\rangle$ is defined by $a_{\mathbf{k}}|0\rangle = 0$. The Hamiltonian operator for $\hat{v}_{\mathbf{k}}$ is given by

$$\hat{H} = \int d^3x \left(\frac{1}{2} \hat{\pi}^2 + \frac{1}{2} (\nabla v)^2 + \frac{1}{2} m_{eff}^2(\eta) \hat{v}^2 \right), \quad (1.50)$$

$$\text{where } \hat{\pi} = \hat{v}'_{\mathbf{k}}, \quad (1.51)$$

$$m_{eff}^2(\eta) = \frac{z''}{z} = \frac{2}{\eta^2}. \quad (1.52)$$

On substituting the mode expansion eq. (1.45) in the Hamiltonian eq. (1.50) and taking the vacuum expectation value, we get

$$\langle 0 | \hat{H} | 0 \rangle = \frac{1}{4} \delta^3(0) \int d^3k E_{\mathbf{k}}(\eta), \quad (1.53)$$

$$\text{where } E_{\mathbf{k}} = |\hat{v}'_{\mathbf{k}}|^2 + \omega_k^2 |v_{\mathbf{k}}|^2, \quad (1.54)$$

$$\omega_k = k^2 - \frac{2}{\eta^2}. \quad (1.55)$$

The $\delta^3(0)$ is an artefact of integrating over an infinite volume. At early times, we have

$$\lim_{\eta \rightarrow -\infty} v_{\mathbf{k}}(\eta) = c_1 e^{-ik\eta} + c_2 e^{ik\eta}. \quad (1.56)$$

Thus, we have

$$\lim_{\eta \rightarrow -\infty} E_{\mathbf{k}}(\eta) = 4(|c_1|^2 + |c_2|^2) k^2. \quad (1.57)$$

We demand that the vacuum state is the ground state of the Hamiltonian. Thus, we minimize eq. (1.57) given the constraint eq. (1.49), which gives us

$$|c_1| = \frac{1}{2k}, \quad c_2 = 0. \quad (1.58)$$

Upto an irrelevant phase, this uniquely determines the Bunch-Davies mode function-

$$v_{\mathbf{k}}(\eta) = \frac{1}{2k} \left(1 - \frac{i}{k\eta} \right) e^{-ik\eta}. \quad (1.59)$$

This is our initial condition for $v_{\mathbf{k}}(\eta)$ at sufficiently early times, meaning at times when k is sufficiently

greater than \mathcal{H} .

We can carry out an identical analysis for $u_k(\eta)$ and arrive at the same Bunch-Davies mode function.

Thus, when we solve for primordial perturbations during inflation, for each scale k at which we want the primordial perturbations, we start evolution at a sufficiently early time, when $k \gg \mathcal{H}$, and at this time we impose the Bunch-Davies initial conditions:

$$\lim_{k/\mathcal{H} \rightarrow \infty} (v_k(\eta), u_k(\eta)) \rightarrow \frac{1}{\sqrt{2k}} e^{-ik\eta}. \quad (1.60)$$

1.5 Primordial power spectra

When it comes to making contact of the analysis of the previous sections with observations, what we observe is the correlation functions of the variables \mathcal{R}, v, u etc. Theoretical calculations give these correlations as vacuum expectation values. We observe two point correlations from the CMB, which in Fourier space means we observe the power spectra. The scalar power spectrum is defined by

$$\int_0^\infty \frac{dk}{k} \mathcal{P}_{\mathcal{R}}(k) = \int d^3k \int \frac{d^3(x-x')}{(2\pi^3)} \langle 0 | \hat{\mathcal{R}}(\eta, \mathbf{x}) \hat{\mathcal{R}}(\eta, \mathbf{x}') | 0 \rangle e^{-ik(x-x')}. \quad (1.61)$$

On using the mode decomposition eq. (1.45), we get

$$\mathcal{P}_{\mathcal{R}}(k) = \frac{k^3}{2\pi^2} |\mathcal{R}_k|^2 = \left(\frac{k^3}{2\pi^2} \right) \left(\frac{|v_k|}{z} \right)^2. \quad (1.62)$$

Similarly, the tensor power spectrum is

$$\mathcal{P}_T(k) = \frac{k^3}{2\pi^2} |h_k|^2 = \left(\frac{k^3}{2\pi^2} \right) \left(\frac{|u_k|}{a} \right)^2. \quad (1.63)$$

However, it is a convention to multiply the RHS by a factor of $4/M_{\text{Pl}}^2$. Thus, in this thesis we will use the following definition:

$$\mathcal{P}_T(k) = \left(\frac{4}{M_{\text{Pl}}^2}\right) \left(\frac{k^3}{2\pi^2}\right) \left(\frac{|u_k|}{a}\right)^2. \quad (1.64)$$

Note that since the primordial tensor power spectrum has not been detected yet, in this thesis when we use 'primordial power spectrum' or 'power spectrum' we will always mean the scalar power spectrum. The scalar and tensor spectral indices are defined by

$$n_s = 1 + \frac{d \ln \mathcal{P}_{\mathcal{R}}}{d \ln k}, \quad (1.65)$$

$$n_T = \frac{d \ln \mathcal{P}_T}{d \ln k}. \quad (1.66)$$

The tensor to scalar ratio r is defined as

$$r(k) = \frac{\mathcal{P}_T(k)}{\mathcal{P}_{\mathcal{R}}(k)}. \quad (1.67)$$

1.5.1 Calculating the power spectra in slow roll inflation

In this subsection, we calculate the power spectra assuming that the slow roll approximation holds well during inflation, meaning ϵ_1 and ϵ_2 are $\ll 1$ during inflation.

z is related to a in terms of the slow roll parameters as

$$z = \sqrt{2\epsilon_1} M_{\text{Pl}}. \quad (1.68)$$

Also, we can express the slow roll parameters in terms of \mathcal{H} and η as

$$\epsilon = 1 - \frac{\mathcal{H}'}{\mathcal{H}^2}, \quad (1.69)$$

$$\epsilon_2 = \frac{\epsilon'}{\mathcal{H}\epsilon}. \quad (1.70)$$

Using the above three equations, we get

$$\frac{z''}{z} = \mathcal{H}^2 \left[2 - \epsilon_1 + \frac{3}{2}\epsilon_2 + \frac{\epsilon_2'}{2\mathcal{H}} \right], \quad (1.71)$$

$$\frac{a''}{a} = \mathcal{H}^2(2 - \epsilon_1). \quad (1.72)$$

We can rewrite eq. (1.69) as

$$\eta = - \int \left(\frac{1}{1 - \epsilon_1} \right) d \left(\frac{1}{\mathcal{H}} \right). \quad (1.73)$$

On integrating this by parts and substituting ϵ_2 from (1.82), we get

$$\eta = - \frac{1}{(1 - \epsilon)\mathcal{H}} - \int \left(\frac{\epsilon_1 \epsilon_2}{(1 - \epsilon_1)^3} \right) d \left(\frac{1}{\mathcal{H}} \right). \quad (1.74)$$

Since we assume $\epsilon_1, \epsilon_2 \ll 1$, we can ignore the second term. Thus, at leading order in the slow roll approximation,

$$\mathcal{H} \simeq - \frac{1}{(1 - \epsilon_1)\eta}. \quad (1.75)$$

Using this expression for \mathcal{H} in eq. (1.71) and eq. (1.72), we get

$$\frac{z''}{z} = \frac{4 + 6\epsilon_1 + 3\epsilon_2}{\eta^2}, \quad (1.76)$$

$$\frac{a''}{a} = \frac{2 + 3\epsilon_1}{\eta^2}. \quad (1.77)$$

Also, at leading order in the slow roll parameters, ϵ_1 and ϵ_2 are constants. The solution of the evolution equation eq. (1.35) that satisfies the Bunch-Davies initial condition at $\eta \rightarrow -\infty$ is given by

$$v_k(\eta) \left(\frac{-\pi\eta}{4} \right)^{1/2} \exp \left[\frac{i\pi}{2} \left(v + \frac{1}{2} \right) \right] H_\nu^1(-k\eta). \quad (1.78)$$

H_ν^1 is the Hankel function of the first kind. At late times, when $-k\eta \rightarrow 0$, the scalar and tensor power spectra are time independent, which can be deduced from inserting in eq. (1.78) the asymptotic form

of the Hankel function:

$$\lim_{x \rightarrow 0} H_\nu^2(x) \propto x^{-1/2}. \quad (1.79)$$

Thus, the solutions v_k and u_k are Hankel functions of the same form as eq. (1.78), with ν for v_k and u_k respectively given by

$$\nu_s \equiv \frac{3}{2} + \epsilon_1 + \frac{\epsilon_2}{2}, \quad (1.80)$$

$$\nu_T \equiv \frac{3}{2} + \epsilon_1. \quad (1.81)$$

We now want to evaluate the power spectra in the super Hubble limit $(-k\eta) \rightarrow 0$. We use the asymptotic form of the Hankel function $H_\nu^1(x)$ as $x \rightarrow 0$:

$$H_\nu^1(x) \simeq \frac{-i\Gamma(\nu)}{\pi} \left(\frac{2}{x}\right)^\nu. \quad (1.82)$$

Using this in the definitions of the power spectra eq. (1.62) and eq. (1.64), we get

$$\mathcal{P}_{\mathcal{R}}(k) = \left(\frac{1}{32\pi^2 M_{\text{pl}}^2 \epsilon_1}\right) \left(\frac{\Gamma(\nu_s)}{\Gamma(3/2)}\right) \left(\frac{k}{a}\right)^2 \left(\frac{-k\eta}{2}\right)^{1-2\nu_s}, \quad (1.83)$$

$$\mathcal{P}_T(k) = \left(\frac{1}{2\pi^2 M_{\text{pl}}}\right) \left(\frac{\Gamma(\nu_T)}{\Gamma(3/2)}\right) \left(\frac{k}{a}\right)^2 \left(\frac{-k\eta}{2}\right)^{1-2\nu_T}. \quad (1.84)$$

We evaluate the above expressions *in terms of* quantities at Hubble exit (not *at* Hubble exit), i.e. when $k = \mathcal{H}$. We do this by noting that, from eq. (1.75) and the fact that ϵ_1 is constant at leading order in slow roll,

$$1 - \epsilon_1 = (-k\eta)_{k=\mathcal{H}}^{-1} = (-\mathcal{H}\eta)_{k=\mathcal{H}}^{-1}. \quad (1.85)$$

Using this, we get

$$\mathcal{P}_{\mathcal{R}}(k) = \left(\frac{H^2}{2\pi\dot{\phi}} \right)^2 \left(\frac{\Gamma(\nu_S)}{\Gamma(3/2)} \right) 2^{2\nu_S-3} (1 - \epsilon_1)^{2\nu_S-1}, \quad (1.86)$$

$$\mathcal{P}_T(k) = \left(\frac{2H^2}{\pi^2 M_{\text{Pl}}^2} \right)^2 \left(\frac{\Gamma(\nu_S)}{\Gamma(3/2)} \right) 2^{2\nu_S-3} (1 - \epsilon_1)^{2\nu_S-1}, \quad (1.87)$$

where the RHS are evaluated at $k = \mathcal{H}$. Using the slow roll approximation $\epsilon_1, \epsilon_2 \ll 1$, we can further simplify the above as

$$\mathcal{P}_{\mathcal{R}}(k) \simeq \left(\frac{H^2}{2\pi\dot{\phi}} \right)_{k=\mathcal{H}}^2, \quad (1.88)$$

$$\mathcal{P}_T(k) \simeq \left(\frac{8}{M_{\text{Pl}}^2} \right) \left(\frac{H}{2\pi} \right)_{k=\mathcal{H}}^2. \quad (1.89)$$

To get the spectral indices in terms of the slow roll parameters, we use the below result.

$$\left(\frac{dy}{d \ln k} \right)_{k=\mathcal{H}} = \left(\frac{dy}{dt} \frac{dt}{d \ln a} \frac{d \ln a}{d \ln k} \right)_{k=\mathcal{H}} = \left(\frac{\dot{y}}{H} \frac{d \ln a}{d \ln k} \right)_{k=\mathcal{H}} \simeq \left(\frac{\dot{y}}{H} \right)_{k=\mathcal{H}}. \quad (1.90)$$

Using this result in the definitions of n_s and n_T we get

$$n_s \simeq 1 - 2\epsilon_1 - \epsilon_2, \quad (1.91)$$

$$n_T = -2\epsilon_1. \quad (1.92)$$

It is to be understood that the LHS is a function of k and the RHS is evaluated at the Hubble exit of that k . The tensor to scalar ratio r is

$$r \simeq 16\epsilon_1. \quad (1.93)$$

$$\text{Thus, } r = -8n_T. \quad (1.94)$$

$r = -8n_T$ is dubbed the 'consistency relation' of single field slow roll inflation.

1.6 Ultra slow roll

The usefulness of the slow roll approximation comes from the fact that it results in a smooth, nearly scale-invariant power spectrum, which matches very well with observations by Planck. However, we have not measured the power spectrum on scales smaller than those probed by Planck. It is interesting to consider departures from slow roll when these small scales are exiting the Hubble radius during inflation. In this section, we discuss the mechanism of ultra slow roll (USR), which is a popular mechanism employed for generating a peak in the primordial power spectrum on small scales.

The slow roll approximation may fail not only if $V(\phi)$ becomes steep, which is what happens at the end of inflation, but also if it suddenly becomes extremely flat. This scenario is called USR.

In USR, in the field equation of motion eq. (1.12), it is the term $V'(\phi)$ term that is negligible, instead of $\ddot{\phi}$. Thus, the evolution of the field becomes insensitive to the potential, except for the fact that the expansion rate of the universe still depends on $V(\phi)$. The field equation of motion during USR is

$$\ddot{\phi} + 3H\dot{\phi} = 0. \quad (1.95)$$

For USR, the primordial power spectrum is not exactly given by eq. (1.88), but it is a good approximation and enough to convey our next point. Using the definition of ϵ_1 , in eq. (1.88), we can write

$$\mathcal{P}_{\mathcal{R}}(k) \simeq \frac{H^2}{8\pi^2 M_{\text{pl}}^2 \epsilon_1}. \quad (1.96)$$

From eq. (1.20), we have

$$\epsilon_1 \simeq \frac{3\dot{\phi}^2}{2V}. \quad (1.97)$$

From eq. (1.95), we find that $\dot{\phi}^2 \propto e^{-6N}$ where N is the number of e-folds. Thus, we have $\epsilon_1 \propto e^{-6N}$ and we find from eq. (1.96) that the power spectrum rises as it has ϵ_1 in the denominator. This results in an upward slope in the power spectrum for scales k which exited the Hubble radius in the USR

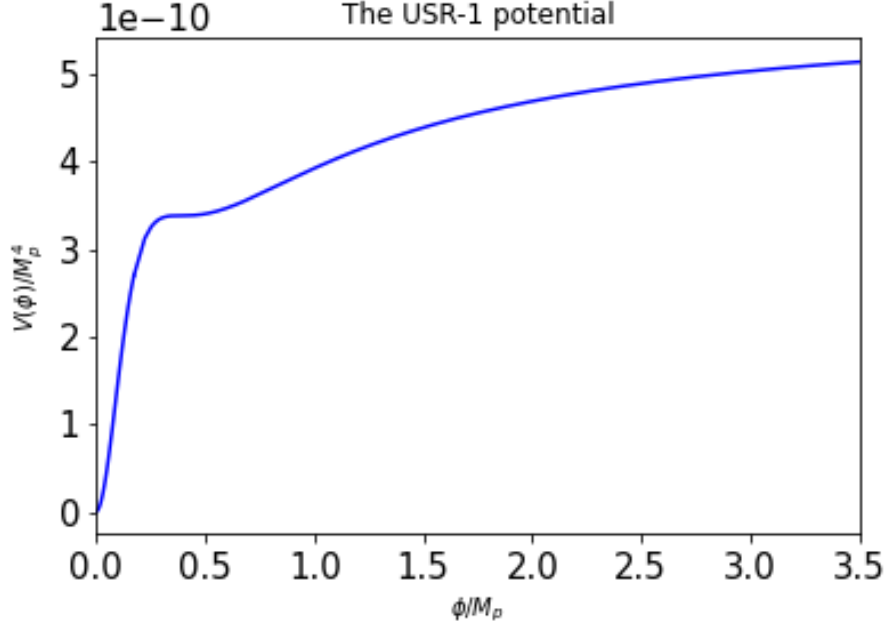


Figure 1.1: The USR-1 potential of eq. (1.98). It has an inflection point at $\phi \simeq 0.39M_{\text{pl}}$.

era. After USR ends, the power spectrum again starts to drop, and thus we get a peak in the power spectrum due to USR.

We end with an example of a USR model and its power spectrum. We consider the phenomenological potential called 'USR-1' discussed in [6]. The potential is given by

$$V(\phi) = V_0 \frac{3x^4 - 4\alpha x^3 + 6x^2}{(1 + \beta x^2)^2}, \quad (1.98)$$

where $x = \phi/v$. To evaluate the power spectrum, following [6], we fix the parameters $V_0 = 4 \times 10^{-10} M_{\text{pl}}^4$, $v = \sqrt{0.108} M_{\text{pl}}$, $\alpha = 1$, $\beta = 1.4349$, and $N_{\text{pivot}} = 50$.

1.7 Aims of this thesis

In the above sections, we encountered strong motivations from cosmology for an inflationary scenario and the physics of such an era. The most important thing now is to confront inflation with observations.

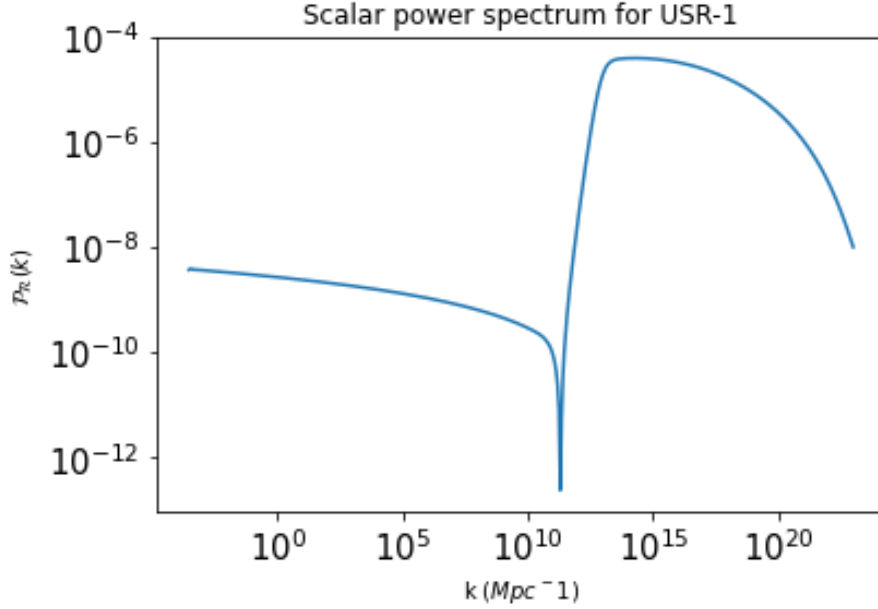


Figure 1.2: Power spectrum for the USR-1 potential, for the parameter values mentioned in the text. It is consistent with the Planck observations on the largest scales and has a peak at smaller scales.

The best constraints on inflation by far come from CMB anisotropies [1], which probe the largest observable scales. However, they cannot probe scales smaller than roughly $k \simeq 0.2 \text{ Mpc}^{-1}$, which is the limit beyond which the CMB anisotropies are too damped by diffusion to be detectable by Planck. To probe much smaller scales, one can turn to structures in the universe, but the problem here is that the overdensity needed for a galaxy to start forming is so large that it cannot be treated with perturbation theory, so modeling the physics and mapping a model of primordial perturbations to the corresponding observables becomes much more difficult.

In the recent past, some new avenues have emerged for constraining primordial perturbations on small scales while still staying in the framework of perturbation theory. In this thesis, we will look at a few of these—namely CMB spectral distortions, primordial black holes, and secondary gravitational waves. We will then use these probes to derive constraints on parametric forms of the primordial scalar power spectrum and an inflationary model.

Chapter 2

Probes of primordial perturbations on small scales

2.1 Introduction

What do we gain by probing smaller scales? There are two main benefits. The first is that probing a new range of scales means that we are probing a new period in the inflationary era, the period when those scales left the Hubble radius. This point is illustrated in fig. 2.1. The second benefit is that the constraints on primordial perturbations from small scale probes are far weaker than the constraints from CMB anisotropies, as we will see in later chapters. This allows for the possibility of the power spectrum having a large amplitude on smaller scales, which can lead to interesting phenomenology during inflation such as ultra slow roll, and interesting phenomenology in the universe after inflation, like primordial black holes and secondary gravitational waves.

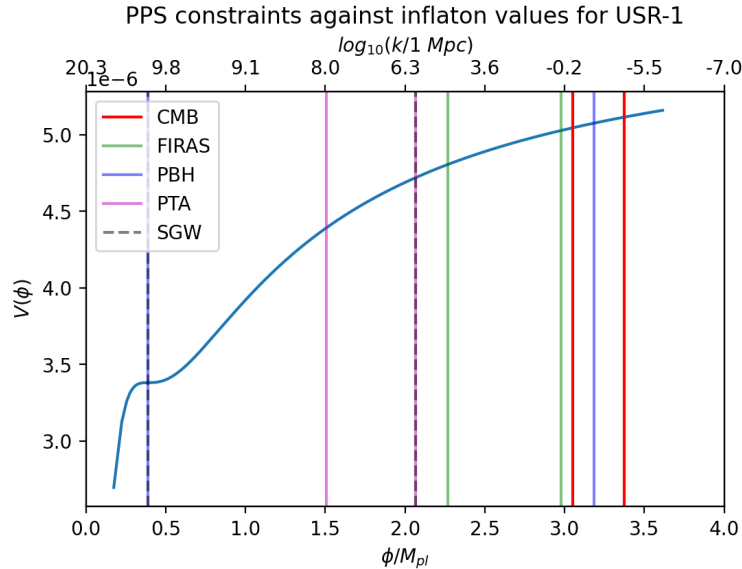


Figure 2.1: A rough picture of the scales probed by various cosmological observables for the USR-1 inflation scenario discussed in section 1.6. The upper x-axis shows the scales that exit the Hubble radius at the corresponding field value on the lower x-axis. Probing a new range of scales allows us to probe a new part of the inflaton potential. In the figure, FIRAS refers to constraints from the sensitivity of the FIRAS instrument to CMB spectral distortions, PTA refers to constraints from pulsar timing arrays, PBH refers to constraints from bounds on primordial black hole abundance, and SGW refers to constraints from the sensitivities of future gravitational wave observatories like the Laser Interferometer Space Antenna and Big Bang Observer. These probes are discussed in the forthcoming sections.

2.2 CMB spectral distortions

The CMB is the most perfect blackbody seen in the sky. This observation, however, is surprisingly old. The latest measurement of the CMB frequency spectrum dates back to 1996, by the FIRAS instrument on board the COBE satellite[7].

The FIRAS measurement did not detect any deviations from the blackbody spectrum in the CMB. This non-detection places substantial constraints on the primordial power spectrum, and the sensitivity of the proposed experiment called PIXIE is expected to strengthen these constraints by more than three orders of magnitude [8].

There are an infinite number of ways for a spectrum to be different from the blackbody shape. We are helped by the fact that we have a robust understanding of the thermodynamics of the radiation era, through which we have classified CMB spectral distortions into 3 broad types: the μ distortion, the y distortion, and the r distortion. We describe below how the dissipation of primordial perturbations leads to spectral distortions in the CMB. The discussion here follows that in [9].

2.2.1 Thermalization in the radiation era

In the radiation era, after neutrino decoupling and electron-positron annihilation, thermal equilibrium is maintained by interactions between electrons and photons, which balance the energy and particle number between photons and baryons. Three processes are involved in maintaining this equilibrium:

1. Double Compton scattering:



2. Bremsstrahlung

3. Compton scattering:



When all three processes are efficient, they quickly thermalize the baryon-photon plasma after any energy injection. As the temperature of the baryon-photon plasma drops, these processes lose their efficiency and perturbations in the energy or particle number form spectral distortions in the CMB.

The first process to become inefficient is Double Compton scattering. This happens at the redshift of $z \simeq 1.98 \times 10^6$. As it is a process that changes photon number, it becomes harder to balance the particle number. At high enough redshifts Compton scattering is still very efficient, so energy redistribution between photons and baryons is easy. Thus, deviation from thermal equilibrium in this period arises from a particle number imbalance between baryons and photons, which shows up as a nonzero chemical potential in the CMB spectrum. The chemical potential is denoted by the usual symbol μ , and this is called the μ distortion.

Below the redshift of $z \simeq 5 \times 10^4$, Compton scattering also becomes inefficient. This leads to deviations from thermal equilibrium appearing as a y distortion in the CMB spectrum. The same signature is in fact produced by galaxy clusters through the Sunyaev-Zeldovich effect. At the level of the global average over all directions, the signal from the Sunyaev-Zeldovich effect dominates over the signal we would expect from the radiation era, so the y distortion is not useful for our purpose of constraining primordial perturbations.

There is one oversimplification in the above picture of thermalization physics. The transition between the eras that lead to μ and y distortions is not sharp at $z \simeq 5 \times 10^4$, but occurs over an extended period $3 \times 10^5 \geq z \geq 10^4$. In this period, the shape of the distortion matches neither that of the μ or y distortions, and unlike the μ and y distortions, it also depends on the redshift at which energy is injected. This is known as the r distortion.

2.2.2 Formation of the μ distortion

It can be shown that most of the fluctuations at a given wavenumber k dissipate when k equals the diffusion damping scale k_D : [10]

$$k_D \simeq 4.0 \times 10^{-6} (1+z)^{3/2} \text{Mpc}^{-1}. \quad (2.3)$$

Equivalently, most of the fluctuations dissipate at the redshift

$$z \simeq 4.5 \times 10^5 \left(\frac{k}{10^3 \text{ Mpc}} \right)^{2/3}. \quad (2.4)$$

Approximating the transition between the μ and y distortion eras to be sharp, we can express the value of μ with the help of a window function:

$$\mu \simeq \int_{k_{min}}^{\infty} \frac{k^2 dk}{2\pi^2} \mathcal{P}_{\mathcal{R}}(k) W(k). \quad (2.5)$$

The cutoff k_{min} is introduced both because an analytical treatment of thermalization becomes harder for larger scales, and because it can be numerically verified that the contribution of the large scale primordial power spectrum obtained from the CMB to spectral distortions is small.

The expression for the window function, from a numerical fit, is [9]

$$W(k) \simeq 2.27 \left[\exp \left(- \frac{\left[\frac{\hat{k}}{1360} \right]^2}{1 + \left[\frac{\hat{k}}{260} \right]^{0.3} + \frac{\hat{k}}{340}} \right) - \exp \left(- \left[\frac{\hat{k}}{32} \right]^2 \right) \right] \quad (2.6)$$

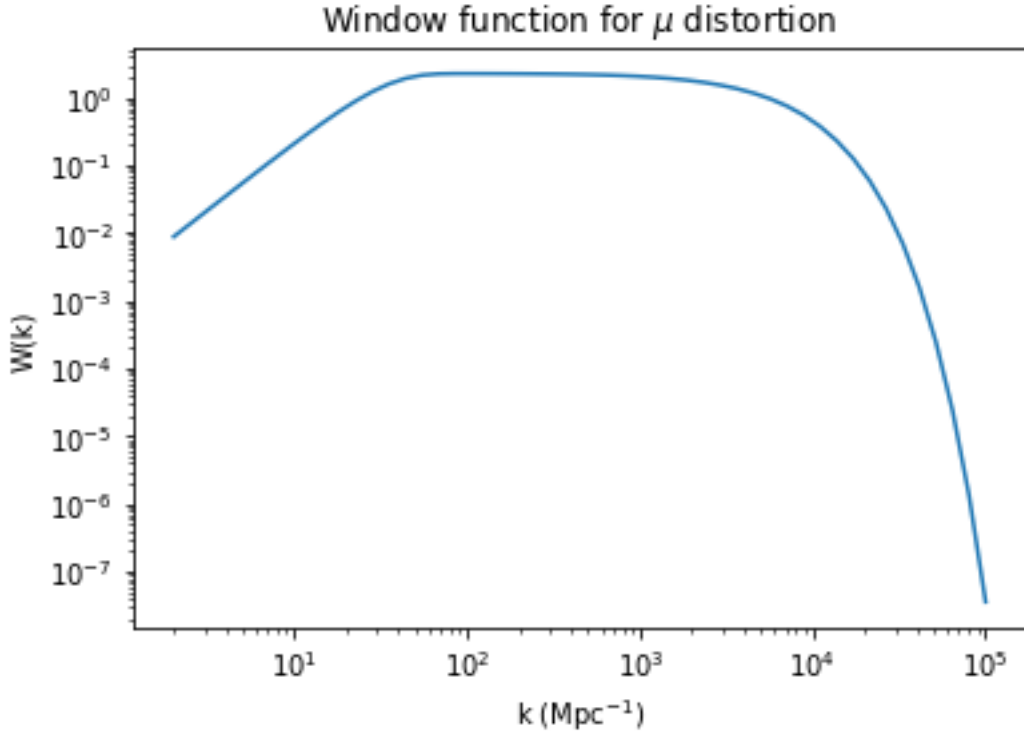


Figure 2.2: Window function for the μ distortion. It rises as k^2 for $k \lesssim 10^2 \text{ Mpc}^{-1}$ and falls off as $\exp(-k)$ for $k \gtrsim 10^4 \text{ Mpc}^{-1}$.

There is a nonzero value of μ expected from the Planck Λ CDM model itself, which corresponds to extrapolating the observed power law spectrum on CMB scales upto $k \lesssim 0.2 \text{ Mpc}^{-1}$ to smaller scales $k \geq 1 \text{ Mpc}^{-1}$. This value is of the order of 10^{-8} [11]. This is much below the sensitivity of FIRAS and even PIXIE and is not important when deriving constraints on the small scale power spectrum.

The most recent measurement of the frequency spectrum of the CMB dates back to 1996. It was

done by the FIRAS instrument on board the COBE satellite. It did not detect any spectral distortions from the CMB, but it placed an upper bound of $\mu < 9 \times 10^{-5}$ (95% CL) [7]. The most popular plan for a future observatory is the PIXIE experiment, with a sensitivity of $\mu \lesssim 3 \times 10^{-8}$ [8].

2.3 Primordial black holes from inflation

There is an interesting possibility that black holes (PBHs) formed in the very early universe before the scales that we observe in the CMB reentered the Hubble radius. If these black holes possess mass in a certain range, they can explain a fraction of or even all of the dark matter in the universe today ([12], [13]).

A promising mechanism for forming PBHs is provided by inflation. When a large overdensity enters the Hubble radius, it can undergo gravitational collapse to form a black hole. USR inflation, discussed in section 1.6, can generate such large overdensities, which correspond to peaks in the primordial power spectrum. We will discuss PBHs in detail in the following sections.

2.3.1 Formation of PBHs

When a mode enters the Hubble radius, if its amplitude is sufficiently large, there are overdense Hubble patches which collapse to form PBHs. Thus, the statistics of the primordial density perturbations will tell us what abundance of PBHs to expect and their masses. For a Gaussian field of density perturbations, its entire statistics is captured by its power spectrum which is called the matter power spectrum. We will use the so called Press-Schechter formalism to predict the PBH abundance from the matter power spectrum.

We assume that the density contrast after inflation $\delta(k) = \delta\rho/\rho$ is a Gaussian random field. We write it as

$$P(\delta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\delta^2}{2\sigma^2}}. \quad (2.7)$$

We consider that only perturbations with a δ larger than a certain δ_c will collapse to form PBHs. A discussion of the theoretical modelling needed to derive the value of δ_c is out of our scope. The

fraction of perturbations that have an overdensity above δ_c is denoted β and given by

$$\beta = \int_{\delta_c}^1 P(\delta) d\delta. \quad (2.8)$$

The upper limit of the integral is 1 and not infinity, because our analysis is valid only for linear perturbation theory which holds for $\delta < 1$. However, in our context, 1 lies far into the tail of the Gaussian due to the smallness of $\sigma^2(k)$. So, it is a good approximation to set the upper limit of the integral to infinity, and we obtain

$$\beta \simeq \int_{\delta_c}^{\infty} P(\delta) d\delta = \frac{1}{2} \left[1 - \operatorname{erf} \left(\frac{\delta_c}{\sqrt{2\sigma^2}} \right) \right]. \quad (2.9)$$

The matter power spectrum is related in a simple way to the primordial scalar power spectrum $\mathcal{P}_{\mathcal{R}}(k)$ on super-Hubble scales. In the radiation dominated era,

$$P_{\delta}(k) = \frac{16}{81} \left(\frac{k}{aH} \right)^4 \mathcal{P}_s(k). \quad (2.10)$$

We now introduce a smoothing scale R in the problem, with the help of a window function $W(kR)$ as follows:

$$\sigma^2(R) = \int_0^{\infty} \frac{dk}{k} P_{\delta}(k) W^2(kR). \quad (2.11)$$

It is common to use a Gaussian window function

$$W(kR) = e^{-\frac{k^2 R^2}{2}}. \quad (2.12)$$

Owing to the absence of any other length scale in the problem that differs considerably from the Hubble radius $(aH)^{-1}$, we set

$$R = (aH)^{-1}. \quad (2.13)$$

Now we consider the result of the gravitational collapse. The mass of the resulting PBH is related to the mass inside the Hubble radius at that time M_H , the only mass scale in the problem. The entire

mass inside the Hubble radius does not collapse to a black hole. We model the finite efficiency of the collapse saying that the PBH mass M is a fraction γ of M_H ,

$$M = \gamma M_H . \quad (2.14)$$

Using eq. (2.13), we find that R and M are related as

$$R = \frac{2^{1/4}}{\gamma^{1/2}} \left(\frac{g_{*, k}}{g_{*, \text{eq}}} \right)^{1/12} \left(\frac{1}{k_{\text{eq}}} \right) \left(\frac{M}{M_{\text{eq}}} \right)^{1/2} , \quad (2.15)$$

where eq denotes the instant of matter radiation equality. Since we know the redshift of matter radiation equality observationally, we know the mass M_{eq} . Substituting this above, we obtain

$$R = 4.72 \times 10^{-7} \left(\frac{\gamma}{0.2} \right)^{-1/2} \left(\frac{g_{*, k}}{g_{*, \text{eq}}} \right)^{1/12} \left(\frac{M}{M_{\odot}} \right)^{1/2} \text{ Mpc} . \quad (2.16)$$

Using the above results we arrive at the present fraction of PBHs of mass M in dark matter:

$$f_{\text{PBH}}(M) = 2^{1/2} \gamma^{3/2} \beta(M) \left(\frac{\Omega_m h^2}{\Omega_c h^2} \right) \left(\frac{g_{*, k}}{g_{*, \text{eq}}} \right)^{-1/4} \left(\frac{M}{M_{\text{eq}}} \right)^{-1/2} . \quad (2.17)$$

On using known values of $\Omega_m h^2$, $\Omega_c h^2$ and M_{eq} we get

$$f_{\text{PBH}}(M) = \left(\frac{\gamma}{0.2} \right)^{3/2} \left(\frac{\beta(M)}{1.46 \times 10^{-8}} \right) \left(\frac{g_{*, k}}{g_{*, \text{eq}}} \right)^{1/12} \left(\frac{M}{M_{\odot}} \right)^{1/2} . \quad (2.18)$$

To summarize, given a primordial power spectrum, we obtain the matter power spectrum eq. (2.10), and its variance eq. (2.11) using the window function eq. (2.12). We write $\sigma^2(R)$ as $\sigma^2(M)$ using the relation between R and M eq. (2.15), which gives us the fraction $\beta(M)$ from eq. (2.8). We can interpret $\beta(M)$ as the fraction of the total energy density of the universe contributed by PBHs of mass M . Finally, after substituting $\beta(M)$ in eq. (2.18) we obtain the abundance of PBHs of mass M .

2.3.2 Observational constraints on f_{PBH}

PBHs lead to a variety of astrophysical phenomenology, which allows us to place observational constraints on them for a wide range of PBH masses. The most elementary observational constraint that we should consider is the fact that since the energy density of PBHs scales as cold dark matter (CDM), this energy density can at most be that of CDM, which is $\Omega_c h^2$, measured from CMB observations. In short, $f_{PBH} \leq 1$. If the equality holds, PBHs are the full answer to the puzzle of the nature of CDM. We have mentioned various other ways to independently constrain the abundance of PBHs below, though the list is not exhaustive. Our discussion here follows that in [12].

1. Evaporation: There are two ways in which evaporation through Hawking radiation constrains PBHs. The first is that it severely constrains the abundance of PBHs that formed so long ago and are so light that the age of the universe exceeds their lifetime, i.e.

$$H_0^{-1} \gtrsim t_H \sim 10^{63} \left(\frac{M}{M_\odot} \right) \text{ yr} . \quad (2.19)$$

As a consequence, PBHs with $M \lesssim 10^{-18} M_\odot$ would have completely evaporated by now. Since most of the contribution to the age of the universe comes from after matter-radiation equality, we do not need to consider the tiny correction due to the time of formation of PBHs in the radiation era.

Another way in which evaporation constrains PBH abundance is through observational constraints on the background of Hawking radiation expected from PBHs that have been emitting it since they formed. An example of such a constraint is a constraint on the extragalactic γ ray background.

We should note that constraints from evaporation ignore the effects of the evolution of the mass of a PBH after its formation, e.g. due to accretion or mergers.

2. Gravitational microlensing: This is an observable effect that does not rely on either the primordial or the black hole nature of PBHs, but just on the fact that they are compact objects.

There are observational constraints on f_{PBH} from stellar microlensing from within and outside the Milky Way. Stellar microlensing occurs when a compact object in the mass range $5 \times$

$10^{-10} \lesssim M/M_{\odot} \lesssim 10$ crosses our line of sight to a star. There are also constraints on PBHs with $10 \lesssim M/M_{\odot} \lesssim 100$ from non-observation of strong lensing of fast radio bursts.

3. Dynamical interactions: These refer to the disruptions in stellar orbits due to encounters with compact objects like PBHs. There are observational constraints on f_{PBH} from dynamical effects on the stellar populations of ultra-faint dwarf galaxies, and on wide binaries in the Milky Way.
4. Accretion: Accretion onto PBHs in the radiation era can lead to spectral distortions in the CMB. PBHs with mass in the range $10^3 \lesssim M/M_{\odot} \lesssim 10^{12}$ are constrained from the FIRAS constraint on the CMB μ distortion.

2.4 Secondary gravitational waves

The word "secondary" in secondary gravitational waves (SGWs) is due to the fact that these are a prediction of second order perturbation theory. In contrast to linear perturbation theory, at second order the equation of motion of the tensor perturbations h_{ij} acquires a source term that is second order in the scalar perturbations. We shall state the equation of motion in Fourier space without proof:

$$h^{\lambda\prime\prime} + 2\mathcal{H}h^{\lambda\prime} + k^2 h^{\lambda} = S^{\lambda} , \quad (2.20)$$

where h is the Fourier mode of the tensor perturbation, \mathcal{H} is the comoving Hubble radius, S is the source term, and $\lambda \in \{+, \times\}$ is the polarization index. The position space perturbation $h_{ij}(\eta, \mathbf{x})$ is decomposed into its polarizations as

$$h_{ij}(\eta, \mathbf{x}) = \int \frac{d^3 \mathbf{k}}{(2\pi)^{3/2}} [e_{ij}^+(\mathbf{k})h_{\mathbf{k}}^+(\eta) + e_{ij}^{\times}(\mathbf{k})h_{\mathbf{k}}^{\times}(\eta)] e^{i\mathbf{k}\cdot\mathbf{x}} , \quad (2.21)$$

where the polarization tensors are

$$\begin{aligned} e_{ij}^+(\mathbf{k}) &= \frac{1}{\sqrt{2}} [e_i(\mathbf{k})e_j(\mathbf{k}) - \bar{e}_i(\mathbf{k})\bar{e}_j(\mathbf{k})] \\ e_{ij}^\times(\mathbf{k}) &= \frac{1}{\sqrt{2}} [e_i(\mathbf{k})\bar{e}_j(\mathbf{k}) + \bar{e}_i(\mathbf{k})e_j(\mathbf{k})] \end{aligned} \quad (2.22)$$

and the polarization vectors are defined such that $e_i(\mathbf{k})$, $\bar{e}_i(\mathbf{k})$, and $\hat{\mathbf{k}}$ form an orthonormal set of vectors.

2.4.1 Solving for the SGW spectrum

We are now looking to solve eq. (2.20) in a radiation dominated background. We will make a further approximation by neglecting anisotropic stress and setting $\Phi = \Psi$. This allows us to write the source term as [14]

$$\begin{aligned} S_k^\lambda(\eta) &= 4 \int \frac{d^3p}{(2\pi)^{3/2}} e_{ij}^\lambda(\mathbf{k}) p^i p^j \{ 2\Phi_p(\eta)\Phi_{k-p}(\eta) \\ &\quad + [\Phi_p(\eta) + \eta\Phi'_p(\eta)] [\Phi_{k-p}(\eta) + \eta\Phi'_{k-p}(\eta)] \} . \end{aligned} \quad (2.23)$$

We consider inflation to be driven by a single canonical scalar field. There is no source of polarization of the tensor perturbations in this case, so the equations of motion for the two polarizations are in fact identical. Owing to this fact, we will drop the polarization label λ from our subsequent expressions unless polarization is important. Also, we will multiply the tensor power spectrum resulting from eq. (2.20) without the polarization index with a factor of 2 to account for the 2 polarizations.

Given all the above, we proceed to solve eq. (2.20) through the Green's function method. But first, since the primordial quantity whose power spectrum we compute is the comoving curvature perturbation \mathcal{R} and not Φ , we substitute \mathcal{R} for Φ in the above using the relation

$$\Phi_{\mathbf{k}}(\eta) = \mathcal{T}(k\eta)\mathcal{R}_{\mathbf{k}(\eta)} , \quad (2.24)$$

where the transfer function in the radiation era is

$$\mathcal{T}(k\eta) = \frac{9}{(k\eta)^2} \left[\frac{\sin(k\eta/\sqrt{3})}{k\eta/\sqrt{3}} - \cos(k\eta/\sqrt{3}) \right]. \quad (2.25)$$

The Green's function method tells us that the solution for the tensor mode is

$$h_{\mathbf{k}}^s(\eta) = \frac{1}{a(\eta)} \int^\eta d\tilde{\eta} g_{\mathbf{k}}(\eta, \tilde{\eta}) a(\tilde{\eta}) \widehat{S}^s(\tilde{\eta}, \mathbf{k}), \quad (2.26)$$

with the Green's function for eq. (2.20) during radiation domination is given by

$$g_{\mathbf{k}}(\eta, \tilde{\eta}) = \frac{\sin(k(\eta - \tilde{\eta}))}{k} \theta(\eta - \tilde{\eta}). \quad (2.27)$$

The next few steps are straightforward but very tedious. We substitute the Green's function and the source term explicitly into eq. (2.26), which gives us the solution for $h_{\mathbf{k}}(\eta)$. We obtain the tensor power spectrum using eq. (1.63). We state the expression for $\mathcal{P}_h(k)$ omitting its algebraic derivation:

$$\begin{aligned} \overline{\mathcal{P}_h(k, \eta)} &= \frac{2}{81k^2\eta^2} \int_0^\infty dv \int_{|1-v|}^{1+v} du \left[\frac{4v^2 - (1 + v^2 - u^2)^2}{4uv} \right]^2 \mathcal{P}_S(kv) \mathcal{P}_S(ku) \\ &\times [\mathcal{I}_c^2(u, v) + \mathcal{I}_s^2(u, v)], \end{aligned} \quad (2.28)$$

The overline in $\overline{\mathcal{P}_h(k, \eta)}$ is because we have averaged over time scales of a few $k\eta$, assuming that we are working with small scales $k\eta \gg 1$. In the above expression, the functions \mathcal{I}_c and \mathcal{I}_s are given by

$$\mathcal{I}_c(v, u) = -\frac{27\pi}{4v^3u^3} \Theta(v + u - \sqrt{3}) (v^2 + u^2 - 3)^2, \quad (2.29)$$

$$\mathcal{I}_s(v, u) = -\frac{27}{4v^3u^3} (v^2 + u^2 - 3) \left[4vu + (v^2 + u^2 - 3) \ln \left| \frac{3 - (v - u)^2}{3 - (v + u)^2} \right| \right]. \quad (2.30)$$

For brevity, we introduce the kernel $K(u, v)$ using

$$\overline{\mathcal{P}_h(k, \eta)} = \frac{2}{81k^2\eta^2} \int_0^\infty dv \int_{|1-v|}^{1+v} du K(u, v) \mathcal{P}_S(kv) \mathcal{P}_S(ku). \quad (2.31)$$

$K(u, v)$ is plotted below.

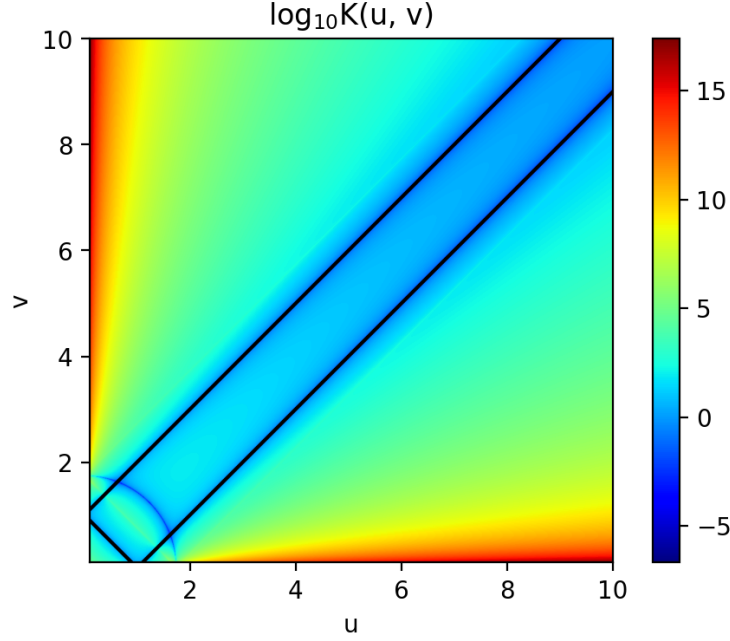


Figure 2.3: The kernel $K(u, v)$ appearing in eq. (2.31). The black lines enclose the region over which it is integrated for obtaining the SGW spectrum. The kernel vanishes along the limit points of the du integral, and diverges as $u \rightarrow 0$ or $v \rightarrow 0$.

The energy density of gravitational waves $\Omega_{GW}(k, \eta)$ is given by

$$\Omega_{GW}(k, \eta) = \frac{\rho_{GW}(k, \eta)}{\rho_{cr}(\eta)} = \frac{1}{24} \left(\frac{k}{\mathcal{H}} \right)^2 \overline{\mathcal{P}_h(k, \eta)} = \frac{k^2 \eta^2}{24} \overline{\mathcal{P}_h(k, \eta)}. \quad (2.32)$$

Our target is to obtain $\Omega_{GW}(k, \eta_0)h^2$, the present day energy density in gravitational waves. To do this, we evolve $\Omega_{GW}(k, \eta)$ to the present day, which gives (we omit the η_0 hereafter)

$$\Omega_{GW}(k)h^2 = 1.38 \times 10^{-5} \left(\frac{g_{*,k}}{106.75} \right)^{-1/3} \left(\frac{\Omega_r h^2}{4.16 \times 10^{-5}} \right) \Omega_{GW}(k, \eta). \quad (2.33)$$

2.4.2 Observing SGWs in pulsar timing arrays

Pulsars have a very robust periodicity, meaning that the arrival times of their pulses can be accurately predicted by theoretical modeling. This enables their use as a GW observatory, as extra noise in the

times of arrival (TOAs) of pulses from the corresponding theoretical model for the particular pulsar can be used to infer a GW passing through the line of sight. The sensitivity to $\Omega_{GW}(f)$ is proportional to the variance to which a pulsar's period can be determined.

Since we are interested in an isotropic background of GWs, we need to cross-correlate pulsars on different lines of sight to detect a common isotropic noise source in their TOAs, and the more pulsars we observe the better our constraints on the GW background. This is how pulsar timing arrays (PTAs) are used as GW observatories. The frequency range a PTA is most sensitive to corresponds to the total time span of its observations. This time span is of the order of years, corresponding to frequencies in the nanohertz range, and wavenumbers in the range $10^6 - 10^7 \text{ Mpc}^{-1}$.

The most simplified version of the constraints on a gravitational wave from a PTA is obtained by assuming a power law spectral shape of $\Omega_{GW}(f)h^2$, and then reporting the upper bound or a detection on it at a reference frequency f_{ref} . Typically f_{ref} is taken to be $1 \text{ yr}^{-1} = 31.7 \times 10^{-9} \text{ Hz}$. The bound on $\Omega_{GW}(f)h^2$ turns out not to be highly sensitive to the spectral index of the power law and lies in the order of magnitude $\Omega_{GW}h^2 \lesssim 10^{-10}$. The lack of sensitivity to the spectral index enables us to use the same bound as a reasonable approximation in our case where the SGW frequency spectrum is more complex than a power law. For concreteness, we take the bound at the value corresponding to the NANOGrav 2020 detection for a scale-invariant $\mathcal{P}_h(k)$: [15]

$$\Omega_{GW}(1 \text{ yr}^{-1})h^2 \simeq 6.4 \times 10^{-10} . \quad (2.34)$$

Note that the NANOGrav publication reports the value for $\Omega_{GW}(1 \text{ yr}^{-1})h^2$ for $\Omega_{GW}(f)$ assuming it has a spectral index of 13/3. We have converted that upper bound to that for a scale invariant $\Omega_{GW}(f)$, with the reasoning that if we expect the GW spectrum to have a peak in the relevant frequency band, it would have both a positive and a negative slope on either side of the peak and so it should not be approximated to have a positive or negative slope over the entire observational frequency range. The conversion scheme is taken from Figure 3 in [16].

For our purposes the equality eq. (2.34) is no different than an upper bound, since for both an equality and an upper bound the interpretation of our constraints is that only power spectra that generate an $\Omega_{GW}h^2$ below the RHS value are consistent with observations.

2.4.3 Observing SGWs in interferometers

The mechanism of detecting GWs from an interferometer is well-known. A GW passing through the apparatus of an interferometer changes its relative arm lengths, which is detected as a phase difference between beams of laser light passing through the different arms. Like for PTAs, in order to detect an isotropic GW background we need to cross-correlate the signals of more than one interferometers and identify the common contribution of the GW background. Parts of the discussion here follow that in [17].

The quantity measured in an interferometer is the GW strain $h(t)$, defined as the relative difference in arm length. Without going into details, we will state that $h(t)$ is given by convolving the tensor perturbations at the interferometer location with the instrument response:

$$h(t) = \int_{-\infty}^{\infty} df \int d^2\Omega_{\hat{k}} \sum_A R^\lambda(f, \hat{k}) h_\lambda(f, \hat{k}) e^{i2\pi f(t - \hat{k} \cdot \vec{x}/c)}. \quad (2.35)$$

Equivalently, in the frequency domain,

$$\tilde{h}(f) = \int d^2\Omega_{\hat{k}} \sum_A R^A(f, \hat{k}) h_A(f, \hat{k}) e^{-i2\pi f \hat{k} \cdot \vec{x}/c}. \quad (2.36)$$

Now we come to the cross-correlation between two detectors. This is given by

$$\langle \tilde{h}_1(f) \tilde{h}_2^*(f') \rangle = \frac{1}{2} \delta(f - f') \Gamma_{12}(f) S_h(f), \quad (2.37)$$

where the quantity $S_h(f)$ is defined as

$$S_h(f) = \frac{3H_0^2}{2\pi^2} \frac{\Omega_{GW}}{f^3}, \quad (2.38)$$

and the quantity $\Gamma_{12}(f)$, called the overlap reduction function, is given by

$$\Gamma_{12}(f) = \frac{1}{8\pi} \int d^2\Omega_{\hat{k}} \sum_A R_1^A(f, \hat{k}) R_2^{A*}(f, \hat{k}) e^{-i2\pi f \hat{k} \cdot (\vec{x}_1 - \vec{x}_2)}, \quad (2.39)$$

where \vec{x}_1 and \vec{x}_2 are the positions of the interferometers. The overlap reduction function depends on

both the distance between the interferometers and their orientation.

The noise for an interferometer is modeled as a noise spectral density $P_n(f)$. It has the same dimensions as $S_h^2(f)$. We can now write the expression for the signal-to-noise ratio (SNR) ρ for a cross-correlation between two interferometers:

$$\rho = \sqrt{2T} \left[\int_{f_{\min}}^{f_{\max}} df \frac{\Gamma_{IJ}^2(f) S_h^2(f)}{P_{nI}(f) P_{nJ}(f)} \right]^{1/2}, \quad (2.40)$$

where T is the total time of observation. A GW spectrum is said to be detected if it results in an SNR larger than the chosen threshold, which we choose to be $\rho = 1$. Also, we choose $T = 1$ yr. Thus, our observational upper bounds correspond to a GW spectrum which will lead to an SNR of unity after one year of interferometer observations.

Chapter 3

Deriving small scale constraints on the power spectrum

In this chapter, we use the theory of primordial perturbations and their phenomenological consequences on small scales to derive constraints on the power spectrum. We also show how constraints on the power spectrum can be interpreted as constraints on the inflaton potential when we have an underlying model of single field inflation.

3.1 Constraining parametric power spectra

The key difference between the current observational status of primordial perturbations on CMB scales and on the scales of spectral distortions, PBHs and SGWs is that on the CMB scales we have directly observed the power spectrum in the data, while for the small scale probes we only have upper bounds on their magnitudes from instrument sensitivities. Thus, the difference is that on CMB scales the constraints are from effects of the primordial perturbations that are already observed in the data, while on small scales the constraints are from the non-detection of these effects.

3.1.1 Parametrization of the power spectrum

On small scales we have much less information available to derive constraints, as all the small scale constraints we have discussed reduce the power spectrum to a single observable quantity by integrating over it, viz. μ for the μ distortion cf. eq. (2.5), $\Omega_{GW}(1 \text{ yr}^{-1} \text{h}^2)$ for PTAs, and the SNR ρ for interferometers cf. eq. (2.40). This is in contrast to the CMB data which has one data point per multipole for both temperature and polarization.

We treat the power spectrum at CMB scales and at small scales separately. We are able to do this because the amplitude of the power spectrum at the CMB scales is measured to be of the order of 10^{-9} , which is so small that it has a negligible effect on all observables discussed in the previous chapter. We treat large and small scales separately by approximating the power spectrum as a sum of spectra for large and small scales:

$$\mathcal{P}_{\mathcal{R}}(k) \simeq \mathcal{P}_{\mathcal{R},\text{large}}(k) + \mathcal{P}_{\mathcal{R},\text{small}}(k) , \quad (3.1)$$

where we classify large scales as $k \lesssim 1 \text{ Mpc}^{-1}$ and small scales as $k \gtrsim 1 \text{ Mpc}^{-1}$. In the next chapter, we will work with profiles for $\mathcal{P}_{\mathcal{R},\text{small}}(k)$ which have a peak on small scales, thus we can neglect $\mathcal{P}_{\mathcal{R},\text{large}}(k)$ on small scales. Also, for these profiles for $\mathcal{P}_{\mathcal{R},\text{small}}(k)$, the amplitude falls off at small scales. Thus, it is a good approximation to take the behaviour of $\mathcal{P}_{\mathcal{R}}(k)$ to be given by

$$\mathcal{P}_{\mathcal{R}}(k) \simeq \begin{cases} \mathcal{P}_{\mathcal{R},\text{large}}(k), & k \lesssim 1 \text{ Mpc}^{-1} , \\ \mathcal{P}_{\mathcal{R},\text{small}}(k), & k \gtrsim 1 \text{ Mpc}^{-1} . \end{cases} \quad (3.2)$$

Since we will be concerned only with $\mathcal{P}_{\mathcal{R},\text{small}}(k)$ from now, we will drop the subscript "small" from now. In our subsequent expressions, $\mathcal{P}_{\mathcal{R}}(k)$ should be understood to mean $\mathcal{P}_{\mathcal{R},\text{small}}(k)$, and the expressions should be understood to be valid only for small scales $k \gtrsim 1 \text{ Mpc}^{-1}$.

Since we have only a single observable quantity that can be used to constrain the power spectrum, we cannot derive proper constraints if we choose a highly tunable parametrization of the spectrum, e.g. a series expansion where we want to constrain each coefficient. The most useful parametrization for which we can derive proper constraints is when we vary the amplitude at the peak of the power spectrum and the wavenumber k_p at which this peak occurs, keeping the shape of the spectrum fixed.

For simplicity, we have so far only worked with spectral shapes that have a single peak, but this restriction is not necessary. Our parametrization is expressed as

$$\mathcal{P}_{\mathcal{R}}(k) = \mathcal{P}_{\mathcal{R}}(k_p) f\left(\frac{k}{k_p}\right). \quad (3.3)$$

By definition of k_p , we have

$$\begin{cases} f\left(\frac{k}{k_p}\right) = 1, & k = k_p, \\ f\left(\frac{k}{k_p}\right) < 1, & k \neq k_p. \end{cases} \quad (3.4)$$

For convenience, we denote

$$A \equiv \mathcal{P}_{\mathcal{R}}(k_p). \quad (3.5)$$

Thus, we work with a power spectrum of the form

$$\mathcal{P}_{\mathcal{R}}(k) = \mathcal{P}_{\mathcal{R}}(k, A, k_p) = A f\left(\frac{k}{k_p}\right). \quad (3.6)$$

Note that sometimes a slightly different normalization for $f(k/k_p)$ is used in the literature:

$$\int_{-\infty}^{\infty} f\left(\frac{k}{k_p}\right) d \ln k = 1. \quad (3.7)$$

We have used this parametrization for the lognormal spectrum in the next section because it conforms with literature and makes it easy to compare our plots with those in the literature. However, we have not used this parametrization for other power spectra because it is in general difficult to perform the integral eq. (3.7) analytically.

3.1.2 Constraints from the CMB μ distortion

The value of the chemical potential μ in the CMB is given by eq. (2.5). For our parametrization of the power spectrum, we can write

$$\mu = \mu(A, k_p) \simeq \int_{1 \text{ Mpc}^{-1}}^{\infty} dk Af\left(\frac{k}{k_p}\right) W(k). \quad (3.8)$$

Note that we can take A out of the integral, so $\mu(A, k_p)$ is simply proportional to A .

$$\mu(A, k_p) \propto A. \quad (3.9)$$

From the FIRAS observations we have the following upper bound:

$$\mu(A, k_p) \leq 9 \times 10^{-5}. \quad (3.10)$$

Using this inequality, we can generate constraints on the A vs k_p plane. The constraint will be in the form of a curve $A_{max}(k_p)$. This means that for each value of the peak wavenumber k_p , there is a peak amplitude $A_{max}(k_p)$ at which the upper bound eq. (3.10) is saturated, and the peak amplitude needs to be equal to or below $A_{max}(k_p)$ in order to be consistent with observations.

Now, our task is to find the function $A_{max}(k_p)$ from the property that it saturates the upper bound eq. (3.10). This is made easy by eq. (3.9), the fact that $\mu(A, k_p)$ simply depends on A by a linear proportionality. Thus, we can evaluate $\mu(A, k_p)$ at any arbitrary value of A , and then using the linear proportionality we obtain $A_{max}(k_p)$ through

$$A_{max}(k_p) = A \left(\frac{9 \times 10^{-5}}{\mu(A, k_p)} \right), \quad (3.11)$$

for any value of A . The curve $A_{max}(k_p)$ is what is shown on power spectrum constraint plots (fig. 4.1 and fig. 4.2, discussed in the next chapter).

3.1.3 Constraints from f_{PBH} bounds

The observational bounds on f_{PBH} for various mass ranges can be collected and represented as a function $f_{PBH,max}(M)$. In the minimal scenario where we do not consider any astrophysical constraints on f_{PBH} (citing the astrophysical uncertainties which are involved in most of them), we have $f_{PBH,max}(M) = 1$. Otherwise, is smaller than unity where astrophysical constraints exist.

The remaining uncertainties are those in the formation mechanism of PBHs, namely the uncertainty in the exact value of δ_c and the uncertainty in the window function eq. (2.12). However, once we look at the constraints from PBHs, we will find both the above kinds of uncertainties do not drastically change the power spectrum constraints from f_{PBH} bounds.

To obtain constraints from f_{PBH} bounds, we first start by mapping a PBH mass to a scale k_p . When a scale k enters the Hubble radius ($k = aH$), if the amplitude $\mathcal{P}_{\mathcal{R}}(k)$ is large, a region of size $r = aH$ can collapse into a PBH. Thus, we can map a peak wavenumber k_p to a PBH mass M using eq. (2.16) and $k_p = 1/R$. For default values of the parameters in eq. (2.16), this gives

$$k_p = 1.59 \times 10^6 \left(\frac{M}{M_\odot} \right)^{-1/2} \text{Mpc}^{-1}. \quad (3.12)$$

Then, after choosing a peak scale k_p , we numerically invert the function $f_{PBH}(A, k_p)$ to obtain the value of A_{max} such that

$$f_{PBH}(A, k_p) = f_{PBH,max}(M(k_p)). \quad (3.13)$$

By repeating this for different scales k_p , we construct the A_{max} vs k_p curve which is the constraint plot. Due to the restrictions and approximations in our theoretical modeling, our power spectrum constraints from f_{PBH} bounds are valid for the range $0.01 \text{ Mpc}^{-1} \lesssim k \lesssim 10^{16} \text{ Mpc}^{-1}$. This is because scales $k \leq 0.01 \text{ Mpc}^{-1}$ enter the Hubble radius after matter-radiation equality, so we cannot assume a radiation background as we have done in our model. Moreover, the power spectrum is already constrained to be small on these large scales, so constraints from f_{PBH} on these scales will not reveal anything new. For small scales $k \gtrsim 10^{16} \text{ Mpc}^{-1}$, from eq. (3.12) we see that the PBHs will be so light that they will evaporate before the present day, so we cannot use f_{PBH} bounds to constrain the power spectrum on these scales.

We will now assess the impact of the uncertainty in δ_c on the constraints. We have considered the f_{PBH} bounds from [12].

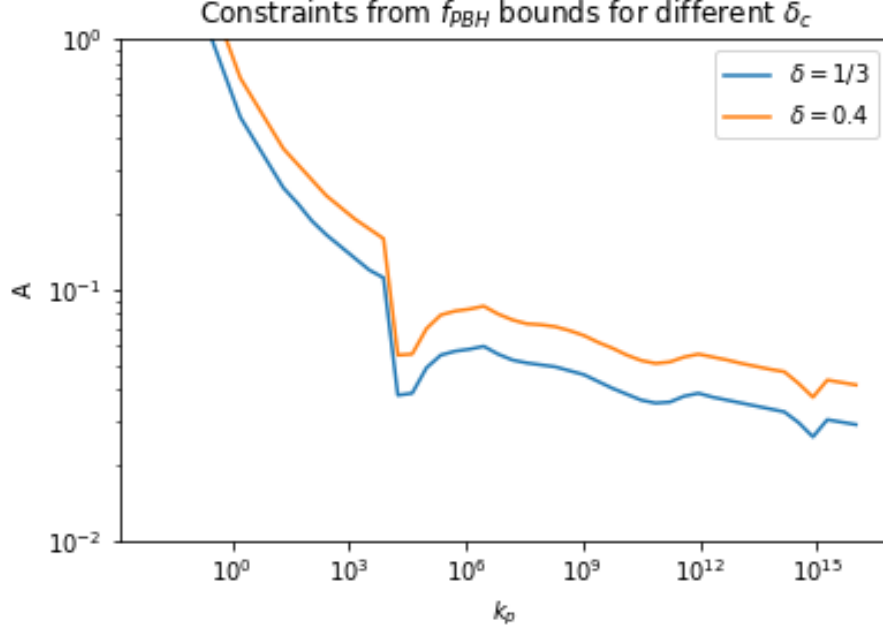


Figure 3.1: Constraints from f_{PBH} bounds on the same power spectrum with $\delta_c = 1/3$ and 0.4. The power spectrum is taken to be a lognormal power spectrum eq. (4.1) with $\sigma = 0.5$. A on the y-axis is defined such that $A/(\sqrt{2\pi}\sigma)$ is the amplitude of the power spectrum at the peak k_p .

We find that the two constraint plots do not differ by even a single order of magnitude. This is a consequence of the very high sensitivity of f_{PBH} to the overall constant A multiplying the power spectrum. As we should expect, the constraint plot for the smaller value of δ_c is slightly below the one with the larger value. This is because a smaller δ_c allows primordial perturbations with a smaller amplitude to form PBHs, which can be seen quantitatively from eq. (2.9) and eq. (2.18). In the next chapter, we choose to show constraints for $\delta_c = 1/3$.

3.1.4 Constraints from Pulsar Timing Arrays

When deriving constraints from GW observatories, we assume that the GW signal is entirely due to SGWs. The observational upper bound for PTAs is taken to be eq. (2.34). Due to our parametrization

eq. (3.6) of the power spectrum, we have

$$\Omega_{GW}(f, A, k_p)h^2 \leq 6.4 \times 10^{-10} . \quad (3.14)$$

From eq. (2.31) and eq. (2.32), we obtain

$$\Omega_{GW}(f, A, k_p)h^2 \propto A^2 . \quad (3.15)$$

We again derive constraints in the form of a function $A_{max}(k_p)$, where A_{max} is the value of A which satisfies eq. (3.14).

Using eq. (3.14) and eq. (3.15) we construct the constraint curve $A_{max}(k_p)$:

$$A_{max}(k_p) = A \left(\frac{6.4 \times 10^{-10}}{\Omega_{GW}(1\text{yr}^{-1}, A, k_p)h^2} \right)^{1/2} , \quad (3.16)$$

for any value of A .

3.1.5 Constraints from GW interferometers

For GW interferometers our observational upper bound comes from eq. (2.40), with our convention $T = 1 \text{ yr}^{-1}$ and our imposed threshold being $\rho < 1$.

From the power spectrum parametrization eq. (3.6) we have

$$\rho = \rho(A, k_p) , \quad (3.17)$$

and from substituting the parametrization in the expression eq. (2.40) we have

$$\rho(A, k_p) \propto A^2 . \quad (3.18)$$

Again, our constraint plot is the curve $A_{max}(k_p)$, where A_{max} is the value of A that gives $\rho = 1$.

We construct $A_{max}(k_p)$ using the above results as:

$$A_{max}(k_p) = A \left(\frac{1}{\rho(A, k_p)} \right)^{1/2}, \quad (3.19)$$

for any value of A .

In principle, we can apply this analysis to any interferometer system where there are more than one interferometer configurations that can be correlated. However, in practice this is not useful for current LIGO-like interferometers, since their sensitivity is too low to provide meaningful constraints on the GW background. Concept interferometers envisioned to operate in the future have the required sensitivity. We have considered the Laser Interferometer Space Antenna (LISA) [18] and the Big Bang Observer (BBO) [19] as examples.

Since LISA and BBO are future experiments, the interpretation of these constraints is different from the constraints from μ distortions and PTAs. Now the interpretation is that for a given k_p , only values of A equal to or above A_{max} will be detectable by LISA or BBO (subject to our imposed values on the threshold on ρ and the observing time T). Or, in other words, only values of A below A_{max} can be ruled out in the absence of a detection.

3.2 Constraining an inflationary potential

In this section, we consider the case when we have an inflaton potential that gives rise to the power spectrum. As an example, we consider the USR-1 potential introduced in eq. (1.98):

$$V(\phi) = V_0 \frac{3x^4 - 4\alpha x^3 + 6x^2}{(1 + \beta x^2)^2}, \quad (3.20)$$

$$\text{where } x \equiv \phi/v. \quad (3.21)$$

The parameters of the potential are V_0 , α , β , and v . We do not have enough observational information to constrain all of them simultaneously. We can proceed by doing something very analogous to how we parametrized the power spectrum.

Analogous to how we varied $A \equiv \mathcal{P}_{\mathcal{R}}(k_p)$ in the power spectrum, we choose to vary V_0 in the

inflaton potential. The mathematical meaning of both these parameters is identical—they are an overall constant multiplying the function. In fact, V_0 and A are proportional to each other if all other parameters are fixed.

The other parameter that we choose to vary is not explicit in eq. (3.20). However, it is still a quantity that holds important information about the inflationary era. We are referring to the number of e-folds of inflation after a chosen scale k_{pivot} exits the Hubble radius during inflation. This quantity, often denoted N_{pivot} or N_* in the literature (we will stick with the former), is an important quantity for CMB observations. The values of the scalar spectral index n_s and the tensor-to-scalar ratio r at any particular k_{pivot} depend on the value of N_{pivot} . This is because varying N_{pivot} does not change the shape of the inflaton potential at all, it only changes the range of scales that exited the Hubble radius during inflation. Increasing N_{pivot} moves the Hubble crossing of each scale during inflation to an earlier time, and thus it does not change the shape of the power spectrum but just increases the scale at which it peaks. We thus conclude that the parameters V_0 and N_{pivot} are in direct correspondence with A and k_p in the power spectrum respectively, and so we can use the same methodology that we used to constrain the power spectrum to constrain V_0 and N_{pivot} .

There are two more important considerations that apply when we are trying to constrain V_0 and N_{pivot} . These arise from the need to respect existing observational constraints from the CMB data. The first is that CMB observations constrain N_{pivot} to lie roughly between 50 to 60. Thus, when we are constraining the power spectrum, we want constraints over all k_{peak} values, but when we are instead constraining the inflaton potential and pivot scale e-folds, we only need constraints over the range $50 \lesssim N_{pivot} \lesssim 60$.

The second point is that in the simple scenario we are considering, changing the value of V_0 changes the power spectrum on all scales, including the large scales probed by CMB anisotropies. But we already have very strong constraints on the amplitude as well as the slope of the power spectrum on these large scales. Thus, if we assume our inflationary model to describe both the largest and smallest scales, then it is not very useful to constrain V_0 from small scale observables because the strongest constraint on V_0 by far comes from the CMB anisotropies probed on large scales. In this thesis, we have not attempted to combine constraints from small scale probes with constraints from CMB anisotropies.

Chapter 4

Results and Discussion

4.1 Constraining a lognormal power spectrum

A lognormal power spectrum is a useful parametrization of a power spectrum with a single peak. It is given by

$$\mathcal{P}_{\mathcal{R}}(k) = \frac{A}{\sqrt{2\pi}\sigma} \exp \left[\frac{-\ln(k/k_{peak})^2}{2\sigma^2} \right]. \quad (4.1)$$

The parameter σ determines the width of the peak of the power spectrum, increasing σ increases the width.

Below, we have shown the constraints obtained for the lognormal spectrum. We have considered three values of σ in order to see how the constraints depend on the width of the spectrum. Previous work on constraining lognormal spectra on small scales has been done by [20], but they have not considered constraints from f_{PBH} bounds. We have developed upon their results by considering constraints from f_{PBH} bounds and implications for PBH formation.

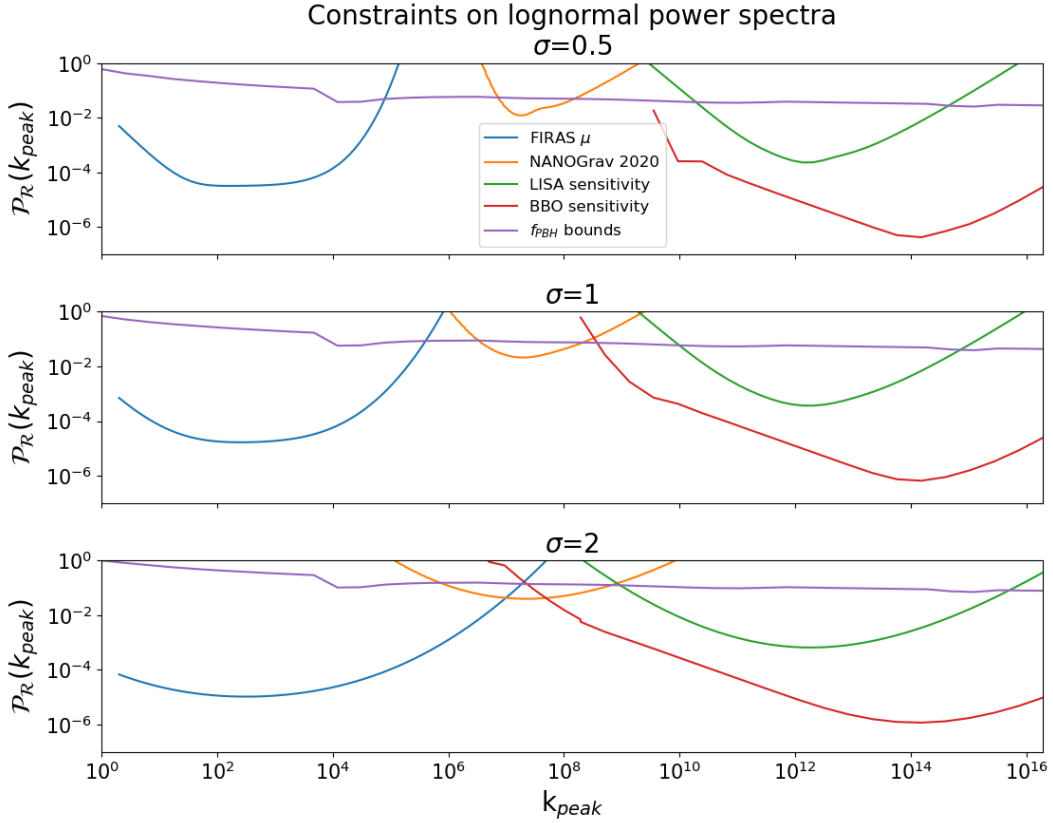


Figure 4.1: Constraints on the lognormal spectrum for three different values of σ . We have not shown constraints below $k = 1 \text{ Mpc}^{-1}$ as larger scales are constrained by CMB anisotropies. Also, we have not shown constraints above $k \sim 10^{16} \text{ Mpc}^{-1}$ because of a numerical difficulty—the time required to calculate the overlap reduction function from eq. (2.39) for the BBO increases exponentially due to the oscillatory nature of the integral involved, and we could not find a satisfactory analytic approximation for the function at large frequencies.

These plots indicate that constraints on the power spectrum depend significantly on its width. Recall fig. 2.1 at the beginning of Chapter 2. In that figure, only CMB constraints come from measurements rather than upper bounds. These are free-form constraints, so they do not depend on the shape of the spectrum as they are constraints on the shape of the spectrum itself in a sense. However, the other constraints, such as the μ distortion, PBHs, and SGWs do depend on the spectrum’s shape. These are constraints on quantities that are obtained by integrating over the power spectrum over corresponding ranges of scales. Integrals of the power spectrum cannot constrain its shape to an

appreciable extent, thus we should expect these constraints to be dependent on the spectrum’s shape.

The length scales corresponding to small scale constraints in fig. 2.1 come from constraint plots in the literature that are derived for fixed shapes of the power spectrum. Specifically, constraints from μ distortions assume a power spectrum that grows as k^4 and vanishes after a cutoff [13], and SGW constraints assume a δ function power spectrum. However, fig. 4.1 indicates that the length scales probed by any observable depends on the width of the peak of the power spectrum, and so the scales shown in fig. 2.1 are not universal, they depend on the width of the power spectrum being considered.

An important implication of the dependence of power spectrum constraints on its width is on PBHs. As seen in fig. 4.1, for $\sigma = 0.5$ there is a window of scales between $k \sim 10^5 - 10^7 \text{ Mpc}^{-1}$ where the power spectrum can have a peak and form appreciable PBHs. However, this window completely disappears for $\sigma = 2$. This indicates that it is harder to form appreciable PBHs while remaining consistent with observations. fig. 4.1 also indicates that the reason for this does not have to do with the physics of PBHs (as the power spectrum constraints from f_{PBH} don’t change substantially with the shape of the spectrum), but it has to do with the physics of the other small scale probes of the power spectrum: μ distortions and SGWs (as the constraints from these probes depend substantially on the width of the spectrum).

4.2 Constraining a broken power law power spectrum

A broken power law is another parametrization for a peaked power spectrum. It has three parameters, one more than the lognormal. These parameters are a constant multiplier, and the slopes on each side of the peak. Thus, like the lognormal, it allows control over the amplitude of the peak and the peak width, and unlike the lognormal it also allows control of how one-sided the peak is. The broken power law spectrum is given by [21]

$$\mathcal{P}_{\mathcal{R}}(k) = \frac{A_s(\alpha + \beta)}{\beta(k/k_p)^\alpha + \alpha(k/k_p)^\beta}. \quad (4.2)$$

Here, α is the logarithmic slope of the spectrum on large scales $k < k_p$, and β is the slope on small scales $k > k_p$.

The broken power law parametrization was introduced in [21], and [22] constrained it on small

CHAPTER 4. RESULTS AND DISCUSSION

scales from f_{PBH} bounds and PTA observations. We have developed upon their results by considering constraints from the μ distortion and GW interferometers.

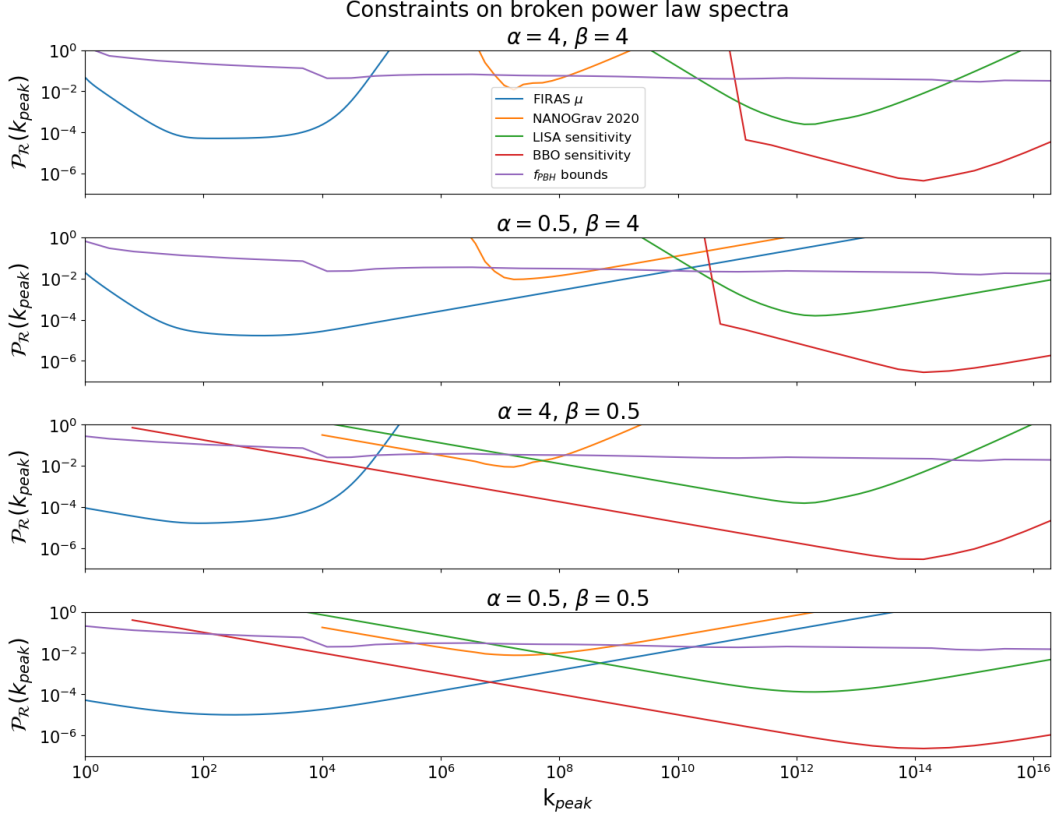


Figure 4.2: Constraints on the broken power law spectrum. The values 0.5 and 4 for α and β are chosen to conform with the values chosen in [22].

fig. 4.2 indicates that a smaller value of the slope at large scales α extends constraints from all probes towards smaller scales, and a smaller value of the slope at small scales β extends all constraints towards larger scales. This might seem counterintuitive at first; it seems like changing the small scale behaviour of the spectrum affects its observability on small scales instead of large scales. To understand this, we should recognize that the observational constraints from any observable depend on the behaviour of the power spectrum over the range of scales which the observable is sensitive to. Thus, when the peak of the power spectrum is at a smaller scale than the scales which the observable is sensitive to, the scales probed by the observable lie in the large-scale (i.e. $k < k_p$) part

of the spectrum. Thus, it is the small-scale part of the spectrum which determines the observational constraints. The same explanation holds for why the small-scale (i.e. $k > k_p$) slope of the spectrum extends observational constraints towards larger scales.

4.3 Constraining the USR-1 potential

We are constraining the USR-1 potential eq. (1.98):

$$V(\phi) = V_0 \frac{3x^4 - 4\alpha x^3 + 6x^2}{(1 + \beta x^2)^2}, \tag{4.3}$$

where $x = \phi/v$. The parameters that we fix are $v = \sqrt{0.108}M_{\text{Pl}}$, $\alpha = 1$, and $\beta = 1.4349$. The parameters that we choose to vary are V_0 and N_{pivot} , as discussed in the previous section. We have shown our constraints on the USR-1 model below.

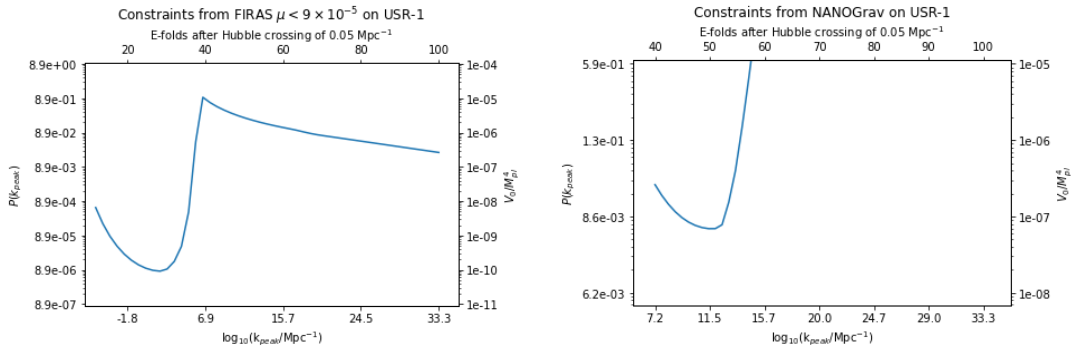


Figure 4.3: Constraints on the USR-1 model from the FIRAS upper bound on μ and the NANOGrav detection of $\Omega_{\text{GW}}h^2$.

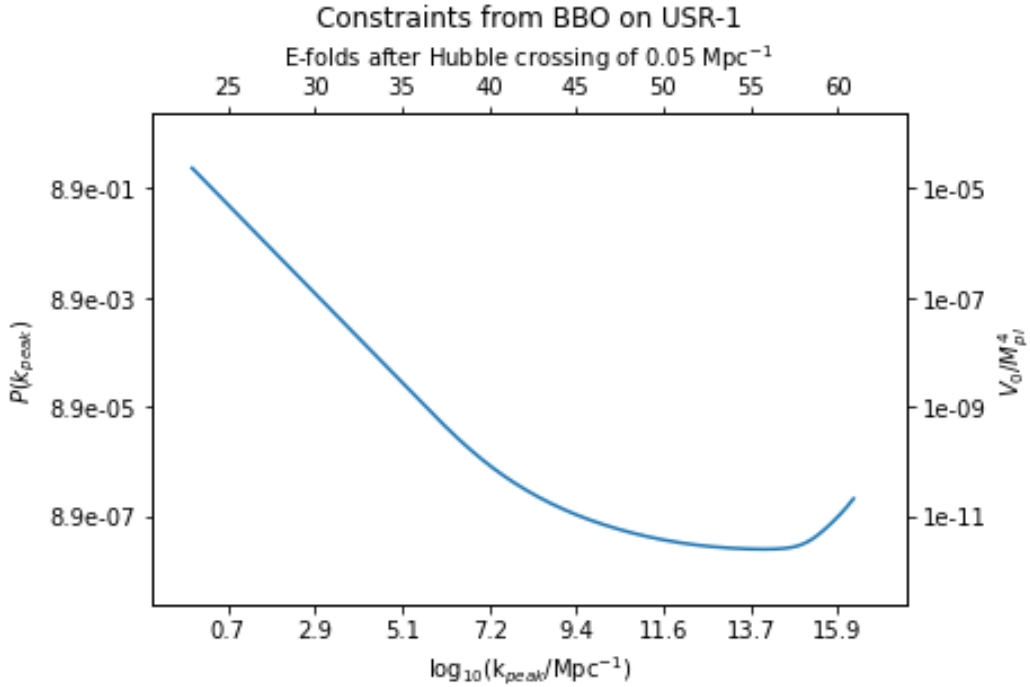


Figure 4.4: Constraints on BBO on the power spectrum for the USR-1 model.

There are two points to note about the above plots:

1. We have not shown constraints stating that the power spectrum is bounded above by a value greater than unity, since when the power spectrum is above unity we cannot rely on linear perturbation theory that has been assumed to be valid in all our theoretical calculations.
2. When we are looking to constrain the inflationary potential rather than its resulting power spectrum, the peak wavenumber k_{peak} is determined by the number of e-folds after pivot scale crossing. Since the number of e-folds is constrained to be roughly between 50 to 60 from CMB observations, we only need constraints on the power spectrum over the corresponding range of k_{peak} values.

Chapter 5

Conclusions and outlook

The objectives achieved through the work in this thesis are:

1. An understanding of the phenomenology on various ranges of small scales that are sensitive to primordial scalar perturbations.
2. A demonstration of how to derive constraints on the primordial perturbations on small scales while still staying in the realm of perturbation theory.
3. An understanding of the degree to, and the manner in which constraints on the power spectrum depend on its assumed spectral shape.
4. An understanding of the implications of the dependence of power spectrum constraints on its shape for PBH formation.
5. An understanding of how to interpret constraints on primordial perturbations as constraints on the underlying model of inflation.
6. A simple proof of the decomposition theorem which is a key result of cosmological perturbation theory.

Appendix A

A simple proof of the decomposition theorem

In this section we will discuss linear perturbation theory around a background FLRW cosmology.

In the FLRW background metric, the metric perturbations can be decomposed according to their behaviour under local spatial rotations on an equal time hypersurface. This property leads to the classification of perturbations as scalars, vectors, and tensors under the group of rotations in 3 dimensions. Scalar perturbations are invariant under rotations. Vector and tensor perturbations, as their names indicate, transform as vectors and tensors respectively.

The decomposition theorem: According to the decomposition theorem in general relativity, which we will prove in this section, at linear order the scalar, vector and tensor perturbations are decoupled from each other, i.e. each type of metric perturbation is only affected by the same type of source, namely scalar, vector and tensor perturbations in the stress energy tensor. This allows us to analyze scalar, vector and tensor perturbations separately.

A.1 Defining the metric and matter perturbations

We first obtain the most general linear perturbations to the flat background FLRW metric eq. (1.2) and the background perfect fluid stress energy tensor. We are working in Cartesian coordinates.

The general linear perturbation to this metric is a symmetric rank 2 tensor in 4 dimensions. Thus,

APPENDIX A. A SIMPLE PROOF OF THE DECOMPOSITION THEOREM

it has 10 components or 10 degrees of freedom. The SVT decomposition distributes these degrees of freedom among quantities that transform as scalars, vectors and rank 2 tensors under rotations in the 3 spatial dimensions. The line element of the perturbed metric is

$$\begin{aligned}
 ds^2 = & (1 + 2A)dt^2 \\
 & -2a(t)(B_{,i} + S_i)dtdx^i \\
 & -a^2(t)[(1 - 2\psi)\delta_{ij} + 2E_{,ij} \\
 & + F_{i,j} + F_{j,i} + h_{ij}]dx^i dx^j .
 \end{aligned} \tag{A.1}$$

This allows us to compute the linear perturbations to the background Einstein tensor eq. (1.3). They are found to be

$$\delta G_0^0 = -6H(\dot{\psi} + HA) + \frac{2}{a^2}\nabla^2[\psi - aH(B - a\dot{E})] , \tag{A.2a}$$

$$\delta G_i^0 = 2(\dot{\psi} + HA)_{,i} - \frac{1}{2a}\nabla^2 V_i , \tag{A.2b}$$

$$\begin{aligned}
 \delta G_j^i = & -\delta_j^i \left[2\ddot{\psi} + 2H(3\dot{\psi} + \dot{A}) + 2(2\dot{H} + 3H^2)A \right. \\
 & \left. + \frac{1}{a^2}\nabla^2 \left(A - \psi + \frac{1}{a} \left[a^2(B - a\dot{E}) \right]' \right) \right] \\
 & + \frac{1}{a^2} \left(A - \psi + \frac{1}{a} \left[a^2(B - a\dot{E}) \right]' \right)'_{,j}{}^i \\
 & + \frac{1}{2a} [\partial^i (\dot{V}_j + 2HV_j) + \partial_j (\dot{V}^i + 2HV^i)] \\
 & - \frac{1}{2} \left(\ddot{h}_j^i + 3H\dot{h}_j^i - \frac{1}{a^2}\nabla^2 h_{ij} \right) .
 \end{aligned} \tag{A.2c}$$

Now we turn to the stress energy tensor. It turns out that the most general perturbation to the background stress energy tensor allowed in a perturbed flat FLRW metric is not simply a general perturbation to the perfect fluid stress energy tensor eq. (1.4). In order to see this, let us first write the

general perturbation to the perfect fluid eq. (1.4) which has the 4-velocity eq. (1.8):

$$\delta T_0^0 = \delta \rho , \tag{A.3a}$$

$$\delta T_i^0 = (\rho + p)\delta u_i , \tag{A.3b}$$

$$\delta T_j^i = -\delta p \delta_j^i . \tag{A.3c}$$

We can write the 4-velocity perturbation as

$$\delta u_i = v_{,i} + \delta u_i^\perp , \tag{A.4}$$

where $\delta u_{i,i}^\perp = 0$. From these expressions we see that δT_j^μ has 3 scalar degrees of freedom ($\delta \rho$, δp and v), and 2 vector degrees of freedom δu_i^\perp . However, a general symmetric tensor has 4 scalar, 4 vector and 2 tensor degrees of freedom, which means we are missing something in eq. (A.3).

In addition to the perturbation of the perfect fluid eq. (1.4), the spatial components δT_j^i contain extra degrees of freedom in the form of one scalar which has a traceless contribution, one transverse vector and one transverse traceless tensor. These extra degrees of freedom represent anisotropic stress, we denote them as π^S , π_i^V and π_{ij}^T respectively. The general perturbed stress energy tensor is given by

$$\delta T_0^0 = \delta \rho , \tag{A.5a}$$

$$\delta T_i^0 = (\rho + p)\delta u_i , \tag{A.5b}$$

$$\begin{aligned} \delta T_j^i = & -\delta p \delta_j^i - \left(\partial^i \partial_j - \delta_j^i \frac{\nabla^2}{3} \right) \pi^S \\ & - \pi_{,j}^V{}^i - \pi_{,i}^V{}^j - \pi_j^T{}^i . \end{aligned} \tag{A.5c}$$

We can see that it has the correct number of degrees of freedom of each class.

The Einstein equations obeyed by the linear metric and matter perturbations are

$$\delta G_\nu^\mu = \delta T_\nu^\mu . \tag{A.6}$$

A.2 Gauge transformations and gauge invariant quantities

The metric perturbations eq. (A.1) and matter perturbations eq. (A.5) are gauge dependent. We will avoid the issue of fixing a gauge and take the approach of working with gauge invariant quantities. In order to construct gauge invariant quantities out of the perturbations in eq. (A.1), we will first derive how each of the perturbation variables transforms under a gauge transformation. A gauge transformation is defined by

$$\tilde{x}^\mu = x^\mu + \delta x^\mu(x^\alpha). \quad (\text{A.7})$$

We can break up the spatial part of δx^μ into a transverse vector and the gradient of a scalar and write

$$\delta x^\mu = (\delta t, \delta x_{,i} + \delta x_i^\perp), \quad (\text{A.8})$$

where $\delta x_{i,i}^\perp = 0$. We denote all functions in the new coordinates with tildes. \tilde{x}^μ are coordinates of the same point in the new coordinates. Consider a rank 2 tensor field $H_{\mu\nu}(x^\alpha)$, which is split into a background and perturbation as

$$H_{\mu\nu}^{\text{tot}}(x^\alpha) = H_{\mu\nu}(x^\alpha) + \delta H_{\mu\nu}(x^\alpha). \quad (\text{A.9})$$

We write

$$(H_{\mu\nu} + \delta H_{\mu\nu})(x^\alpha) = (H_{\mu\nu} + \delta H_{\mu\nu})(x^\alpha(\tilde{x}^\alpha)) \quad (\text{A.10a})$$

$$= \tilde{x}_{,\mu}^\rho \tilde{x}_{,\nu}^\sigma (\tilde{H}_{\rho\sigma} + \delta \tilde{H}_{\rho\sigma})(\tilde{x}^\alpha) \quad (\text{A.10b})$$

$$= (\delta_\mu^\rho + \delta x_{,\mu}^\rho) (\delta_\nu^\sigma + \delta x_{,\nu}^\sigma) (\tilde{H}_{\rho\sigma} + \delta \tilde{H}_{\rho\sigma})(\tilde{x}^\alpha) \quad (\text{A.10c})$$

$$= \tilde{H}_{\mu\nu}(\tilde{x}^\alpha) + \delta x_{,\mu}^\rho \tilde{H}_{\rho\nu}(\tilde{x}^\alpha) + \delta x_{,\nu}^\sigma \tilde{H}_{\mu\sigma}(\tilde{x}^\alpha) + \delta \tilde{H}_{\mu\nu}(\tilde{x}^\alpha) \quad (\text{A.10d})$$

$$= \tilde{H}_{\mu\nu}(x^\alpha) + \delta x^\beta \tilde{H}_{\mu\nu,\beta}(\tilde{x}^\alpha) + \delta x_{,\mu}^\rho \tilde{H}_{\rho\nu}(\tilde{x}^\alpha) + \delta x_{,\nu}^\sigma \tilde{H}_{\mu\sigma}(\tilde{x}^\alpha) + \delta \tilde{H}_{\mu\nu}(\tilde{x}^\alpha). \quad (\text{A.10e})$$

We have neglected terms beyond linear order in the perturbations above. We define the background

part of the tensor to be the same in both coordinate systems, i.e.

$$\tilde{H}_{\mu\nu} = H_{\mu\nu} . \quad (\text{A.11})$$

Thus, we get the gauge transformation law

$$\delta H_{\mu\nu} = \delta \tilde{H}_{\mu\nu} + \tilde{H}_{\mu\nu,\alpha} \delta x^\alpha + \delta x^\rho_{,\mu} \tilde{H}_{\rho\nu} + \delta x^\sigma_{,\nu} \tilde{H}_{\mu\sigma} . \quad (\text{A.12})$$

Equivalently,

$$\delta \tilde{H}_{\mu\nu} = \delta H_{\mu\nu} - H_{\mu\nu,\alpha} \delta x^\alpha - \delta x^\rho_{,\mu} H_{\rho\nu} - \delta x^\sigma_{,\nu} H_{\mu\sigma} . \quad (\text{A.13})$$

On applying this transformation law to the perturbed metric $\delta g_{\mu\nu}$ from the line element eq. (A.1), we obtain the gauge transformation laws for the metric perturbations.

$$\tilde{A} = A - \dot{\delta}t , \quad (\text{A.14})$$

$$\tilde{B} = B + \frac{\delta t}{a} - a \dot{\delta}x , \quad (\text{A.15})$$

$$\tilde{\psi} = \psi + H \delta t , \quad (\text{A.16})$$

$$\tilde{E} = E - \delta x , \quad (\text{A.17})$$

$$\tilde{S}_i = S_i - a \dot{\delta}x_i^\perp , \quad (\text{A.18})$$

$$\tilde{F}_i = F_i - \delta x_i^\perp , \quad (\text{A.19})$$

$$\tilde{h}_{ij} = h_{ij} . \quad (\text{A.20})$$

We find that the scalar and vector perturbations are affected only by the scalar and vector degrees of freedom in the gauge transformations respectively. The tensor perturbations are gauge invariant. Since there are 4 scalar and 4 vector degrees of freedom in the perturbed metric eq. (A.1) and 2 scalar and 2 vector degrees of freedom in the gauge choice eq. (A.8), we should have 2 scalar and 2 vector degrees of freedom that are gauge invariant. Using eq. (A.14) - eq. (A.20), we can construct the gauge

invariant metric perturbations as follows:

$$\Phi \equiv A + [a(B - a\dot{E})] , \quad (\text{A.21})$$

$$\Psi \equiv \psi - [aH(B - a\dot{E})] , \quad (\text{A.22})$$

$$V_i \equiv S_i - a\dot{F}_i . \quad (\text{A.23})$$

where $V_{i,i} = 0$.

The next step is to write the Einstein equation in terms of gauge invariant variables. We can derive the gauge transformation law for δG_ν^μ in the same way as we derived the law for a tensor with two lower indices eq. (A.13), the result is

$$\tilde{\delta G}_\nu^\mu = \delta G_\nu^\mu - G_{\nu,\alpha}^\mu \delta x^\alpha - \delta x_{,\nu}^\rho G_\rho^\mu + \delta x_{,\sigma}^\mu G_\nu^\sigma . \quad (\text{A.24})$$

On substituting the expression eq. (A.8) for δx^μ in the above and using the property $G_i^i = G_j^j$ (i and j not summed over), we get

$$\tilde{\delta G}_0^0 = \delta G_0^0 - \dot{G}_0^0 \delta t , \quad (\text{A.25a})$$

$$\tilde{\delta G}_i^0 = \delta G_i^0 - \left(G_0^0 - \frac{G_k^k}{3} \right) \delta t_{,i} , \quad (\text{A.25b})$$

$$\tilde{\delta G}_j^i = \delta G_j^i - \dot{G}_j^i \delta t . \quad (\text{A.25c})$$

Using these transformation laws, we can construct the gauge invariant version of the Einstein tensor:

$$\delta \mathcal{G}_0^0 = \delta G_0^0 + \dot{G}_0^0 [a(B - a\dot{E})] , \quad (\text{A.26a})$$

$$\delta \mathcal{G}_i^0 = \delta G_i^0 + \left(G_0^0 - \frac{G_j^j}{3} \right) [a(B - a\dot{E})]_{,i} , \quad (\text{A.26b})$$

$$\delta \mathcal{G}_j^i = \delta G_j^i + \dot{G}_j^i [a(B - a\dot{E})] . \quad (\text{A.26c})$$

On substituting the gauge invariant variables eq. (A.21)-eq. (A.23) and the perturbed Einstein tensor

eq. (A.2), we get

$$\delta\mathcal{G}_0^0 = -6H\dot{\Psi} - 6H^2\Phi + \frac{2}{a^2}\nabla^2\Psi, \quad (\text{A.27a})$$

$$\delta\mathcal{G}_i^0 = 2\dot{\Psi}_{,i} + 2H\Phi_{,i} - \frac{1}{2a}\nabla^2 V_i, \quad (\text{A.27b})$$

$$\begin{aligned} \delta\mathcal{G}_j^i = & -\delta_j^i \left[2\ddot{\Psi} + 2H(3\dot{\Psi} + \dot{\Phi}) + 2(2\dot{H} + 3H^2)\Phi \right. \\ & \left. + \frac{1}{a^2}\nabla^2(\Phi - \Psi) \right] + \frac{1}{a^2}\partial^i\partial_j(\Phi - \Psi) \\ & + \frac{1}{2a}[(\dot{V}_j + 2HV_{j,i}) + (\dot{V}^i + 2HV^i)_{,j}] \\ & - \frac{1}{2} \left(\ddot{h}_j^i + 3H\dot{h}_j^i - \frac{1}{a^2}\nabla^2 h_j^i \right). \end{aligned} \quad (\text{A.27c})$$

Now we turn to the stress energy tensor. Owing to the Einstein equations, δT_ν^μ has the same gauge transformation law eq. (A.25) as δG_ν^μ , and the same operations that make δG_ν^μ gauge invariant must also make δT_ν^μ also gauge invariant. i.e. the gauge invariant stress energy tensor is given by

$$\delta\mathcal{T}_0^0 = \delta T_0^0 + \dot{T}_0^0[a(B - a\dot{E})], \quad (\text{A.28})$$

$$\delta\mathcal{T}_i^0 = \delta T_i^0 + \left(T_0^0 - \frac{T_j^j}{3} \right) [a(B - a\dot{E})]_{,i}, \quad (\text{A.29})$$

$$\delta\mathcal{T}_j^i = \delta T_j^i + \dot{T}_j^i[a(B - a\dot{E})]. \quad (\text{A.30})$$

Using this, we obtain the gauge invariant density, pressure and velocity perturbations, denoted with

an overbar:

$$\bar{\delta}\rho = \delta\rho + \dot{\rho}[a(B - a\dot{E})], \quad (\text{A.31})$$

$$\bar{\delta}p = \delta p + \dot{p}[a(B - a\dot{E})], \quad (\text{A.32})$$

$$\bar{v} = v + a(B - a\dot{E}), \quad (\text{A.33})$$

$$\delta\bar{u}_i^\perp = \delta u_i^\perp, \quad (\text{A.34})$$

$$\bar{\pi}^S = \pi^S, \quad (\text{A.35})$$

$$\bar{\pi}_i^V = \pi_i^V, \quad (\text{A.36})$$

$$\bar{\pi}_j^T{}^i = \pi_j^T{}^i. \quad (\text{A.37})$$

We see that δu_i^\perp and the anisotropic stresses gauge invariant. Thus, the gauge invariant perturbed stress energy tensor looks like

$$\delta\mathcal{T}_0^0 = \bar{\delta}\rho, \quad (\text{A.38a})$$

$$\delta\mathcal{T}_i^0 = (\rho + p)\delta\bar{u}_i, \quad (\text{A.38b})$$

$$\begin{aligned} \delta\mathcal{T}_j^i = & -\bar{\delta}p\delta_j^i - \left(\partial^i\partial_j - \delta_j^i\frac{\nabla^2}{3} \right) \pi^S \\ & -\pi_j^V{}^i - \pi_j^V{}^{,i} - \pi_j^T{}^i{}^i. \end{aligned} \quad (\text{A.38c})$$

Finally, the gauge invariant Einstein equations are

$$\delta\mathcal{G}_\nu^\mu = \delta\mathcal{T}_\nu^\mu. \quad (\text{A.39})$$

For clarity, we set $8\pi G = 1$ in this section. In terms of the metric and stress energy perturbations,

$$-6H\dot{\Psi} - 6H^2\Phi + \frac{2}{a^2}\nabla^2\Psi = \bar{\delta}\rho, \quad (\text{A.40a})$$

$$2\dot{\Psi}_{,i} + 2H\Phi_{,i} - \frac{1}{2a}V_i = (\rho + p)\bar{\delta}u_i, \quad (\text{A.40b})$$

$$\begin{aligned} & -\delta_j^i \left[2\ddot{\Psi} + 2H(3\dot{\Psi} + \dot{\Phi}) + 2(2\dot{H} + 3H^2)\Phi \right. \\ & \quad \left. + \frac{1}{a^2}\nabla^2(\Phi - \Psi) \right] + \frac{1}{a^2}\partial^i\partial_j(\Phi - \Psi) \\ & \quad + \frac{1}{2a}[(\dot{V}_j + 2HV_{j,i}) + (\dot{V}^i + 2HV^i)_{,j}] \\ & \quad - \frac{1}{2} \left(\ddot{h}_j^i + 3H\dot{h}_j^i - \frac{1}{a^2}\nabla^2 h_j^i \right) \end{aligned} \quad (\text{A.40c})$$

$$= \bar{\delta}p\delta_j^i - \left(\partial^i\partial_j - \delta_j^i\frac{\nabla^2}{3} \right) \pi^S - \pi_{,j}^V{}^i - \pi_{,i}^V{}^j - \pi_j^T{}^i. \quad (\text{A.40d})$$

A.3 Proving the decomposition theorem

Having obtained the gauge invariant Einstein equations, we can start proving the decomposition theorem. First, we will prove the decomposition theorem for scalars. The idea is to separate out the scalar terms in the Einstein equations by using the properties $V_{,i}^i = 0$, $h_{j,j}^i = 0$ and $h_i^i = 0$.

The 00 component eq. (A.40a) is already entirely in terms of scalars. The 0*i* components eq. (A.40b) contain scalars and divergenceless vectors. We consider the three 0*i* equations as a vector equation and take its divergence to obtain

$$\delta\mathcal{G}_{i,i}^0 = \delta\mathcal{T}_{i,i}^0, \quad (\text{A.41})$$

which in terms of the perturbations is

$$2\nabla^2(\dot{\Psi} + H\Phi) = (\rho + p)\nabla^2 v. \quad (\text{A.42})$$

The *ij* components eq. (A.40d) contain scalars, divergenceless vectors and transverse traceless tensors. We can consider the *ij* components as a single rank 2 tensor equation. On taking its trace we

obtain

$$\delta\mathcal{G}_i^i = \delta\mathcal{T}_i^i, \quad (\text{A.43})$$

which in terms of the perturbations is

$$\begin{aligned} & -3 \left[2\ddot{\Psi} + 2H(3\dot{\Psi} + \dot{\Phi}) + 2(2\dot{H} + 3H^2)\Phi \right. \\ & \left. + \frac{1}{a^2}\nabla^2(\Phi - \Psi) \right] + \frac{1}{a^2}\nabla^2(\Phi - \Psi) = -3\bar{\delta}p. \end{aligned} \quad (\text{A.44})$$

It seems like we have used all the Einstein equations to obtain the equations eq. (A.40a), eq. (A.42) and eq. (A.44). However, the trace of the ij components depends only on the ij components where $i = j$, so we have not utilized the ij components where $i \neq j$ yet. To eliminate the vector and tensor components from these equations, we can simply take the divergence twice.

$$\delta\mathcal{G}_{j,i}^{i,j} = \delta\mathcal{T}_{j,i}^{i,j}, \quad (\text{A.45})$$

which in terms of the perturbations is

$$\frac{1}{a^2}\nabla^2[\nabla^2(\Phi - \Psi)] = -\nabla^2(\nabla^2\pi^S). \quad (\text{A.46})$$

Consider eq. (A.42) and eq. (A.45). They equate the Laplacians of two quantities. This does not imply that the quantities themselves are equal—they can differ by any solution of Laplace’s equation. However, the Fourier transforms of the two quantities are the same (a fact that demonstrates that the Fourier transform is not an invertible operation in general). In Fourier space it is also simpler to invert derivative operations. These facts indicate that it is simpler to work with the Einstein equations in Fourier space than in real space. Hence, we write eq. (A.40a), eq. (A.42), eq. (A.44) and eq. (A.45)

in Fourier space:

$$-6H\dot{\Psi} - 6H^2\Phi - \frac{2}{a^2}k^2\Psi = \delta\bar{\rho}, \quad (\text{A.47a})$$

$$2(\dot{\Psi} + H\Phi) = (\rho + p)v, \quad (\text{A.47b})$$

$$-3 \left[2\ddot{\Psi} + 2H(3\dot{\Psi} + \dot{\Phi}) + 2(2\dot{H} + 3H^2)\Phi \right] + 2\frac{1}{a^2}k^2(\Phi - \Psi) = -3\delta\bar{p}, \quad (\text{A.47c})$$

$$\frac{1}{a^2}(\Phi - \Psi) = -\pi^S. \quad (\text{A.47d})$$

From the last equation above it is immediately apparent that $\Phi \neq \Psi$ if and only if there is a scalar source of anisotropic stress.

Next, we prove the independent evolution of the vector perturbations. For this, consider two equations: the curl of the $0i$ components of the Einstein equations eq. (A.39), and the curl of the divergence of their ij components. In Fourier space these equations are:

$$\epsilon^{ijk}k_j\delta\mathcal{G}_k^0 = \epsilon^{ijk}k_j\delta\mathcal{T}_k^0, \quad (\text{A.48a})$$

$$\epsilon^{ijk}k_j\delta\mathcal{G}_{k,l}^l = \epsilon^{ijk}k_j\delta\mathcal{T}_{k,l}^l. \quad (\text{A.48b})$$

In terms of the metric and stress energy perturbations,

$$-\frac{1}{2a}\epsilon^{ijk}k_j(-k^2V_k) = (\rho + p)\epsilon^{ijk}k_j\delta u_k^\perp, \quad (\text{A.49a})$$

$$-\frac{1}{2a}\epsilon^{ijk}k_j(\dot{V}_k + 2HV_k) = \epsilon^{ijk}k_j\pi_k^V. \quad (\text{A.49b})$$

We can rewrite eq. (A.49a) as

$$\epsilon^{ijk}k_j [k^2V_k - 2a(\rho + p)u_k^\perp] = 0. \quad (\text{A.50})$$

Thus, $[k^2V_k - 2a(\rho + p)u_k^\perp]$ is a vector whose divergence, i.e. inner product with \mathbf{k} is zero (by definition of V_k and δu_k^\perp) and curl, i.e. cross product with \mathbf{k} is also zero. This implies that it is a constant vector. But generally we assume statistical isotropy in cosmology, which is a good assumption as far as we

APPENDIX A. A SIMPLE PROOF OF THE DECOMPOSITION THEOREM

know. This means that there is no preferred direction in Fourier space, so the constant vector must be the zero vector, i.e.

$$V_i = \frac{2a}{k^2}(\rho + p)\delta u_i^\perp . \quad (\text{A.51})$$

It should be kept in mind that this equality is in Fourier space—it does not necessarily imply that V_i and δu_i^\perp are equal in position space unless some boundary conditions are imposed on both of them. But no matter the relation between them in position space, our analysis of the vectors in Fourier space remains valid.

Using the same analysis as above for eq. (A.49b), we obtain

$$\epsilon^{ijk}k_j(\dot{V}_k + 2HV_k + 2a\pi_k^V) = 0 . \quad (\text{A.52})$$

Thus, we have

$$\dot{V}_i + 2HV_i = -2a\pi_i^V . \quad (\text{A.53})$$

It is apparent that the vector perturbation in the metric decays as a^{-2} if there is no vector source of anisotropic stress. This applies to e.g. a perfect fluid as it has $\pi_i^V = 0$.

$$V_i \propto a^{-2} \text{ when } \pi_i^V = 0 . \quad (\text{A.54})$$

Note that this is also a component-wise conclusion- it holds for each individual component i of the vector equation eq. (A.53). On substituting eq. (A.51) in the above, we have a different physical interpretation:

$$(\rho + p)\delta u_i^\perp \propto a^{-3} , \quad (\text{A.55})$$

when there is no vector source of anisotropic stress. This equation can also be written as

$$[(\rho + p)\delta u_i^\perp]' + 3H(\rho + p)\delta u_i^\perp = 0 . \quad (\text{A.56})$$

We now use the definitions for the sound speed and the equation of state of the matter:

$$c_s^2 \equiv \frac{\dot{p}}{\dot{\rho}}, \quad (\text{A.57})$$

$$w \equiv \frac{p}{\rho}. \quad (\text{A.58})$$

Using these in eq. (A.56) we get

$$\begin{aligned} 0 &= (\rho + p)\delta u_i^\perp + (\dot{\rho} + \dot{p})\delta u_i^\perp + 3H(\rho + p)\delta u_i^\perp \\ &= \rho(1 + w)\delta u_i^\perp + \dot{\rho}(1 + c_s^2)\delta u_i^\perp + 3H\rho(1 + w)\delta u_i^\perp. \end{aligned} \quad (\text{A.59})$$

We now divide the equation by $\rho(1 + w)$ and use the relation

$$\frac{\dot{\rho}}{\rho} = -3H(1 + w) \quad (\text{A.60})$$

to get

$$\delta u_i^\perp - 3Hc_s^2\delta u_i^\perp = 0. \quad (\text{A.61})$$

Thus,

$$\delta u_i^\perp \propto a^{3c_s^2}. \quad (\text{A.62})$$

To interpret this dependence of δu_i^\perp on the scale factor, let us take a moment to understand what the 4-velocity components δu_i^\perp actually mean. Firstly, δu_i^\perp is a one-form, but 4-velocity refers to the corresponding vector $\delta u^{i\perp}$, so we actually want to deal with $\delta u^{i\perp}$. In our case, $\delta u^{i\perp} = \delta u_i^\perp$ since we have been using the spatial metric δ_{ij} to raise and lower spatial indices.

The components of 4-velocity are $u^\mu = \frac{dx^\mu}{d\tau}$, where τ is the proper time defined by

$$\tau = \int \sqrt{d\tau^2} = \int \sqrt{ds^2}, \quad (\text{A.63})$$

where the last equality is because we use the $(+, -, -, -)$ metric signature, else we would have $d\tau^2 = -ds^2$. Now, the spatial velocity $\frac{dx^i}{dt}$ is the velocity measured by an observer at rest in the

comoving coordinates. We have

$$\frac{dx^i}{dt} = \frac{dx^i}{d\tau} \frac{d\tau}{dt} = \frac{u^i}{u^0}. \quad (\text{A.64})$$

There is one more subtlety remaining, which arises due to the nature of our coordinates. To see it, let us derive the value of the speed of light, which we can do by setting $d\tau^2 = 0$. Assuming that light travels in the $i = 1$ direction without loss of generality, we have

$$0 = d\tau^2 = dt^2 - a^2(dx^1)^2. \quad (\text{A.65})$$

Thus, the speed of light is

$$\frac{dx^1}{dt} = \frac{1}{a}. \quad (\text{A.66})$$

This dependence of the speed of light on the scale factor is only a result of our choice of coordinates. If we want to match our interpretation of velocity in flat space, we want to have coordinates in which the speed of light is unity. This can be achieved by using conformal time η instead of t as the time coordinate, since

$$\frac{dx^i}{d\eta} = a \frac{dx^i}{dt}. \quad (\text{A.67})$$

For this reason we refer to $v_i \equiv \frac{dx^i}{d\eta}$ as the physical velocity. From eq. (A.62), we have

$$\delta v_i^\perp \propto a^{-1+3c_s^2}. \quad (\text{A.68})$$

It is interesting that for a hypothetical component of the universe with $w > 1/3$ and no vector anisotropic stress, the divergenceless part of the physical velocity perturbation δv_i^\perp increases with time. However, as soon as $\delta v_i^\perp \sim 1$, this increase will stop, simply because speeds greater than the speed of light are not possible. In this limit, we have

$$\delta v_i^\perp \simeq \text{const.} \quad (\text{A.69})$$

This can be taken to mean that any component of the universe having $c_s^2 \equiv \dot{\rho}/\dot{p} > 1/3$ will eventually tend to $c_s^2 \leq 1/3$.

We can interpret this in terms of the equation of state by writing c_s^2 in terms of w as follows:

$$\dot{w} = \left(\frac{p}{\rho} \right) \dot{\quad} \quad (\text{A.70a})$$

$$= \frac{\dot{p}}{\rho} - \frac{p\dot{\rho}}{\rho^2} \quad (\text{A.70b})$$

$$= (c_s^2 - w) \frac{\dot{\rho}}{\rho} \quad (\text{A.70c})$$

$$= 3H(w - c_s^2)(1 + w) . \quad (\text{A.70d})$$

Thus, we have

$$c_s^2 = w - \frac{\dot{w}}{3H(1 + w)} . \quad (\text{A.70e})$$

Finally, we prove the independence of the tensors from scalars and vectors. Here the procedure is a slightly more algebraically involved, but conceptually straightforward. It is possible to uniquely project out the transverse traceless part of a tensor, which is what we are going to do here. We know that this will completely isolate the terms with h_j^i in $\delta\mathcal{G}_j^i$ and the term with $\pi_j^{T\ i}$ in $\delta\mathcal{T}_j^i$, since by definition these are the transverse traceless parts of the metric and stress energy perturbations respectively.

The transverse projection operator to a vector \mathbf{v} in Euclidean space is

$$P_j^i = \delta_j^i - \frac{r^i r_j}{r^2} . \quad (\text{A.71})$$

This operator subtracts out the component of any vector along \mathbf{r} . In vector notation, the action of this operator on any vector \mathbf{v} is

$$P\mathbf{v} = \mathbf{v} - (\mathbf{v} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}} . \quad (\text{A.72})$$

We thus obtain the transverse part $H_j^{i\ \text{tr}}$ of any tensor H_j^i as

$$H_j^{i\ \text{tr}} = P_l^i P_j^m H_m^l . \quad (\text{A.73})$$

APPENDIX A. A SIMPLE PROOF OF THE DECOMPOSITION THEOREM

Next, to obtain the transverse traceless part, we subtract the trace of $H_j^{i \text{ tr}}$ from $H_j^{i \text{ tr}}$. The result is

$$H_j^{i \text{ TT}} = P_l^i P_j^m H_m^l - \frac{1}{2} P_j^i P_l^m H_m^l . \quad (\text{A.74})$$

In our case, when analyzing the Einstein equations, by the transverse traceless part we mean the part that is transverse to the wave vector \mathbf{k} in Fourier space, and traceless. On applying eq. (A.74) to $\delta\mathcal{G}_j^i = \delta\mathcal{T}_j^i$ in Fourier space, with the vector \mathbf{r} in eq. (A.72) set to \mathbf{k} , we find the transverse traceless part of the spatial Einstein equations to be

$$\delta\mathcal{G}_j^{i \text{ TT}} = \delta\mathcal{T}_j^{i \text{ TT}} , \quad (\text{A.75})$$

which is found to be

$$\frac{1}{2} \left(\ddot{h}_j^i + 3H\dot{h}_j^i - \frac{1}{a^2} \nabla^2 h_j^i \right) = \pi_j^{i T} . \quad (\text{A.76})$$

We have carried out the algebra leading to the above equation in some detail in Appendix A.

Note that we can also use the transverse projection operator eq. (A.73) to obtain the vector evolution equation from the $0i$ component of eq. (A.39), since the terms involving vectors are the only terms transverse to \mathbf{k} . To obtain the vector evolution equation from the ij component, we would first take the divergence along the i index which will remove the terms involving tensors and keep the terms involving scalars, and half of the terms involving vectors. Then we would again apply the transverse projection operator to isolate the terms involving vectors. Here we opted to take the curl of these equations because it seems less algebraically involved and more connected to our intuition about vectors.

Thus, we have proved that the evolution of each of the 3 classes of metric perturbations—scalars, vectors and tensors, is independent of the other two classes, and depends only on stress energy perturbations of the same class.

A.4 Generalizing to curved FLRW universes

Here, we consider curved universes with the metric

$$ds^2 = dt^2 - a^2(t) \left(\frac{dr^2}{1 - \kappa r^2} + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2 \right). \quad (\text{A.77})$$

Here, $-\infty < \kappa < \infty$. $\kappa < 0$ corresponds to an open universe and $\kappa > 0$ corresponds to a closed universe. We cannot use Cartesian coordinates $x = r \cos \theta$, $y = r \sin \theta \cos \phi$, $z = r \sin \theta \sin \phi$ as this transformation is not invertible for $\kappa \neq 0$. The background Einstein tensor is

$$G_0^0 = 3 \left(H^2 + \frac{\kappa}{a^2} \right), \quad (\text{A.78a})$$

$$G_i^0 = 0, \quad (\text{A.78b})$$

$$G_j^i = 2\dot{H} + 3H^2 + \frac{\kappa}{a^2}. \quad (\text{A.78c})$$

The form of the background stress energy tensor is the same as that for the flat universe.

The metric perturbations can again be taken of the form eq. (A.1):

$$\begin{aligned} ds^2 = & (1 + 2A)dt^2 \\ & - 2a(t)(B_{,i} + S_i)dt dx^i \\ & - a^2(t)[(1 - 2\psi)\gamma_{ij} + 2E_{,ij} \\ & + F_{i,j} + F_{j,i} + h_{ij}]dx^i dx^j. \end{aligned} \quad (\text{A.79})$$

The difference from the flat case is that the spatial metric that we use to raise and lower spatial indices is not δ_{ij} , but γ_{ij} , which is a diagonal metric given by

$$\gamma_{rr} = \frac{1}{1 - \kappa r^2}, \gamma_{\theta\theta} = r^2, \gamma_{\phi\phi} = r^2 \sin^2 \theta. \quad (\text{A.80a})$$

Note that in this section, unless mentioned otherwise, all covariant derivatives correspond to the spatial metric γ_{ij} , not the FLRW metric.

It is a bit tedious to prove that the metric perturbations have exactly the same gauge transformation

APPENDIX A. A SIMPLE PROOF OF THE DECOMPOSITION THEOREM

laws as in the case of the flat FLRW metric— eq. (A.14)-eq. (A.20), we show some slightly tricky parts of the calculation in Appendix B. The proof is in fact general and applies to any diagonal spatial metric γ_{ij} , i.e. any 4 dimensional metric

$$ds^2 = dt^2 - a^2(t)\gamma_{ij}dx^i dx^j, \quad (\text{A.81})$$

where γ_{ij} is diagonal and does not depend on t .

Due to the perturbations having the same gauge transformation laws, the gauge invariant perturbations are the same as that for the flat case, and given by

$$\Phi \equiv A + [a(B - a\dot{E})]', \quad (\text{A.82})$$

$$\Psi \equiv \psi - [aH(B - a\dot{E})], \quad (\text{A.83})$$

$$V_i \equiv S_i - a\dot{F}_i. \quad (\text{A.84})$$

Since the background stress energy tensor is the same as the flat case, the gauge invariant stress energy tensor is found by the same laws eq. (A.28), and is the same as eq. (A.38) with the partial derivatives replaced by covariant derivatives.

The gauge invariant Einstein tensor is given by

$$\delta\mathcal{G}_0^0 = -6H\dot{\Psi} - 6H^2\Phi + \frac{2}{a^2}\nabla^2\Psi + \frac{6\kappa}{a^2}\Psi, \quad (\text{A.85})$$

$$\delta\mathcal{G}_i^0 = 2\dot{\Psi}_{;i} + 2H\Phi_{;i} - \frac{1}{2a}\nabla^2 V_i - \frac{\kappa}{a}V_i, \quad (\text{A.86})$$

$$\begin{aligned} \delta\mathcal{G}_j^i &= -\delta_j^i \left[2\ddot{\Psi} + 2H(3\dot{\Psi} + \dot{\Phi}) + 2(2\dot{H} + 3H^2)\Phi \right. \\ &\quad \left. + \frac{1}{a^2}\nabla^2(\Phi - \Psi) - 2\frac{\kappa}{a^2}\Psi \right] + \frac{1}{a^2}\nabla^i\nabla_j(\Phi - \Psi) \\ &\quad + \frac{1}{2a}[(\dot{V}_j + 2HV_j)_{;i} + (\dot{V}^i + 2HV^i)_{;j}] \\ &\quad - \frac{1}{2} \left(\ddot{h}_j^i + 3H\dot{h}_j^i - \frac{1}{a^2}\nabla^2 h_j^i + \frac{2\kappa}{a^2}h_j^i \right). \end{aligned} \quad (\text{A.87})$$

Thus, the gauge invariant Einstein equations are

$$-6H\dot{\Psi} - 6H^2\Phi + \frac{2}{a^2}\nabla^2\Psi + \frac{6\kappa}{a^2}\Psi = \bar{\delta}\rho, \quad (\text{A.88a})$$

$$2\dot{\Psi}_{;i} + 2H\Phi_{;i} - \frac{1}{2a}\nabla^2V_i - \frac{\kappa}{a}V_i = (\rho + p)(\bar{v}_{;i} + \delta u_i^\perp), \quad (\text{A.88b})$$

$$\begin{aligned} & -\delta_j^i \left[2\ddot{\Psi} + 2H(3\dot{\Psi} + \dot{\Phi}) + 2(2\dot{H} + 3H^2)\Phi \right. \\ & \left. + \frac{1}{a^2}\nabla^2(\Phi - \Psi) - 2\frac{\kappa}{a^2}\Psi \right] + \frac{1}{a^2}\nabla^i\nabla_j(\Phi - \Psi) \\ & + \frac{1}{2a}[(\dot{V}_j + 2HV_j)_{;i} + (\dot{V}^i + 2HV^i)_{;j}] \\ & - \frac{1}{2} \left(\ddot{h}_j^i + 3H\dot{h}_j^i - \frac{1}{a^2}\nabla^2h_j^i \right) \end{aligned} \quad (\text{A.88c})$$

$$= \bar{\delta}p\delta_j^i - \left(\nabla^i\nabla_j - \delta_j^i\frac{\nabla^2}{3} \right) \pi^S - \pi_{;j}^V{}^i - \pi_{;i}^V{}^j - \pi_j^T{}^i. \quad (\text{A.88d})$$

Now, the procedure to decouple the scalar, vector and tensor equations is the same as that used for the flat metric. We start with separating the scalar equations from the Einstein equations:

$$\delta\mathcal{G}_0^0 = \delta\mathcal{T}_0^0, \quad (\text{A.89a})$$

$$\delta\mathcal{G}_i^{0;i} = \delta\mathcal{T}_i^{0;i}, \quad (\text{A.89b})$$

$$\delta\mathcal{G}_i^i = \delta\mathcal{T}_i^i, \quad (\text{A.89c})$$

$$\delta\mathcal{G}_{j;i}^{i;j} = \delta\mathcal{T}_{j;i}^{i;j}. \quad (\text{A.89d})$$

The resulting evolution equations for the scalar variables are

$$-6H\dot{\Psi} - 6H^2\Phi + \frac{2}{a^2}\nabla^2\Psi + \frac{6\kappa}{a^2}\Psi = \bar{\delta}\rho, \quad (\text{A.90})$$

$$\nabla^2(2\dot{\Psi} + 2H\Phi) = (\rho + p)\nabla^2 v, \quad (\text{A.91})$$

$$\begin{aligned} & -3 \left[2\ddot{\Psi} + 2H(3\dot{\Psi} + \dot{\Phi}) + 2(2\dot{H} + 3H^2)\Phi \right. \\ & \left. + \frac{1}{a^2}\nabla^2(\Phi - \Psi) - 2\frac{\kappa}{a^2}\Psi \right] + \frac{1}{a^2}\nabla^2(\Phi - \Psi) = -3\bar{\delta}p, \end{aligned} \quad (\text{A.92})$$

$$\frac{1}{a^2}\nabla^2[\nabla^2(\Phi - \Psi)] = -\nabla^2(\nabla^2\pi^S). \quad (\text{A.93})$$

We find that the only modification from the flat universe is that a term $\frac{6\kappa}{a^2}$ gets added to the 00 component and the trace of the ij components of the Einstein equations.

Now we consider the vector equations. We can still take the curl like we did for the δ_{ij} spatial metric, but we need to give some extra justification on how we can make taking the curl a tensor operation and why it works for decoupling the vector equations.

In a curved space, the Levi-Civita symbol ϵ^{ijk} is not a tensor. However, it can be easily modified to make the result a tensor while preserving its property of being antisymmetric in all its indices— it turns out that the quantity

$$\bar{\epsilon}^{ijk} \equiv \frac{\epsilon^{ijk}}{\sqrt{|g|}} \quad (\text{A.94})$$

behaves like a tensor. Here, g refers to the determinant of the metric γ_{ij} , which is positive, but the expression contains $|g|$ because it also applies to metrics where g can be negative. The curl operation on a vector V looks like

$$(\nabla \times V)^i = \bar{\epsilon}^{ijk}\nabla_j V_k. \quad (\text{A.95})$$

Secondly, we can explain why the curl operation decouples the vector equations from the $0i$ and ij components of Einstein equations. The reason why the curl operation works for the flat universe is because the terms involving scalars in the $0i$ component of the Einstein equation and the divergence of

the ij component are of the form $\partial_i s$, where s is a scalar perturbation. On taking the curl of these terms we get $\epsilon^{ijk} \partial_j \partial_k s$, which vanishes because the partial derivatives commute and ϵ^{ijk} is antisymmetric in j and k , this is the reason why the curl of a gradient is zero. In a curved space, we have covariant derivatives instead of partial derivatives. Covariant derivatives do not commute in general, so the operator $\epsilon^{ijk} \nabla_j \nabla_k$ does not vanish in general. But it turns out that covariant derivatives commute on a scalar, i.e.

$$\nabla_i \nabla_j s = \nabla_j \nabla_i s . \quad (\text{A.96})$$

This is why taking the curl in the equations

$$\epsilon^{ijk} k_j \delta \mathcal{G}_k^0 = \epsilon^{ijk} k_j \delta \mathcal{T}_k^0 , \quad (\text{A.97})$$

$$\epsilon^{ijk} k_j \delta \mathcal{G}_{k,l}^l = \epsilon^{ijk} k_j \delta \mathcal{T}_{k,l}^l , \quad (\text{A.98})$$

still decouples the evolution equations of the vector perturbations. These evolution equations in Fourier space are

$$V_i = \frac{2a}{k^2 - 2\kappa} (\rho + p) \delta u_i^\perp , \quad (\text{A.99})$$

$$\dot{V}_i + 2H V_i = -2a \pi_i^V . \quad (\text{A.100})$$

It can be seen that our analysis on the behaviour of $(\rho + p) \delta u_i^\perp$ and the physical velocity δv_i^\perp obtained for the flat universe also holds here.

Finally, we can decouple the tensor evolution equations by taking the transverse traceless component of the ij component of the Einstein equations.

$$\frac{1}{2} \left(\ddot{h}_j^i + 3H \dot{h}_j^i - \frac{1}{a^2} \nabla^2 h_j^i + \frac{2\kappa}{a^2} h_j^i \right) = \pi_j^{iT} . \quad (\text{A.101})$$

Bibliography

- [1] Y. Akrami et al. (Planck Collaboration), “Planck 2018 results - X. Constraints on inflation”, *A&A* **641**, A10 (2020).
- [2] J. M. Bardeen, “Gauge-invariant cosmological perturbations”, *Phys. Rev. D* **22**, 1882–1905 (1980).
- [3] H. Kodama and M. Sasaki, “Cosmological Perturbation Theory”, *Progress of Theoretical Physics Supplement* **78**, 1–166 (1984).
- [4] V. Mukhanov, H. Feldman, and R. Brandenberger, “Theory of cosmological perturbations”, *Physics Reports* **215**, 203–333 (1992).
- [5] D. Baumann, *TASI Lectures on Inflation*, 2012.
- [6] H. V. Ragavendra, P. Saha, L. Sriramkumar, and J. Silk, “Primordial black holes and secondary gravitational waves from ultraslow roll and punctuated inflation”, *Phys. Rev. D* **103**, 083510 (2021).
- [7] D. J. Fixsen, E. S. Cheng, J. M. Gales, J. C. Mather, R. A. Shafer, and E. L. Wright, “The Cosmic Microwave Background Spectrum from the Full COBE* FIRAS Data Set”, *The Astrophysical Journal* **473**, 576 (1996).
- [8] A. Kogut, D. Fixsen, D. Chuss, J. Dotson, E. Dwek, M. Halpern, G. Hinshaw, S. Meyer, S. Moseley, M. Seiffert, D. Spergel, and E. Wollack, “The Primordial Inflation Explorer (PIXIE): a nulling polarimeter for cosmic microwave background observations”, *Journal of Cosmology and Astroparticle Physics* **2011**, 025 (2011).

BIBLIOGRAPHY

- [9] J. Chluba, J. Hamann, and S. P. Patil, “Features and new physical scales in primordial observables: Theory and observation”, *International Journal of Modern Physics D* **24**, 1530023 (2015).
- [10] J. Chluba, A. L. Erickcek, and I. Ben-Dayan, “PROBING THE INFLATON: SMALL-SCALE POWER SPECTRUM CONSTRAINTS FROM MEASUREMENTS OF THE COSMIC MICROWAVE BACKGROUND ENERGY SPECTRUM”, *The Astrophysical Journal* **758**, 76 (2012).
- [11] J. Chluba, “Which spectral distortions does CDM actually predict?”, *Monthly Notices of the Royal Astronomical Society* **460**, 227–239 (2016).
- [12] A. M. Green and B. J. Kavanagh, “Primordial black holes as a dark matter candidate”, *Journal of Physics G: Nuclear and Particle Physics* **48**, 043001 (2021).
- [13] C. T. Byrnes, P. S. Cole, and S. P. Patil, “Steepest growth of the power spectrum and primordial black holes”, *Journal of Cosmology and Astroparticle Physics* **2019**, 028 (2019).
- [14] H. V. Ragavendra and L. Sriramkumar, “Observational Imprints of Enhanced Scalar Power on Small Scales in Ultra Slow Roll Inflation and Associated Non-Gaussianities”, *Galaxies* **11**, 10.3390/galaxies11010034 (2023).
- [15] Z. Arzoumanian et al. (The NANOGrav Collaboration), “The NANOGrav 12.5 yr Data Set: Search for an Isotropic Stochastic Gravitational-wave Background”, *The Astrophysical Journal Letters* **905**, L34 (2020).
- [16] Z. Arzoumanian et al. (The NANOGrav Collaboration), “THE NANOGrav NINE-YEAR DATA SET: LIMITS ON THE ISOTROPIC STOCHASTIC GRAVITATIONAL WAVE BACKGROUND”, *The Astrophysical Journal* **821**, 13 (2016).
- [17] T. L. Smith and R. R. Caldwell, “LISA for cosmologists: Calculating the signal-to-noise ratio for stochastic and deterministic sources”, *Phys. Rev. D* **100**, 104055 (2019).
- [18] Pau Amaro-Seoane et al. (LISA Collaboration), *Laser Interferometer Space Antenna*, 2017.
- [19] J. Crowder and N. J. Cornish, “Beyond LISA: Exploring future gravitational wave missions”, *Phys. Rev. D* **72**, 083005 (2005).
- [20] K. Inomata and T. Nakama, “Gravitational waves induced by scalar perturbations as probes of the small-scale primordial spectrum”, *Phys. Rev. D* **99**, 043511 (2019).

- [21] V. Vaskonen and H. Veermäe, “Did NANOGrav See a Signal from Primordial Black Hole Formation?”, *Phys. Rev. Lett.* **126**, 051303 (2021).
- [22] Yi, Z., Fei, Q., “Constraints on primordial curvature spectrum from primordial black holes and scalar-induced gravitational waves”, *Eur. Phys. J. C* **83**, 82 (2023), doi.org/10.1140/epjc/s10052-023-11233-3.