

Multi-field inflation and cosmological perturbations

Jinn-Ouk Gong

*Asia Pacific Center for Theoretical Physics, Pohang 37673, Korea and
Department of Physics, Postech, Pohang 37673, Korea*

Abstract

We provide a concise review on multi-field inflation and cosmological perturbations. We discuss convenient and physically meaningful bases in terms of which perturbations can be systematically studied. We give formal accounts on the gauge fixing conditions and present the perturbation action in two gauges. We also briefly review non-linear perturbations.

Contents

1	Introduction	3
2	Inflation	4
2.1	Background equations	4
2.2	Cosmic microwave background	6
2.2.1	Generation of the CMB	6
2.2.2	Horizon problem	7
2.3	Inflation	10
2.3.1	Inflation: what and how	10
2.3.2	Horizon problem revisited	11
2.3.3	Single field inflation and slow-roll approximation	12
3	Background dynamics	14
4	Dynamics of perturbations	16
4.1	Solutions of constraints	16
4.2	Quadratic action for perturbations	17
4.3	Explicit calculations: two-field case	19
4.3.1	Kinematic basis	19
4.3.2	Mass basis	20
5	Formulation of perturbations	22
5.1	First-order form	23
5.2	Gauge fixing conditions	24
5.3	Quadratic action in the comoving gauge	27
5.4	Two-field case	29
5.4.1	Effective single field theory	29
6	Power spectrum	30
6.1	Free solutions of mode functions and power spectrum	30
6.2	Corrections to the power spectrum	33
6.2.1	Two-field case	35
6.3	Alternative solutions	36
7	Towards non-linear perturbations	38
7.1	Issue of mapping	39
7.2	Cubic order action	40
7.3	Large-scale approximation	42
7.3.1	Non-linear equations	43
7.3.2	Scale dependence in the δN formalism	46
7.3.3	Correlation functions in the δN formalism	47
8	Conclusions	50

1 Introduction

Cosmology, a comprehensive study on the origin and evolution of the universe, has become a branch of physical sciences only in twentieth century. With its theoretical foundation being provided by general relativity [1], the hot big bang cosmology emerged in 1920s in which the universe has ever been expanding from an extremely hot and dense initial state in very far past [2]. Supported by a series of observational discoveries including the expansion of the universe from the relation between distance and redshift of galaxies [3], and especially the cosmic microwave background (CMB) in 1965 [4], now the hot big bang cosmology is regarded as the standard model of cosmology.

CMB was generated when the universe has become cooled down due to the expansion so that the coupling between electrons and photons by Thompson scattering could not be maintained any longer [5]. This happened about 380,000 years after the big bang, thus the CMB may be regarded as a snapshot of the universe at this very early moment. On the whole observable range, the CMB exhibits an almost perfect black body spectrum corresponding to a homogeneous temperature of 2.725 K with an accuracy of $\mathcal{O}(10^{-5})$. This observation suggests that when the CMB was generated, the observable patch of the universe was in thermal equilibrium in a single causally connected patch. However, as we will see, naively the universe at that time was composed of a huge number of causally disconnected regions and thus there is no reason for them to have the same temperature with $\mathcal{O}(10^{-5})$ accuracy. That is, the hot big bang cosmology is plagued by an extremely finely tuned initial condition to reproduce the universe as we observe now [6], and this is the so-called horizon problem. There are similar problems of initial conditions with fine tuning such as flatness problem.

In early 1980s¹, it was realized that all the necessary initial conditions for successful hot big bang cosmology can be naturally provided by a period of rapid expansion of the universe at very early times, called “inflation” [7]. Furthermore, inflation not only solves the otherwise finely tuned initial conditions like the horizon problem, but also provides the seed of subsequent structure formation in the universe [11]: during inflation, quantum mechanical fluctuations are stretched due to the rapid expansion to become classical perturbations. After inflation, these perturbations become the seed for the temperature anisotropies in the CMB and the inhomogeneous distribution of galaxies on large scales. Thus, by observing them closely we can study the primordial cosmic inflation and the underlying physics. Indeed, an important prediction of inflation is that the primordial perturbations produced during inflation have more or less the same amplitude on different length scales, i.e. scale-invariant, since the expansion is so fast that no appreciable change in inflationary dynamics happened on the whole observable scales relevant for, say, the CMB observations. And this prediction has been verified with very high confidence by most recent CMB observations including the Planck missions [12].

An immediate question that follows would be how to realize inflation concretely. Since the energy scale of the very early universe when cosmic inflation occurred is likely to be extremely high, as large as 10^{15} GeV. This is far exceeding the energy scale we can probe with terrestrial particle accelerators such as the Large Hadron Collider at CERN, with which we have confirmed the standard model of particle physics up to TeV scale. We regard, however, the standard model of particle physics as a low-energy effective field theory of a parent theory relevant for

¹The idea of exponential expansion was known even in 1960s [8], and just before the term “inflation” appeared it was discussed in different contexts, e.g. phase transition [9] or singularity problem [10].

higher energy scales due to a number of reasons: small but non-zero neutrino masses are not explained, gravity is not included, and so on. Thus, inflation is likely to be described in the context of theories beyond the standard model of particle physics [13] such as supersymmetric theories and string theory. One common feature of those theories is that there are a multiple number of degrees of freedom [14] that could be relevant for inflation. Usually inflation is supposed to be driven by a single scalar field dominating the energy density of the universe during inflation, called inflaton. There are many scalar, possibly light, fields as the so-called superpartners in supersymmetric theories and as moduli fields in string theory, which all can in principle participate in inflationary dynamics and thus can be identified as the inflatons. Therefore we do have good theoretical motivations to consider more than single degree of freedom during inflation. Furthermore, the existence of other degrees of freedom can give rise to a phenomenology full of rich and interesting observational consequences that can be further constrained or even detected in near future.

The aim of this article is to provide a concise and easily accessible review for inflation driven by a multiple number of fields, complementing many excellent review articles [13, 15] and textbooks [16] on related topics. The outline of this article is as follows. In Section 2, we briefly recall the basic of inflationary cosmology. In Section 3 we move to our main topic of multi-field inflation by starting with the background dynamics. Section 4 is devoted to the conventional approach to the cosmological perturbations produced during multi-field inflation. In Section 5 we present more careful considerations on the formulation of perturbations with alternative choices of gauge. We devote Section 7 to discuss concisely non-linear perturbations. Then we conclude in Section 8.

2 Inflation

Before we begin our main topic of multi-field inflation, we quickly recall inflation: what it is, why it is attractive and how it occurs.

2.1 Background equations

We begin with the so-called Friedmann-Robertson-Walker (FRW) metric of a flat universe

$$ds^2 = -dt^2 + a^2(t)\delta_{ij}dx^i dx^j . \quad (2.1)$$

This metric describes a flat, homogeneous, isotropic and expanding universe parametrized by the scale factor $a(t)$. The spatial distance with the scale factor being singled out is described by $\delta_{ij}dx^i dx^j$, which is called *comoving* distance. On the contrary, the *physical* distance is multiplied by the scale factor.

We first want the key equations. These are given by the Einstein equation

$$G_{\mu\nu} = \frac{T_{\mu\nu}}{m_{\text{Pl}}^2} . \quad (2.2)$$

We first consider the left hand side of (2.2), namely, the Einstein tensor

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R . \quad (2.3)$$

We can immediately write each component of the metric tensor $g_{\mu\nu}$ and its inverse $g^{\mu\nu}$ as

$$\begin{aligned} g_{00} &= -1, & g_{ij} &= a^2 \delta_{ij}, \\ g^{00} &= -1, & g^{ij} &= a^{-2} \delta^{ij}. \end{aligned} \quad (2.4)$$

To compute the Einstein tensor $G_{\mu\nu}$, we need to calculate the Christoffel symbol, the Ricci tensor and the Ricci scalar:

$$\begin{aligned} \Gamma_{\mu\nu}^\rho &= \frac{1}{2} g^{\rho\sigma} (g_{\mu\sigma,\nu} + g_{\sigma\nu,\mu} - g_{\mu\nu,\sigma}), \\ R_{\mu\nu} &= \Gamma_{\mu\nu,\alpha}^\alpha - \Gamma_{\mu\alpha,\nu}^\alpha + \Gamma_{\sigma\alpha}^\alpha \Gamma_{\mu\nu}^\sigma - \Gamma_{\sigma\nu}^\alpha \Gamma_{\mu\alpha}^\sigma, \\ R &= g^{\mu\nu} R_{\mu\nu}. \end{aligned} \quad (2.5)$$

The non-zero components of the Christoffel symbols are, after some calculations,

$$\Gamma_{ij}^0 = a^2 H \delta_{ij}, \quad (2.6)$$

$$\Gamma_{0j}^i = \Gamma_{j0}^i = H \delta^i_j, \quad (2.7)$$

with $H = \dot{a}/a$ being the Hubble parameter, otherwise zero. Then, easily we have

$$R_{00} = -3 \left(H^2 + \dot{H} \right), \quad (2.8)$$

$$R_{ij} = a^2 \left(3H^2 + \dot{H} \right) \delta_{ij}, \quad (2.9)$$

$$R = 6 \left(\dot{H} + 2H^2 \right). \quad (2.10)$$

Thus, the non-zero components of the Einstein tensor (2.3), or more frequently $G^\mu{}_\nu = g^{\mu\rho} G_{\rho\nu}$, are

$$G_{00} = 3H^2, \quad (2.11)$$

$$G_{ij} = -a^2 \left(2\dot{H} + 3H^2 \right) \delta_{ij}, \quad (2.12)$$

$$G^0{}_0 = -3H^2, \quad (2.13)$$

$$G^i{}_j = - \left(2\dot{H} + 3H^2 \right) \delta^i_j. \quad (2.14)$$

As can be read from (2.2), the Einstein tensor which describes the structure of the space-time should be matched with the energy-momentum tensor which describes the matter residing in the space-time. On the assumption of the homogeneous and isotropic background, we may regard at the background level that the energy-momentum tensor is that of perfect fluid², i.e.

$$T^\mu{}_\nu = \text{diag}(-\rho, p, p, p). \quad (2.15)$$

²In terms of the general hydrodynamical matter fluid, the energy-momentum tensor is written as

$$T_{\mu\nu} = (\rho + p)u_\mu u_\nu + p g_{\mu\nu},$$

where u^μ is the fluid 4-velocity which satisfies

$$u^\mu u_\mu = g_{\mu\nu} u^\mu u^\nu = -1,$$

so that u^μ is a time-like, unit 4-vector. Thus we can set $u^\mu = (1, 0, 0, 0)$. Using these we can trivially find (2.15).

Now we can write each component of (2.2):

$$00 \text{ component: } H^2 = \frac{\rho}{3m_{\text{Pl}}^2}, \quad (2.16)$$

$$ij \text{ component: } -3H^2 - 2\dot{H} = \frac{p}{m_{\text{Pl}}^2}. \quad (2.17)$$

(2.16) is called the Friedmann equation, which relates the Hubble parameter to the energy density. Using (2.16) for (2.17) to replace H^2 with ρ , we can find the time variation of H as

$$\dot{H} = -\frac{\rho + p}{2m_{\text{Pl}}^2}. \quad (2.18)$$

Or, explicitly in terms of the time derivatives of the scale factor,

$$\frac{\ddot{a}}{a} = -\frac{\rho + 3p}{6m_{\text{Pl}}^2}. \quad (2.19)$$

We will refer to this equation soon. Note that by taking a time derivative of (2.16) and using (2.17) to eliminate \dot{H} , we can derive energy conservation equation

$$\dot{\rho} + 3H(\rho + p) = 0. \quad (2.20)$$

This is what we can find from the conservation of energy-momentum tensor: from

$$T^\mu{}_{\nu;\mu} = T^\mu{}_{\nu,\mu} - \Gamma_{\mu\nu}^\rho T^\mu{}_\rho + \Gamma_{\rho\mu}^\mu T^\rho{}_\nu = 0, \quad (2.21)$$

we can trivially check that $\nu = 0$ component gives (2.20). $\nu = i$ component vanishes identically.

2.2 Cosmic microwave background

2.2.1 Generation of the CMB

With the necessary background equations, now let us see what happened in the past when the temperature was high. First, we note that from the conservation equation (2.20) that different species scale differently: ordinary particles (electron, proton, neutron...) have very large rest energy compared to the kinetic energy, so they are called pressureless matter and $p = 0$. Meanwhile, photons, or more generally relativistic particles, have $p = \rho/3$ and are called radiation. Plugging these relations into (2.20), we find

$$\rho_{\text{matter}} \propto a^{-3}, \quad (2.22)$$

$$\rho_{\text{radiation}} \propto a^{-4}. \quad (2.23)$$

We may understand that the energy density of pressureless matter is inversely proportional to the volume $\sim a^3$ which contains the matter particles, and for radiation the energy density is also proportional to the frequency, or the inverse of the wavelength, so we have one more power of the scale factor. What this tells us is that, in the past, the universe was dominated by radiation.

More radiation in the past means, of course, the universe was hotter. It was too hot to maintain neutral molecules, like hydrogen: because of the very hot temperature, electrons were energetic enough to overcome the binding energy to protons, so that the universe was filled by radiation (mostly photons), free electrons and nuclei (and dark matter). During this stage, the mean free path of photons was very short because of the Thomson scattering between free electrons and photons, maintaining thermal equilibrium. Thus, the universe was very “foggy” for photons: exactly like we cannot see very far away when the weather is very foggy. This stage continued until the universe was cooled to a critical temperature $T_c \sim 3000$ K. Below this temperature, the binding energy between electrons and protons could overcome thermal background and there remained no free electron. Thus, from this time on, the universe has become transparent to photons and they could reach us after propagating for a long long time. This situation is depicted in Figure 1. These very old photons, which have traveled since the moment of this “last scattering”, are the cosmic microwave background (CMB). It was observed in 1965 by Penzias and Wilson by chance.

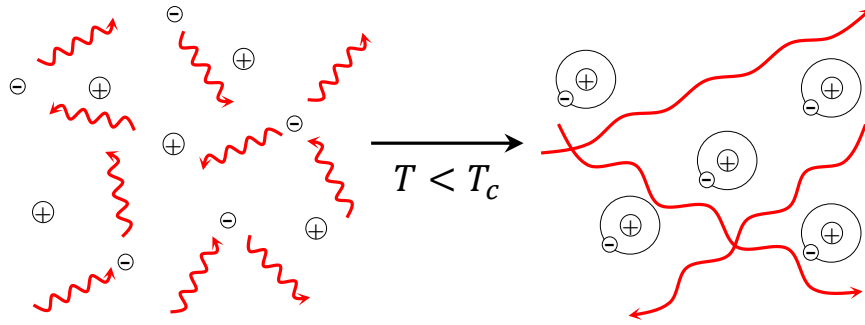


Figure 1: When $T > T_c$, electrons were free and constantly scattered off photons, so that the universe was “foggy”. After the temperature drops below T_c , electrons are all captured by protons, and photons can propagate without scattering.

The observations tell us that the CMB is extremely homogeneous and isotropic, i.e. we observe the same average temperature $T_0 \sim 2.7$ K no matter which part or direction of the sky we observe. Since photons were constantly scattering off free electrons and thus in thermal equilibrium, the temperature spectrum of the CMB exhibits that of almost perfect blackbody radiation. Moreover, the CMB could be generated only when the universe was hotter in the past. Thus the discovery of the CMB was the knockdown blow for the steady state cosmology which was competing against the hot big bang model in 60’s. Note that, after removing all the contaminations and foreground effects, we have genuine temperature fluctuations of the magnitude $\delta T/T_0 \sim 10^{-5}$. We will return to this point later. In Figure 2 we show the background and fluctuation temperature maps of the CMB.

2.2.2 Horizon problem

The CMB has brought, with the triumph of the hot big bang cosmology, big mysteries at the same time. Let us consider 1 of them, namely, why the CMB is so much homogeneous. For

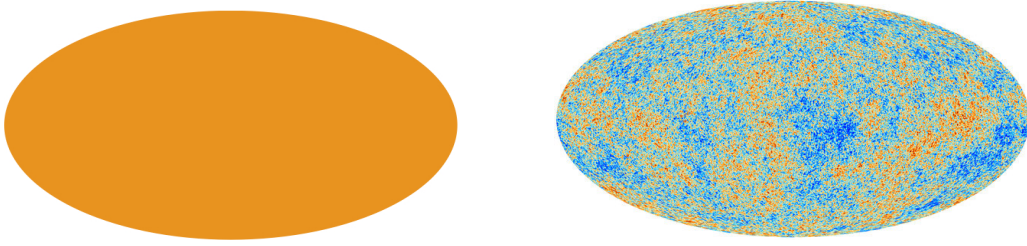


Figure 2: (Left) the cosmic microwave background is observed to be extremely homogeneous and isotropic with the average temperature $T_0 \sim 2.7$ K. (Right) however, it contains genuine temperature fluctuations with respect to T_0 of the magnitude $\delta T/T_0 \sim 10^{-5}$. The temperature fluctuation map is taken by the Planck satellite [17].

this, it is very convenient to introduce the conformal time τ , defined by

$$d\tau \equiv \frac{dt}{a}. \quad (2.24)$$

With τ , the line element (2.1) is written as

$$ds^2 = a^2(\tau) (-d\tau^2 + \delta_{ij} dx^i dx^j), \quad (2.25)$$

so that the metric is written as a product of the static Minkowski metric times the scale factor. What does the conformal time mean? Let us consider the radial propagation of light, which is the null geodesic $ds^2 = 0$. Then, using the spherical coordinate we can write the radial distance r a photon has traveled from some initial moment in terms of the conformal time as

$$r = \tau, \quad (2.26)$$

i.e. the conformal time measures the (comoving) distance a photon has traveled.

Then what's the trouble with the CMB? We can straightforwardly find that from an initial moment i till present 0, the conformal time (i.e. the distance photons have traveled)

$$\tau = \int_i^0 \frac{dt}{a} = \int_i^0 \frac{1}{a} \frac{dt}{da} da = \int_i^0 \frac{1}{aH} d \log a \propto a^{1/2} \Big|_i^0, \quad (2.27)$$

where for each equality we have used 1) the scale factor is a function of time solely, $a = a(t)$, 2) the definition of the Hubble parameter, $\dot{a} = aH$, and 3) assumption of a matter dominated universe, $H \sim \rho_{\text{matter}}^{1/2} \sim a^{-3/2}$. Now, without loss of generality, we can take initial moment as the initial singularity $a(t_i) = 0$, where also $\tau = 0$, so that simply $\tau \propto a^{1/2}$. Further, using the relation between the scale factor which is normalized to $a_0 = 1$ at present and the redshift z

$$a = \frac{1}{1+z}, \quad (2.28)$$

we can find $\tau \propto (1+z)^{-1/2}$. Using $z_0 = 0$ and $z_{\text{CMB}} \sim 1100$, we can easily find

$$\frac{\tau_{\text{CMB}}}{\tau_0} \sim \frac{1}{\sqrt{1100^3}} \sim 0.03. \quad (2.29)$$

Thus, at the moment when the CMB was generated, the past light cones stemming from the two end points do not have any overlapping region initially, i.e. those two points were never in causal communication and thus there is no reason they should have the same temperature with the accuracy of 10^{-5} : we must impose a heavy fine tuning over $10^4 - 10^5$ causally disconnected patches at the moment of the last scattering unless we provide a natural way for them to have the same temperature. This is the so-called horizon problem. It is depicted in the left panel of Figure 3. Note that the spatial distance shown in the figure is the comoving one, thus the physical distance is obtained by multiplying the scale factor $a(t)$ which vanishes as we approach the cosmic singularity, currently at $\tau = 0$.

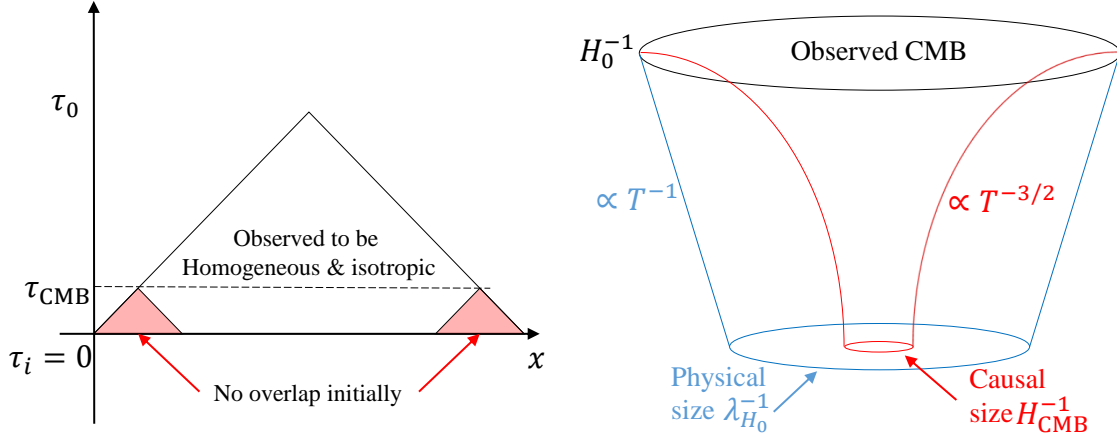


Figure 3: (Left) conformal diagram of the universe. From the cosmic singularity ($\tau_i = 0$) until the moment of the CMB generation (τ_{CMB}) there was no time for the CMB to achieve causal communication to have the same temperature T_0 . (Right) as a sample calculation, we can see that at that time the universe was filled with $10^4 - 10^5$ causally disconnected patches.

To have a more concrete idea, let us assume that the observed CMB size coincides with the current Hubble patch $1/H_0$, within which causal communications are possible. Then let us ask whether they were the same when the CMB was generated, or if different how much they were different. First, what is $\lambda_{H_0^{-1}}$, the *physical* size that corresponds to $1/H_0$? Physical sizes simply scale with the scale factor $a(t)$, which is inversely proportional to the temperature T . Thus, we can easily find

$$\lambda_{H_0^{-1}} = H_0^{-1} \frac{a_{\text{CMB}}}{a_0} = H_0^{-1} \frac{T_0}{T_{\text{CMB}}}. \quad (2.30)$$

Meanwhile, H evolves according to the Friedmann equation (2.16). It is important to notice at this moment that H depends on the energy density, i.e. which types of matter contents are there. For simplicity we assume the universe is dominated by matter that is inversely proportional to the physical volume as can be read from (2.22), and thus H is proportional to T^3 :

$$H^2 \propto \rho_{\text{matter}} \propto a^{-3} \propto T^3, \quad (2.31)$$

so that we can find H_{CMB}^{-1} , the Hubble horizon radius when the CMB was generated, as

$$H_{\text{CMB}}^{-1} = H_0^{-1} \left(\frac{T_0}{T_{\text{CMB}}} \right)^{3/2}. \quad (2.32)$$

Thus, if we compare the ratio of these volumes,

$$\frac{\lambda_{H_0^{-1}}^3}{(H_{\text{CMB}}^{-1})^3} = \left(\frac{T_{\text{CMB}}}{T_0}\right)^{3/2} \sim 4 \times 10^4. \quad (2.33)$$

That is, assuming that at present the Hubble horizon size and the CMB scale are the same, when the CMB was generated, the corresponding physical volume was filled with $10^4 - 10^5$ causally disconnected patches: see the right panel of Figure 3. Thus, it is a tremendous fine tuning that these disconnected patches all turn out to have the same temperature with the accuracy of 10^{-5} as the current observations on the CMB demand.

2.3 Inflation

2.3.1 Inflation: what and how

Thus, we see that at the heart of the horizon problem lies the fact that the Hubble horizon $1/H = 1/(\dot{a}/a)$ always expands faster than the physical length scale $\lambda \sim a$,

$$\frac{d}{dt} \left(\frac{\lambda}{H^{-1}} \right) \sim \frac{d}{dt} \left[\frac{a}{(\dot{a}/a)^{-1}} \right] = \ddot{a} < 0, \quad (2.34)$$

irrespective of whether the universe is dominated by matter or radiation. Thus, we can just turn upside down and make the physical size expands faster than the Hubble horizon: then physical scales expand faster than the horizon so causal communication could be possible during this stage. This tells us

$$\frac{d}{dt} \left(\frac{\lambda}{H^{-1}} \right) > 0 \quad \longleftrightarrow \quad \ddot{a} > 0. \quad (2.35)$$

That is, the universe experiences an *accelerated* expansion. This period of accelerated expansion is called “*inflation*”.

How can we more quantitatively say if it’s inflation or not? We can rewrite (2.19) as

$$\frac{\ddot{a}}{a} = \frac{2\rho}{6m_{\text{Pl}}^2} - \frac{3\rho + 3p}{6m_{\text{Pl}}^2} = H^2 + \dot{H} > 0, \quad (2.36)$$

where the 2nd equality follows by applying (2.16) and (2.18), and the last inequality is the definition of inflation (2.35). Thus, inflation occurs when the following condition is satisfied:

$$\epsilon \equiv -\frac{\dot{H}}{H^2} < 1. \quad (2.37)$$

This parameter, which tells whether it’s inflation or not, is called “slow-roll” parameter, in the context of slow-roll inflation: see the next section.

So with what kind of matter can we have inflation? From (2.19), we see that to have $\ddot{a} > 0$ we need a special form of matter which has a negative pressure, $p < -\rho/3$, or in terms of the equation of state w ,

$$w \equiv \frac{p}{\rho} < -\frac{1}{3}. \quad (2.38)$$

Clearly usual pressureless matter ($w = 0$) or radiation ($w = 1/3$) cannot support inflation. The simplest candidate is the so-called cosmological constant Λ , which has

$$p_\Lambda = -\rho_\Lambda \quad (w_\Lambda = -1) . \quad (2.39)$$

Then the Friedmann equation (2.16) is trivially solved: since Λ is, as the name suggests, a constant and so is H :

$$H^2 = \frac{\Lambda}{3m_{\text{Pl}}^2} = \text{constant} . \quad (2.40)$$

Thus we can see that the scale factor increases exponentially during inflation as

$$a \propto \exp \left(\sqrt{\frac{\Lambda}{3m_{\text{Pl}}^2}} t \right) . \quad (2.41)$$

2.3.2 Horizon problem revisited

So the question is: how does inflation solve the horizon problem? Now we can move to the conformal time to see a clear visualization how inflation solves the horizon problem. During inflation, for convenience driven by a cosmological constant Λ so that H is constant, the conformal time is given by

$$\tau = \int \frac{dt}{a} = \int \frac{e^{-Ht}}{a_0} dt = -\frac{1}{aH} < 0 . \quad (2.42)$$

That is, the conformal time is *negative* during inflation. Further, now the cosmic singularity $a = 0$ can be pushed to $\tau = -\infty$. Thus, even the two end points at $\tau = \tau_{\text{CMB}}$ have no overlap at $\tau = 0$, now τ can be extended to negative infinity during inflation so that there could be ample overlapping region enough to explain the homogeneity of the CMB.

As is clear from Figure 4, the longer inflation lasts, the larger the overlapping region becomes. Thus we need a certain duration of inflation to explain the homogeneous CMB. The amount of inflation is quantified by the number of e -folds N between some initial (i) and final (f) moments, which is given by

$$N \equiv \int_i^f H dt = \int_i^f \frac{da}{a} = \log \left(\frac{a_f}{a_i} \right) . \quad (2.43)$$

Thus, with a given N , the final scale factor is related to the initial scale factor by $a_f = a_i e^N$, i.e. the universe has expanded by e^N times. Now we can compute how large N should be for the CMB. The most natural way is that at the beginning of inflation (or the part of inflation relevant for our observable universe) the physical length scale $\lambda_{H_0^{-1}}$ is smaller than the Hubble horizon during inflation H_I so that causal communication has been established within $\lambda_{H_0^{-1}}$ to have the same temperature. This gives

$$\lambda_{H_0^{-1}} = H_0^{-1} \frac{a_i}{a_0} = H_0^{-1} \frac{a_f}{a_0} \frac{a_i}{a_f} = H_0^{-1} \frac{T_0}{T_f} e^{-N} < H_I^{-1} . \quad (2.44)$$

Thus, solving for N from the last inequality, we obtain

$$N > \log \left(\frac{T_0}{H_0} \right) - \log \left(\frac{T_f}{H_I} \right) \sim 67 - \log \left(\frac{T_f}{H_I} \right) , \quad (2.45)$$

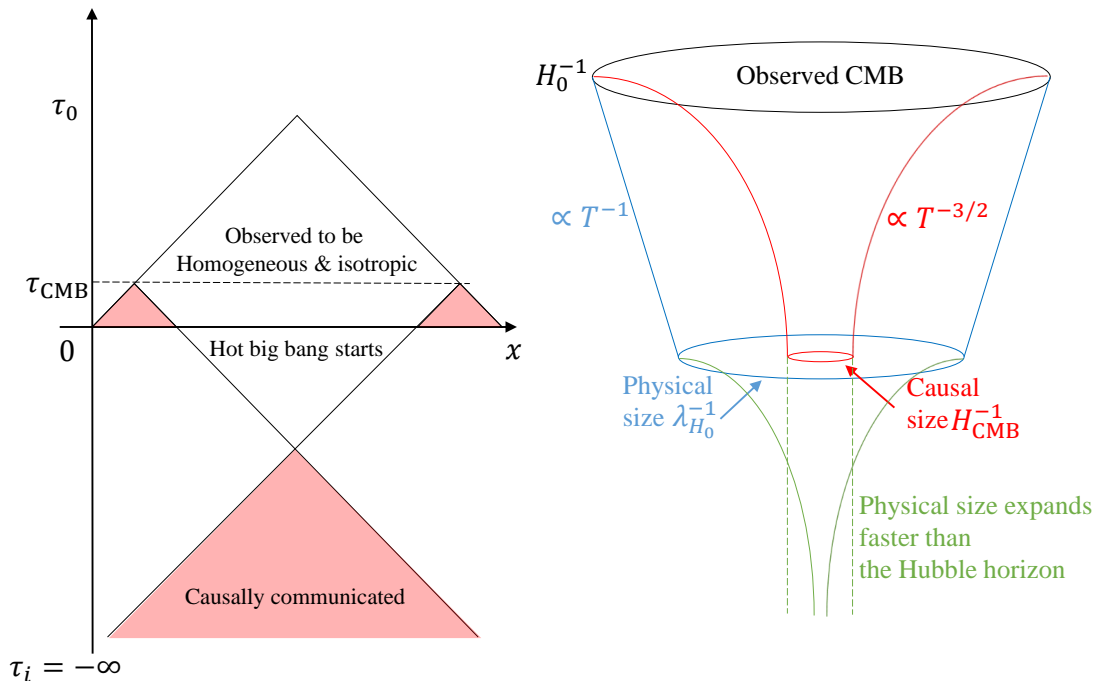


Figure 4: (Left) conformal diagram of the universe, this time including inflation. Inflation extends τ to $-\infty$, giving ample room for causal communication well before the onset of hot big bang evolution at $\tau = 0$. (Right) inflation corresponds to the period when the physical size $\lambda \sim a$ expands faster than the Hubble horizon $1/H$.

where we have used $H_0 \sim 10^{-42}$ GeV and $T_0 \sim 10^{-13}$ GeV. Thus, assuming that the logarithmic term which includes two unknown factors give a number of $\mathcal{O}(1)$, we require that

$$N \gtrsim 60. \quad (2.46)$$

That is, to explain the homogeneity of the CMB, i.e. to solve the horizon problem, we need 60 e -folds of expansion: during inflation the universe should have expanded by $e^{60} \sim 10^{26}$ times.

2.3.3 Single field inflation and slow-roll approximation

The cosmological constant is obviously the simplest candidate that drives inflation, but the problem is that if this is the case, inflation never ends and we cannot recover the universe in which we live with stars, galaxies, clusters of galaxies and so on. Thus, we need some different material which can mimic the cosmological constant and at the same time provide a “graceful exit” from inflation. This is usually achieved by a scalar field ϕ . For simplicity here we assume that this scalar field, named “inflaton” in the sense that it drives inflation, is minimally coupled to gravity and has canonical kinetic term. Then the action is the sum of the gravitational sector, which we take the Einstein-Hilbert action, and the matter sector:

$$S = \int d^4x \sqrt{-g} \frac{m_{\text{Pl}}^2}{2} R + \int d^4x \sqrt{-g} \underbrace{\left[-\frac{1}{2} g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi - V(\phi) \right]}_{\equiv \mathcal{L}_m}. \quad (2.47)$$

The corresponding energy-momentum tensor $T_{\mu\nu}$ of ϕ can be obtained by perturbing the matter Lagrangian with respect to $g^{\mu\nu}$,

$$T_{\mu\nu} = -\frac{2}{\sqrt{-g}} \frac{\delta(\sqrt{-g}\mathcal{L}_m)}{\delta g^{\mu\nu}} = \partial_\mu\phi\partial_\nu\phi - g_{\mu\nu} \left[\frac{1}{2}g^{\rho\sigma}\partial_\rho\phi\partial_\sigma\phi + V(\phi) \right]. \quad (2.48)$$

Then we can easily compute 00 and ii components which can then be matched to the energy density and pressure respectively³ [see (2.15)]:

$$\rho = -T^0_0 = \frac{1}{2}\dot{\phi}^2 + \frac{1}{2}\frac{(\nabla\phi)^2}{a^2} + V, \quad (2.49)$$

$$p = \frac{1}{3}T^i_i = \frac{1}{2}\dot{\phi}^2 - \frac{1}{6}\frac{(\nabla\phi)^2}{a^2} - V. \quad (2.50)$$

Thus, if potential dominates over the kinetic energy ($\partial_\mu\phi\partial^\mu\phi \ll V$) these simplify to $\rho \approx V \approx -p$, thus the inflaton provides a nearly cosmological constant, leading to an exponential expansion of the universe, viz. inflation.

Let us write the background equation of motion for ϕ . From this we can find a number of useful formulae which do not resort to the dynamics of ϕ but to V and its derivatives only. The equation of motion for ϕ can be found from the Euler-Lagrange equation,

$$\partial_\mu \left[\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)} \right] = \frac{\partial\mathcal{L}}{\partial\phi}. \quad (2.51)$$

This gives

$$-\square\phi + \frac{\partial V}{\partial\phi} = 0, \quad (2.52)$$

where

$$\square \equiv \frac{1}{\sqrt{-g}}\partial_\mu(\sqrt{-g}g^{\mu\nu}\partial_\nu) = -\frac{\partial^2}{\partial t^2} - 3H\frac{\partial}{\partial t} + \frac{\Delta}{a^2}, \quad (2.53)$$

with $\Delta \equiv \delta^{ij}\partial_i\partial_j$ being the spatial Laplacian operator. Here for the last equality we have taken the background metric. Thus, the background field $\phi = \phi(t)$ follows the equation of motion

$$\ddot{\phi} + 3H\dot{\phi} + \frac{\partial V}{\partial\phi} = 0. \quad (2.54)$$

So how this equation for ϕ simplifies? From (2.19), using (2.49) and (2.50) we require

$$\frac{\ddot{a}}{a} = -\frac{\dot{\phi}^2 - V}{3m_{\text{Pl}}^2} > 0. \quad (2.55)$$

Note that we can again precisely find (2.37) by using (2.49) and (2.50) for (2.16) and (2.18). Then this means

$$\dot{\phi}^2 < V. \quad (2.56)$$

³The relations between fluid quantities in terms of one or more scalar fields are highly non-trivial. See [18] for example.

Taking a time derivative on both sides, this says $\ddot{\phi} < \partial V/\partial\phi$. Thus, (2.54) is simplified to

$$3H\dot{\phi} + \frac{\partial V}{\partial\phi} = 0. \quad (2.57)$$

Thus we can replace $\dot{\phi}$, or more generally the dynamics of ϕ , with the derivatives of the potential V .

Then now let us consider the slow-roll parameter ϵ , (2.37). Applying (2.57), we find

$$\epsilon = -\frac{\dot{H}}{H^2} \approx \frac{\dot{\phi}^2/(2m_{\text{Pl}}^2)}{V/(3m_{\text{Pl}}^2)} \approx \frac{3}{2} \frac{1}{V} \frac{V'^2}{9H^2} \approx \frac{m_{\text{Pl}}^2}{2} \left(\frac{V'}{V}\right)^2, \quad (2.58)$$

where $V' \equiv \partial V/\partial\phi$. Thus, ϵ in the slow-roll approximation tells us how steep the potential slope is. We can introduce another important slow-roll parameter η , which describes how quickly ϵ evolves:

$$\eta \equiv \frac{\dot{\epsilon}}{H\epsilon} \approx \left[H \frac{m_{\text{Pl}}^2}{2} \left(\frac{V'}{V}\right)^2 \right]^{-1} m_{\text{Pl}}^2 \frac{V'}{V} \left[\frac{V''}{V} - \left(\frac{V'}{V}\right)^2 \right] \dot{\phi} \approx 2m_{\text{Pl}}^2 \frac{V''}{V} + 4\epsilon. \quad (2.59)$$

Also note that in the slow-roll approximation the e -fold N can be written in terms of the potential solely:

$$N = \int_i^f H dt = \int_i^f H \frac{dt}{d\phi} d\phi = \int_i^f \frac{H}{\dot{\phi}} d\phi \approx \frac{1}{m_{\text{Pl}}^2} \int_f^i \frac{V}{V'} d\phi. \quad (2.60)$$

3 Background dynamics

In the previous section, we have briefly recalled the basic of inflation driven by a single inflaton field. Being well acquainted with the prerequisite knowledge, now we move to our main topic of multi-field inflation. Explicitly, we consider a n -dimensional multi-field system with a generic field space metric G_{ab} coupled to the Einstein gravity:

$$S = \int d^4x \sqrt{-g} \left[\frac{m_{\text{Pl}}^2}{2} R - \frac{1}{2} G_{ab} g^{\mu\nu} \partial_\mu \phi^a \partial_\nu \phi^b - V(\phi) \right]. \quad (3.1)$$

We may well consider more general possibilities, such as $f(R)$ gravity [19] or $P(G_{ab}, X^{ab}, \phi^a)$ with $X^{ab} \equiv -g^{\mu\nu} \partial_\mu \phi^a \partial_\nu \phi^b/2$ [20]. But the discussions and considerations presented in this article can be straightly applied and extended, so we restrict our discussions to (3.1).

We begin with the background dynamics. The energy-momentum tensor derived from (3.1) is of the form

$$T_{\mu\nu} = G_{ab} \partial_\mu \phi^a \partial_\nu \phi^b - g_{\mu\nu} \left[\frac{1}{2} G_{ab} g^{\rho\sigma} \partial_\rho \phi^a \partial_\sigma \phi^b + V(\phi) \right], \quad (3.2)$$

from which we can combine the 00 and ij components of the Einstein equation to find

$$H^2 = \frac{1}{3m_{\text{Pl}}^2} \left[\frac{1}{2} \dot{\phi}_0^2 + V(\phi_0) \right], \quad (3.3)$$

$$\dot{H} = -\frac{\dot{\phi}_0^2}{2m_{\text{Pl}}^2}, \quad (3.4)$$

where $\dot{\phi}_0^2 \equiv G_{ab}\dot{\phi}_0^a\dot{\phi}_0^b$ with ϕ_0^a being the background value. The background equation of motion for the scalar fields can be obtained either by the variation of the action (3.1) with respect to ϕ^a or $\nu = 0$ component of the conservation of the energy-momentum tensor $T^\mu{}_{\nu;\mu} = 0$ as

$$D_t\dot{\phi}_0^a + 3H\dot{\phi}_0^a + G^{ab}V_b = 0, \quad (3.5)$$

where

$$D_t\dot{\phi}_0^a \equiv \frac{D\dot{\phi}_0^a}{dt} \equiv \frac{d\dot{\phi}_0^a}{dt} + \Gamma_{bc}^a\dot{\phi}_0^b\dot{\phi}_0^c \quad (3.6)$$

is a covariant time derivative with Γ_{bc}^a being the Christoffel symbol constructed by the field space metric G_{ab} . For more geometric discussions, see e.g. [21].

Before we proceed further to discuss perturbations, it is very useful to consider the change of basis in the field space. As in space-time, ϕ^a plays the role of coordinates in the field space⁴. Thus we are free to choose and/or transform to a convenient basis. One possible choice is a local orthogonal frame: we can introduce a complete set of vielbeins $e_a^I = e_a^I(t)$ which maps the general, arbitrary basis denoted by the index a into a local orthogonal frame denoted by the superscript I as

$$e_a^I e_b^J G^{ab} = \delta^{IJ} \quad \text{and} \quad e_a^I e_b^J \delta_{IJ} = G_{ab}. \quad (3.7)$$

Note that here we have not specified the new, orthogonal IJ basis: we may choose whatever frame we like as long as it is orthogonal. A physically important one is the so-called ‘‘kinematic basis’’, which is set along and perpendicular to the field trajectory [21, 22, 23]. The unit tangent vector T^a is defined by

$$T^a \equiv \frac{\dot{\phi}_0^a}{\dot{\phi}_0}. \quad (3.8)$$

The normal vector N^a which satisfies $G_{ab}T^aN^b = 0$ is naturally proportional to the derivative of T^a , i.e. $D_tT^a \propto N^a$. Taking a time derivative to T^a and using (3.5), we find

$$D_tT^a = -\frac{\ddot{\phi}_0}{\dot{\phi}_0}T^a - \frac{1}{\dot{\phi}_0}\left(3H\dot{\phi}_0^a + V^a\right). \quad (3.9)$$

Projecting this equation along T^a and N^a , we obtain respectively

$$\ddot{\phi}_0 + 3H\dot{\phi}_0 + V_T = 0, \quad (3.10)$$

$$D_tT^a = -\frac{V_N}{\dot{\phi}_0}N^a \equiv \dot{\theta}N^a, \quad (3.11)$$

where $V_T \equiv V_aT^a$ and $V_N \equiv V_aN^a$ are the projection of the potential derivative onto the tangential and perpendicular direction to the trajectory respectively, and we have *defined* the proportionality parameter as the angular velocity $\dot{\theta}$ of the trajectory. Note that the background equation along the tangent direction (3.10) is precisely the same as that in single field inflation (2.54). Thus we are naturally led to identify the tangential component of the field fluctuations as what is associated with the curvature perturbation, as we will see in later sections. Also note that the kinematic basis is not the only sensible choice for an orthogonal basis. We will discuss another physically illuminating choice in the next section.

⁴This is of particular caution when we discuss non-linear perturbations, for which we will further return to this point in Section 7.

4 Dynamics of perturbations

Up to now, we have considered background dynamics only. But as mentioned before, the observables in the universe such as the temperature anisotropies of the CMB and the inhomogeneous distribution of galaxies on large scales are originated from the primordial perturbations. During inflation, the whole observable patch of the universe was once deep inside the horizon, subject to quantum fluctuations around the homogeneous background. These fluctuations exit the horizon and become classical perturbations before they disappear due to rapid expansion. Since the perturbation in the matter sector is equivalent to that in space-time, there are small perturbations in otherwise smooth three-hypersurfaces. Once inflation is over and Hubble horizon expands faster than the physical scales, these perturbations enter the horizon and matter piles up, leading eventually to gravitational collapse when the amplitude of density perturbation exceeds certain critical value. Thus cosmological perturbations produced during inflation lies at the heart of the observational tests. In this section, we first take a conventional approach to the cosmological perturbations in multi-field inflation to gain solid idea. More careful considerations on cosmological perturbations will be given in the following section.

4.1 Solutions of constraints

We begin with the Arnowitt-Deser-Misner (ADM) form of the metric [24],

$$ds^2 = -\mathcal{N}^2 dt^2 + \gamma_{ij}(\beta^i dt + dx^i)(\beta^j dt + dx^j), \quad (4.1)$$

where \mathcal{N} is the lapse function and β^i is the shift vector and γ_{ij} is the spatial metric. With the extrinsic curvature

$$K_{ij} = \frac{1}{2\mathcal{N}} (\partial_t \gamma_{ij} - \beta_{i;j} - \beta_{j;i}), \quad (4.2)$$

with the covariant derivative being with respect to γ_{ij} , the action (3.1) is rewritten as

$$S = \int d^4x \sqrt{\gamma} \mathcal{N} \left[\frac{m_{\text{Pl}}^2}{2} (R^{(3)} + K_{ij} K^{ij} - K^2) - \frac{1}{2} G_{ab} g^{\mu\nu} \partial_\mu \phi^a \partial_\nu \phi^b - V(\phi) \right], \quad (4.3)$$

where $R^{(3)}$ is the curvature of three-dimensional hypersurface of constant t , constructed from γ_{ij} , and $K \equiv K^i_i$.

Varying the action (4.3) with respect to \mathcal{N} and β^i respectively yield the constraints

$$R^{(3)} - (K^{ij} K_{ij} - K^2) = \frac{2}{m_{\text{Pl}}^2} \left[\frac{1}{2} G_{ab} g^{\mu\nu} \partial_\mu \phi^a \partial_\nu \phi^b + V(\phi) \right], \quad (4.4)$$

$$(K^{ij} - \gamma^{ij} K)_{;j} = \frac{G_{ab}}{m_{\text{Pl}}^2 \mathcal{N}} \left(\dot{\phi}^b - \beta^j \partial_j \phi^b \right) \partial^i \phi^a. \quad (4.5)$$

Solving these constraints give the solutions of unphysical perturbation variables. As advertised, in this section we work in the flat gauge in which the scalar sector of the spatial metric γ_{ij} is unperturbed:

$$\gamma_{ij} = a^2(t) (\delta_{ij} + h_{ij}^{TT}), \quad (4.6)$$

where the pure tensor h_{ij}^{TT} is transverse and traceless:

$$h^{TT^i}_i = h^{TT^ij}_{;j} = 0. \quad (4.7)$$

For our discussion in this section at linear order, it is sufficient to obtain the linear solutions of the constraints. Writing the perturbation parts of the variables as

$$\mathcal{N} = 1 + \alpha, \quad (4.8)$$

$$\beta_i = \chi_{,i} + \beta_i^T, \quad (4.9)$$

$$\phi^a = \phi_0^a + Q^a, \quad (4.10)$$

where the transverse vector β_i^T satisfies $\partial^i \beta_i^T = 0$, we find the solutions of the constraints as

$$\alpha = \frac{G_{ab}}{2m_{\text{Pl}}^2 H} \dot{\phi}_0^b Q^a \equiv \mathcal{N}_a Q^a, \quad (4.11)$$

$$-2m_{\text{Pl}}^2 H \frac{\Delta}{a^2} \chi = 6m_{\text{Pl}}^2 H^2 \alpha - \alpha \dot{\phi}_0^2 + G_{ab} \dot{\phi}_0^a D_t Q^b + V_a Q^a, \quad (4.12)$$

and $\beta_i^T = 0$. Plugging these solutions back into the action, we obtain the quadratic action for the field fluctuations Q^a relevant for linear perturbation theory.

We can also arrive at the same solutions from the conventional perturbed Einstein equation. We can write the perturbed metric as

$$ds^2 = -(1 + 2A)dt^2 + 2a\mathcal{B}_i dt dx^i + a^2 [(1 + 2\varphi)\delta_{ij} + 2\mathcal{E}_{ij}] dx^i dx^j, \quad (4.13)$$

where the $0i$ and ij components can be decomposed into

$$\mathcal{B}_i = B_{,i} + S_i, \quad (4.14)$$

$$\mathcal{E}_{ij} = H_{T,ij} + F_{(i,j)} + \frac{1}{2} h_{ij}^{TT}, \quad (4.15)$$

with $A = \alpha$ and $B = \chi/a$. Here the pure vectors S_i and F_i are transverse:

$$S^i_{,i} = F^i_{,i} = 0. \quad (4.16)$$

Then at linear order the scalar, vector and tensor equations are all decoupled and we can consider them independently from each other. The flat gauge condition (5.35) corresponds to setting $\varphi = H_T = 0$, the 00 and $0i$ components of the Einstein equation are respectively, with the energy-momentum tensor (3.2),

$$6H^2 A + 2H \frac{\Delta}{a} B = \frac{1}{m_{\text{Pl}}^2} \left(-G_{ab} \dot{\phi}_0^a D_t Q^b + \dot{\phi}_0^2 A - V_a Q^a \right), \quad (4.17)$$

$$2HA_{,i} = \frac{1}{m_{\text{Pl}}^2} G_{ab} \dot{\phi}_0^a \partial_i Q^b, \quad (4.18)$$

from which we can find the same solutions for A and B as (4.11) and (4.12), respectively.

4.2 Quadratic action for perturbations

Having found the linear solutions of the constraints, now we can use them to write the quadratic action of the field fluctuations Q^a and the tensor perturbation h_{ij}^{TT} . After straightforward manipulations, we can obtain the quadratic action as [23, 25, 26, 27, 28]

$$S_2 = \int d^4x \frac{a^3}{2} \left\{ G_{ab} D_t Q^a D_t Q^b - \frac{\delta^{ij}}{a^2} G_{ab} \partial_i Q^a \partial_j Q^b - M_{ab}^2 Q^a Q^b + \frac{m_{\text{Pl}}^2}{4} \left[\left(h_{ij}^{TT} \right)^2 - \frac{1}{a^2} \partial^k h_{ij}^{TT} \partial_k h_{ij}^{TT} \right] \right\}, \quad (4.19)$$

where

$$M_{ab}^2 \equiv V_{ab} - \mathbb{R}_{acdb} \dot{\phi}_0^c \dot{\phi}_0^d + (3 - \epsilon) \frac{\dot{\phi}_{0a} \dot{\phi}_{0b}}{m_{\text{Pl}} m_{\text{Pl}}} + \frac{1}{m_{\text{Pl}}^2 H} \left(\dot{\phi}_{0a} V_b + \dot{\phi}_{0b} V_a \right), \quad (4.20)$$

with $V_{ab} \equiv V_{;ab}$, $\mathbb{R}_{bcd}^a \equiv \Gamma_{bd,c}^a - \Gamma_{bc,d}^a + \Gamma_{ce}^a \Gamma_{bd}^e - \Gamma_{de}^a \Gamma_{bc}^e$ being the Riemann tensor associated with the field space and ϵ being the slow-roll parameter defined in (2.37). The field indices are raised and lowered by the field space metric G_{ab} . The equations of motion we can find by perturbing (4.19) with respect to Q^a and h_{ij}^{TT} are, respectively,

$$D_t^2 Q^a + 3H D_t Q^a - \frac{\Delta}{a^2} Q^a + (M^2)^a_b Q^b = 0, \quad (4.21)$$

$$\ddot{h}_{ij}^{TT} + 3H \dot{h}_{ij}^{TT} - \frac{\Delta}{a^2} h_{ij}^{TT} = 0. \quad (4.22)$$

We see that in general the field fluctuations are coupled to each other through the effective mass matrix M_{ab}^2 . Also note that the tensor sector is exactly the same as that in single field inflation.

While we will solve (4.21) and (4.22) later in Section 6, at the moment for (4.21) it is convenient to make use of the local orthogonal vielbeins e_a^I introduced in Section 3. Using the orthogonal basis is particularly useful to connect the field fluctuations to the curvature perturbation. The field fluctuation Q^a in an arbitrary basis can be transformed into the one in the orthogonal frame by incorporating e_a^I as

$$Q^I = e_a^I Q^a. \quad (4.23)$$

Further, introducing $u^I \equiv aQ^I$ and moving to the conformal time $d\tau = dt/a$, we obtain

$$S_2 = \int d\tau d^3x \frac{1}{2} \left[\delta_{IJ} \left(u'^I u'^J + 2u'^I Z^J_K u^K + Z^I_K Z^J_L u^K u^L - \delta^{ij} \partial_i u^I \partial_j u^J \right) - a^2 M_{IJ}^2 u^I u^J \right], \quad (4.24)$$

where a prime denotes a derivative with respect to the conformal time, $M_{IJ}^2 \equiv e_a^I e_b^J M_{ab}^2 - H^2(2 - \epsilon)\delta_{IJ}$, and $Z^I_J \equiv e_a^I D_\tau e_J^a$. The resulting equation of motion is

$$u^{I''} + 2Z^I_{J'} u'^J + Z^I_J u'^J - Z^I_J Z^J_K u^K - \Delta u^I + a^2 (M^2)^I_J u^J = 0. \quad (4.25)$$

Thus we see that while the basis is orthogonal, the form of the equation of motion at first looks more complicated. This is because the vielbeins that map from an arbitrary basis to the orthogonal one are time-dependent, $e_a^I = e_a^I(t)$, so that the change of the vielbeins itself is reflected in the antisymmetric matrix $Z_{IJ} = -Z_{JI}$. Note, however, that we have used in (4.25) usual partial derivatives, not covariant ones. Indeed, using Z_{IJ} we can define a new covariant derivative \mathcal{D}_τ acting on quantities such as v^I labelled with the I -index as

$$\mathcal{D}_\tau u^I \equiv \frac{du^I}{d\tau} + Z^I_J u'^J. \quad (4.26)$$

Then the quadratic action (4.24) becomes very simple [23]:

$$S_2 = \int d\tau d^3x \frac{1}{2} \left[\delta_{IJ} \left(\mathcal{D}_\tau u^I \mathcal{D}_\tau u^J - \delta^{ij} \partial_i u^I \partial_j u^J \right) - a^2 M_{IJ}^2 u^I u^J \right], \quad (4.27)$$

so is the equation of motion:

$$\mathcal{D}_\tau^2 u^I - \Delta u^I + a^2 (M^2)^I_J u^J = 0. \quad (4.28)$$

4.3 Explicit calculations: two-field case

Having discussed the general aspects, in this section we consider explicitly two-field inflation case. The benefit is three-fold: first, we will perform explicit calculations in detail so that we see how the previous discussions become materialized. Second, two-field inflation is the simplest and thus most intuitive case with multiple number of fields, yet captures many important aspects of multi-field inflation. Lastly, we can visualize the geometric implications very easily.

4.3.1 Kinematic basis

First we consider the kinematic basis. For two-field case, one of the vielbeins corresponds to the tangent vector T^a and the other to the normal vector N^a ,

$$e_T^a = T^a \quad \text{and} \quad e_N^a = N^a. \quad (4.29)$$

Then the parallel and normal perturbations with respect to the inflationary trajectory are given respectively by

$$u^T = aQ^T = aT_a Q^a, \quad (4.30)$$

$$u^N = aQ^N = aN_a Q^a. \quad (4.31)$$

By choosing this frame, we find that $Z_{TN} = -Z_{NT} = -a\dot{\theta} = -\theta'$. Then, the quadratic action (4.24) becomes

$$\begin{aligned} S_2 = \int d\tau d^3x \frac{1}{2} & \left[u^{T'2} - (\nabla u^T)^2 - \left(a^2 M_{TT}^2 - \theta'^2 \right) u^{T2} \right. \\ & + u^{N'2} - (\nabla u^N)^2 - \left(a^2 M_{NN}^2 - \theta'^2 \right) u^{N2} \\ & \left. - 4\theta' u^{T'} u^N - 2 \left(a^2 M_{TN}^2 + \theta'' \right) u^T u^N \right], \end{aligned} \quad (4.32)$$

where the symmetric matrix M_{IJ}^2 consists of the following elements:

$$M_{TT}^2 = V_{TT} + 2H^2\epsilon(3 - \epsilon) + \frac{2}{m_{\text{Pl}}^2 H} \dot{\phi}_0 V_T - (2 - \epsilon)H^2, \quad (4.33)$$

$$M_{NN}^2 = M^2 - (2 - \epsilon)H^2, \quad (4.34)$$

$$M_{TN}^2 = V_{TN} - 2\epsilon\dot{\theta}H, \quad (4.35)$$

where $M^2 \equiv V_{NN} + \epsilon m_{\text{Pl}}^2 H^2 \mathbb{R}$ is the (effective) mass squared of the orthogonal mode u^N with \mathbb{R} being the Ricci scalar parametrizing the geometry of the field space, and the projections of the potential derivatives can be written as

$$V_{TT} = H^2 \left(3\epsilon - 3\delta_1 - \delta_2 + \frac{\dot{\theta}^2}{H^2} \right), \quad (4.36)$$

$$V_{TN} = H\dot{\theta} \left(-3 + 2\epsilon - 2\delta_1 + \frac{\ddot{\theta}}{H\dot{\theta}} \right), \quad (4.37)$$

where we have defined

$$\delta_n \equiv \frac{1}{H^n \dot{\phi}_0} \frac{d^{n+1} \phi_0}{dt^{n+1}}. \quad (4.38)$$

The action (4.32) in the kinematic basis leads to the coupled equations of motion describing the evolution of both modes u^T and u^N :

$$u^{T''} - 2\theta' u^{N'} + \left(-\Delta + a^2 M_{TT}^2 - \theta'^2\right) u^T + (a^2 M_{TN}^2 - \theta'') u^N = 0, \quad (4.39)$$

$$u^{N''} + 2\theta' u^{T'} + \left(-\Delta + a^2 M_{NN}^2 - \theta'^2\right) u^N + (a^2 M_{NT}^2 + \theta'') u^T = 0. \quad (4.40)$$

4.3.2 Mass basis

In the very previous section we have considered the perturbations in the kinematic basis, set along and perpendicular to the background inflationary trajectory. In general for a curved trajectory, which naturally incorporates “heavy” and “light” degrees of freedom, in the mass matrix in the kinematic basis which can be written as

$$V_{IJ} = \begin{pmatrix} V_{TT} & V_{TN} \\ V_{TN} & V_{NN} \end{pmatrix}, \quad (4.41)$$

the off-diagonal component of the mass matrix V_{TN} is non-zero, as can be read from (4.37). Thus, regarding heaviness (or lightness) of the relevant degrees of freedom, it is most convenient to adopt another set of basis in which the mass matrix V_{IJ} becomes diagonal. We may call this as the “mass basis” [29]. Then the two eigenvalues of the mass basis correspond to the light and heavy masses along the trajectory respectively. Notice that if one performs the change-of-basis around the bottom of the potential [30], no kinematic information is required because geometry determines everything. Explicit diagonalization of (4.41) gives two eigenvalues,

$$\lambda_{\pm} = \frac{1}{2} \left[V_{NN} + V_{TT} \pm (V_{NN} - V_{TT}) \sqrt{1 + \beta^2} \right] \quad \text{where} \quad \beta \equiv \frac{2V_{TN}}{V_{NN} - V_{TT}}, \quad (4.42)$$

so that λ_- (λ_+) corresponds to V_{ll} (V_{hh}). The corresponding eigenvectors transformed from the kinematic basis are then

$$\begin{aligned} e_l^a &= T^a \cos \psi - N^a \sin \psi = \begin{pmatrix} \cos \psi \\ -\sin \psi \end{pmatrix}, \\ e_h^a &= T^a \sin \psi + N^a \cos \psi = \begin{pmatrix} \sin \psi \\ \cos \psi \end{pmatrix}, \end{aligned} \quad (4.43)$$

with

$$\cos \psi \equiv \frac{1 + \sqrt{1 + \beta^2}}{\sqrt{2(1 + \beta^2 + \sqrt{1 + \beta^2})}} \quad \text{and} \quad \sin \psi \equiv \frac{\beta}{\sqrt{2(1 + \beta^2 + \sqrt{1 + \beta^2})}}. \quad (4.44)$$

So what does the angle ψ mean? This becomes transparent if we write the change-of-basis matrix P is, with e_l^a and e_h^a on the first and second column and row respectively is

$$P = \begin{pmatrix} \cos \psi & \sin \psi \\ -\sin \psi & \cos \psi \end{pmatrix}. \quad (4.45)$$

Thus, given the kinematic basis $\{T^a, N^a\}$, we can rotate it by ψ to obtain the mass basis. This situation is depicted in Figure 5. Notice that when we discuss about the kinematic basis, we take care of not θ , which is the rotation angle from the general basis $\{\phi^1, \phi^2\}$ to the kinematic one, but only $\dot{\theta}$. This is because the field space coordinates ϕ^a can be set totally arbitrary, so important is only the rate of change of the kinematic basis with respect to the general basis, i.e. the ‘‘angular velocity’’ of the trajectory. But ψ denotes the misalignment between the kinematic and mass bases, thus not only its rate of change but also its value itself are important to describe the dynamics along the trajectory: if $\psi \neq 0$ so that the two bases are misaligned, oscillations of the trajectory may be caused until the misalignment disappears.

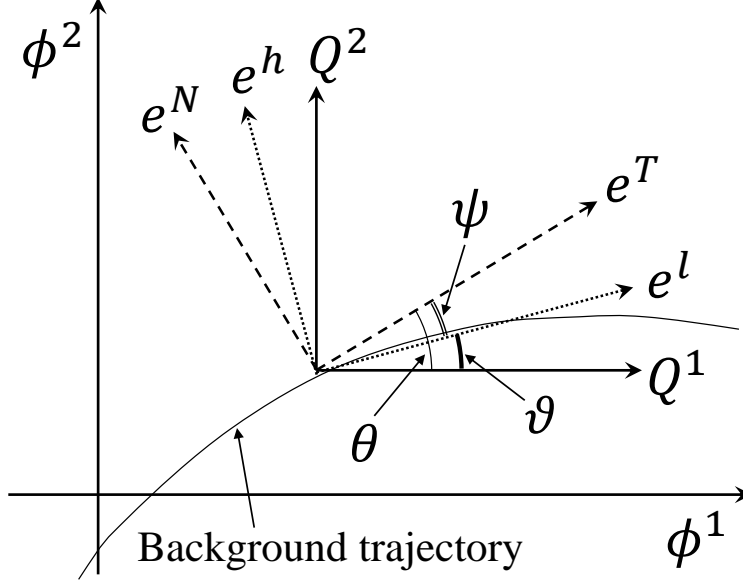


Figure 5: A schematic plot showing the relation between the arbitrary field basis ϕ^1, ϕ^2 , kinematic basis $\{e^T, e^N\}$ and mass basis $\{e^l, e^h\}$. The field fluctuations in the field basis are decomposed into $\{Q^1, Q^2\}$. The angular velocity of the kinematic basis is $\dot{\theta}$ while that of the mass basis is $\dot{\vartheta}$, so that θ, ϑ and ψ are related by $\theta = \vartheta + \psi$, although for θ and ϑ only their rates of change are important since the field basis is arbitrary, with respect to which θ and ϑ are set.

Finally, M_{IJ}^2 in (4.24) in the mass basis is

$$M_{ll}^2 = V_{ll} - 2\epsilon H^2 \left[\left(3 - \epsilon + \frac{\dot{\epsilon}}{H\epsilon} \right) \cos^2 \psi - \frac{\dot{\theta}}{H} \sin(2\psi) + m_{\text{Pl}}^2 \mathbb{R} \sin^2 \psi \right] - H^2(2 - \epsilon), \quad (4.46)$$

$$M_{lh}^2 = 2\epsilon H^2 \left[-\frac{1}{2} \left(3 - \epsilon + \frac{\dot{\epsilon}}{H\epsilon} \right) \sin(2\psi) - \frac{\dot{\theta}}{H} \cos(2\psi) + \frac{1}{2} m_{\text{Pl}}^2 \mathbb{R} \sin(2\psi) \right], \quad (4.47)$$

$$M_{hh}^2 = V_{hh} - 2\epsilon H^2 \left[\left(3 - \epsilon + \frac{\dot{\epsilon}}{H\epsilon} \right) \sin^2 \psi + \frac{\dot{\theta}}{H} \sin(2\psi) + m_{\text{Pl}}^2 \mathbb{R} \cos^2 \psi \right] - H^2(2 - \epsilon). \quad (4.48)$$

Then the quadratic action (4.24) becomes explicitly [31]

$$\begin{aligned}
S_2 = \int d\tau d^3x \frac{1}{2} & \left[u_l'^2 - (\nabla u_l)^2 - \left(a^2 M_{ll}^2 - \vartheta'^2 \right) u_l'^2 \right. \\
& + u_h'^2 - (\nabla u_h)^2 - \left(a^2 M_{hh}^2 - \vartheta'^2 \right) u_h'^2 \\
& \left. - 4\vartheta' u_l' u_h' - 2 \left(a^2 M_{lh}^2 + \vartheta'' \right) u_l u_h \right], \tag{4.49}
\end{aligned}$$

where $\vartheta' = (\theta - \psi)'$ denotes the angular velocity of the mass basis. The first and second lines denote the free terms of the light and heavy modes respectively, while the third line is the interaction between them. Note that the form of the action (4.49) is precisely the same as that in the kinematic basis (4.32). This should be obviously the case, since both the kinematic and mass bases are orthogonal ones.

5 Formulation of perturbations

In the previous section, we have considered the perturbations in the flat gauge to extract physical scalar degrees of freedom. Let us recall the general arguments: on top of the matter sector which contains n scalar fields, the gravitational sector brings additional 4 scalar, 4 vector and 2 tensor degrees of freedom. However, not all of them are physical, as we have the freedom to choose an arbitrary coordinate system with the same physics. That is, gravity allows gauge degrees of freedom. These fictitious gauge degrees of freedom can be removed by imposing appropriate gauge conditions and by solving the constraints – naively we have $n + 4$ scalar variables: n from n scalar field components and 4 from the metric. Since there are 1 temporal and 1 spatial gauge transformations in the scalar sector, we can eliminate 2 of them. In the flat gauge discussed in the previous section, we impose the conditions that the perturbations of three-dimensional spatial metric on each time slice vanish. The remaining metric degrees of freedom are perturbations of the lapse function and the shift vector. They further can be removed by solving 2 constraint equations, so that after all n degrees of freedom are left. Namely, we can write all the physical degrees of freedom solely in terms of the field fluctuations Q^a .

This however is not necessarily the only sensible choice. Especially, we do not directly observe the inflaton fields – the temperature fluctuations in the CMB and the distribution of galaxies on large scales reflect the initial perturbation in the curvature of the constant-time spatial hypersurfaces. This curvature perturbation becomes manifest in other gauge choices, not in the flat gauge we adopted in the previous section in which by definition there is no perturbation on spatial hypersurfaces. But the curvature perturbation is associated with the metric, so in this case we have to give 1 degree of freedom to the gravitational sector while the remaining $n - 1$ to the matter sector that contains n fields. But how can we do this conveniently, since the gravitational sector receives contributions from the total matter contents? Thus, we need to reconsider the formulation of perturbations more formally to establish possible gauge choices on concrete ground. We can discuss formally the issue of gauge fixing in the Hamiltonian formulation, which is also useful for path integral quantization of cosmological perturbations [32]. In this section, we consider the Hamiltonian analysis of cosmological perturbations to study the gauge conditions systematically.

5.1 First-order form

For the Hamiltonian formulation, we begin with the canonical momentum of γ_{ij} as

$$\Pi^{ij} = \frac{\delta S_G}{\delta \partial_t \gamma_{ij}} = \frac{m_{\text{Pl}}^2}{2} \sqrt{\gamma} (K^{ij} - \gamma^{ij} K). \quad (5.1)$$

The pure gravity part of the action then is written in the first-order form,

$$S_G = \int d^4x \left(\Pi^{ij} \partial_t \gamma_{ij} - \mathcal{N} \mathcal{H}_G - \beta_i \mathcal{H}_G^i \right), \quad (5.2)$$

where

$$\mathcal{H}_G = \frac{2}{m_{\text{Pl}}^2 \sqrt{\gamma}} \left(\Pi^{ij} \Pi_{ij} - \frac{1}{2} \Pi^2 \right) - \frac{m_{\text{Pl}}^2}{2} \sqrt{\gamma} R^{(3)}, \quad (5.3)$$

$$\mathcal{H}_G^i = -2 \left(\partial_j \Pi^{ij} + \Gamma_{jk}^i \Pi^{jk} \right), \quad (5.4)$$

with $\Pi \equiv \Pi^i_i$. Here, the lapse \mathcal{N} and the shifts β_i play the role of Lagrangian multipliers for the constraints \mathcal{H} and \mathcal{H}^i respectively, which are interpreted as the generators of diffeomorphism. Likewise, we can write the matter part of the action in the first-order form as

$$S_M = \int d^4x \left(\Pi_a \partial_t \phi^a - \mathcal{N} \mathcal{H}_M - \beta_i \mathcal{H}_M^i \right), \quad (5.5)$$

where the canonical momentum Π_a and the constraints \mathcal{H}_M and \mathcal{H}_M^i are

$$\Pi_a = \frac{\sqrt{\gamma}}{\mathcal{N}} G_{ab} (\partial_t \phi^b - \beta^i \partial_i \phi^b), \quad (5.6)$$

$$\mathcal{H}_M = \frac{G^{ab}}{2\sqrt{\gamma}} \Pi_a \Pi_b + \frac{1}{2} \sqrt{\gamma} \gamma^{ij} G_{ab} \partial_i \phi^a \partial_j \phi^b + \sqrt{\gamma} V(\phi^a), \quad (5.7)$$

$$\mathcal{H}_M^i = \Pi_a \partial^i \phi^a. \quad (5.8)$$

Thus, the action in the first-order form is given by

$$S = \int d^4x \left[\Pi^{ij} \partial_t \gamma_{ij} + \Pi_a \partial_t \phi^a - \mathcal{N} (\mathcal{H}_G + \mathcal{H}_M) - \beta_i (\mathcal{H}_G^i + \mathcal{H}_M^i) \right]. \quad (5.9)$$

It is clearly seen that unphysical \mathcal{N} and β_i become Lagrange multipliers accompanying constraints $\mathcal{H}_G + \mathcal{H}_M$ and $\mathcal{H}_G^i + \mathcal{H}_M^i$, respectively.

We are interested in the cosmological perturbations around the time-dependent classical background. For gravity sector, we separate the FRW background quantities and their perturbations as:

$$\Pi^{ij} = \frac{P(t)}{6a(t)} [\delta^{ij} + \pi^{ij}(t, \mathbf{x})], \quad (5.10)$$

$$\gamma_{ij} = a^2(t) [\delta_{ij} + h_{ij}(t, \mathbf{x})], \quad (5.11)$$

$$\mathcal{N} = \mathcal{N}_0(t) + \alpha(t, \mathbf{x}), \quad (5.12)$$

$$\beta^i = \beta^i(t, \mathbf{x}). \quad (5.13)$$

Note that at the moment we have kept the background lapse function denoted by $\mathcal{N}_0(t)$, not simply setting it to 1. This is because we can obtain a background equation by varying the zeroth order action with respect to \mathcal{N}_0 [see (5.16) and (5.21)]. For the matter sector, the perturbations around the classical backgrounds $\phi_0^a(t)$ are written as

$$\phi^a = \phi_0^a(t) + Q^a(t, \mathbf{x}), \quad (5.14)$$

$$\Pi_a = P_a(t) + \pi_a(t, \mathbf{x}). \quad (5.15)$$

We can obtain the classical solutions from the zeroth order action,

$$S_0 = \int d^4x \left[P \partial_t a + P_a \partial_t \phi_0^a - \mathcal{N}_0 \left(-\frac{P^2}{12m_{\text{Pl}}^2 a} + \frac{G^{ab} P_a P_b}{2a^3} + a^3 V \right) \right]. \quad (5.16)$$

Varying this with respect to the classical backgrounds, we obtain the background equations of motion as:

$$\frac{\delta S_0}{\delta P} = 0 : \quad \dot{a} = -\frac{P}{6m_{\text{Pl}}^2 a}, \quad (5.17)$$

$$\frac{\delta S_0}{\delta a} = 0 : \quad \dot{P} = -\frac{P^2}{12m_{\text{Pl}}^2 a^2} + \frac{3G^{ab} P_a P_b}{2a^4} - 3a^2 V, \quad (5.18)$$

$$\frac{\delta S_0}{\delta P_a} = 0 : \quad \dot{\phi}_0^a = \frac{G^{ab} P_b}{a^3}, \quad (5.19)$$

$$\frac{\delta S_0}{\delta \phi_0^a} = 0 : \quad \dot{P}_a = -a^3 V_a - \frac{1}{2a^3} G^{cd}{}_{,a} P_c P_d, \quad (5.20)$$

$$\frac{\delta S_0}{\delta \mathcal{N}_0} = 0 : \quad \frac{P^2}{12m_{\text{Pl}}^2 a} = \frac{G^{ab} P_a P_b}{2a^3} + a^3 V, \quad (5.21)$$

Combining these equations, we obtain the familiar background equations (3.3), (3.4) and (3.5). Note that from the linear order action

$$\begin{aligned} S_1 = \int d^4x \mathcal{N}_0 & \left[\frac{a}{6} h \left(-\dot{P} + HP + \frac{P^2}{12m_{\text{Pl}}^2 a^2} + \frac{3G^{ab} P_a P_b}{2a^4} - 3a^2 V \right) + \frac{P}{3} \pi \left(\dot{a} + \frac{P^2}{6m_{\text{Pl}}^2 a} \right) \right. \\ & + Q^a \left(\dot{P}_a + \frac{1}{2a^3} G^{cd}{}_{,a} P_c P_d + a^3 V_a \right) + \pi_a \left(\dot{\phi}_0^a - \frac{G^{ab} P_b}{a^3} \right) \\ & \left. + \frac{\alpha}{\mathcal{N}_0} \left(\frac{P^2}{12m_{\text{Pl}}^2 a} - \frac{G^{ab} P_a P_b}{2a^3} - a^3 V \right) \right], \quad (5.22) \end{aligned}$$

where $h \equiv h^i{}_i$ and $\pi \equiv \pi^i{}_i$, we can immediately read the same background equations, which we obtained by perturbing the zeroth order action with respect to the background variables, as the constraints for the perturbation variables.

5.2 Gauge fixing conditions

Having found the first-order form of the action, now we can proceed formal discussion on the possible choice of gauge fixing. Before we begin explicitly with the quadratic action for cosmological perturbations, let us recall how to remove properly the unphysical degrees of

freedom for a constrained system, viz. when constraints are present [33]. Let us consider a Hamiltonian with r constraints in a $2f$ -dimensional phase space $(q_1, \dots, q_f; p_1, \dots, p_f)$,

$$H_{\text{total}}(q_i, p_i) = H(q_1, \dots, q_f; p_1, \dots, p_f) + \sum_{m=1}^r \lambda_m \chi_m(q_1, \dots, q_f; p_1, \dots, p_f), \quad (5.23)$$

where the Lagrangian multipliers λ_m are combinations of unphysical degrees of freedom. Since there are r constraints χ_m , the number of degrees of freedom in the phase space is reduced to $2f - r$ by solving r constraints. But the phase space should be even-dimensional, so total remaining degrees of freedom should be not $2f - r$ but $2f - 2r$. That is, still r unphysical degrees of freedom remain. This is because different values of λ_m correspond to different copies of physical degrees of freedom in the phase space on $\chi_m = 0$. Thus we need r more, *gauge-fixing* conditions $\psi_m(q_1, \dots, q_f; p_1, \dots, p_f) = 0$. In quantum field theory, these gauge fixing conditions are subject to $\det([\chi_m, \psi_n]) \neq 0$ so that just one set of physical degrees of freedom is chosen. By choosing gauge fixing conditions satisfying $[\psi_m, \psi_n] = 0$, we can identify ψ_m as unphysical canonical momenta p_m then the remaining $f - r$ momenta $(p_1^*, \dots, p_{f-r}^*)$ are physical. Then $\det([\chi_m, \psi_n]) = \det([\chi_m, p_n]) = \det(i\partial\chi_m/\partial q_n) \neq 0$ so that it is always possible to identify unphysical variables q_m by inverting $\chi_m(q_1, \dots, q_f; p_1^*, \dots, p_{f-r}^*; p_1 = \dots = p_r = 0) = 0$. Thus, we are finally left with $2f - 2r$ physical, constrained variables $(q_1^*, \dots, q_{f-r}^*; p_1^*, \dots, p_{f-r}^*)$.

Now we return to our discussion on the cosmological perturbations around the classical solutions. At quadratic order, the first-order form of the action is written as

$$S_2 = \int d^4x \left(\pi_a \partial_t Q^a - 2a^3 H \pi^{ij} \partial_t h_{ij} - \mathcal{H}_2 + \alpha C_1^0 + \beta_i C_1^i \right), \quad (5.24)$$

where \mathcal{H}_2 is the quadratic Hamiltonian, and C_1^μ denote the constraints linear in perturbations. They are given by

$$\begin{aligned} \mathcal{H}_2 = & 4a^3 H^2 \frac{m_{\text{Pl}}^2}{2} \left[\frac{1}{2} \pi^{ij} A_{ijkl} \pi^{kl} + \pi^{ij} \left(2h_{ij} - \frac{1}{2} \delta_{ij} h \right) \right] + \frac{G^{ab}}{2a^3} (\pi_a \pi_b - h P_a \pi_b) \\ & + \frac{a}{2} G_{ab} \partial_i Q^a \partial_i Q^b + \frac{a^3}{2} (V_{ab} Q^a Q^b + h V_a Q^a) \\ & + \left(\frac{5m_{\text{Pl}}^2}{4} a^3 H^2 + \frac{G^{ab}}{8a^3} P_a P_b \right) h^{ij} h_{ij} - \left(\frac{3m_{\text{Pl}}^2}{8} a^3 H^2 - \frac{G^{ab}}{16a^3} P_a P_b \right) h^2 \\ & + a \frac{m_{\text{Pl}}^2}{2} \left(\frac{1}{4} h \Delta h - \frac{1}{2} h h^{ij}{}_{,ij} + \frac{1}{2} h^{ij} \partial^k \partial_i h_{jk} - \frac{1}{4} h^{ij} \Delta h_{ij} \right) + a^3 \frac{V}{4} \left(\frac{h^2}{2} - h^{ij} h_{ij} \right), \end{aligned} \quad (5.25)$$

$$C_1^0 = a^3 h \left(\frac{G^{ab} P_a P_b}{4a^6} + \frac{m_{\text{Pl}}^2}{2} H^2 \right) + a \frac{m_{\text{Pl}}^2}{2} (h^{ij}{}_{,ij} - \Delta h) - a^3 V_a Q^a - \frac{G^{ab}}{a^3} P_a \pi_b + 2m_{\text{Pl}}^2 a^3 H^2 \pi - \frac{a^3}{2} h V, \quad (5.26)$$

$$C_1^i = -\frac{1}{a^2} P_a \partial^i Q^a - 2a H m_{\text{Pl}}^2 \left(\partial_j \pi^{ij} + \partial_j h^{ij} - \frac{1}{2} \partial^i h \right), \quad (5.27)$$

with $A_{ijkl} \equiv \delta_{ik} \delta_{jl} + \delta_{il} \delta_{jk} - \delta_{ij} \delta_{kl}$. Notice that going beyond cubic order, which is necessary for non-linear perturbation theory, we can schematically write

$$S_{\geq 3} = \int d^4x \left(-\mathcal{H}_{\geq 3} + \alpha C_{\geq 2}^0 + \beta_i C_{\geq 2}^i \right), \quad (5.28)$$

where the subscript $\geq n$ denotes n -th order and beyond in perturbations.

Since gravity has four constraints, we need four gauge fixing conditions ψ_μ ($\mu = 0, 1, 2, 3$) satisfying

$$\det(\{C^\mu, \psi_\nu\}) \neq 0, \quad (5.29)$$

where curly brackets denote the Poisson brackets. To extract the possible gauge fixing conditions, we need the Poisson brackets of the constraints C^μ with the fluctuations h_{ij} and Q^a . We can explicitly find

$$\begin{aligned} \{h_{ij}(t, \mathbf{x}), C^0(t, \mathbf{y})\} = & -2H\tilde{\gamma}^{-1/2} \left\{ \delta_{ij} + (-h_{ij} + \delta_{ij}h - 2\pi^{ij} + \delta_{ij}\pi) \right. \\ & + [(h + \pi)h_{ij} - 2h_{ik}h_{jk} - 2(h_{il}\pi^{jl} + h_{jl}\pi^{il}) + \pi^{kl}h_{kl}\delta_{ij}] \\ & \left. + (-2h_{il}\pi^{lk}h_{jk} + h_{kl}\pi^{kl}h_{ij}) \right\} \delta^{(3)}(\mathbf{x} - \mathbf{y}), \end{aligned} \quad (5.30)$$

$$\{h_{ij}(t, \mathbf{x}), C^{ik}(t, \mathbf{y})\} = -\frac{2}{a^2} (\delta_{k(i}\partial_{j)}^x - \Gamma_{ij}^k) \delta^{(3)}(\mathbf{x} - \mathbf{y}), \quad (5.31)$$

$$\{Q^a(t, \mathbf{x}), C^0(t, \mathbf{y})\} = -\frac{\tilde{\gamma}^{-1/2}}{a^3} (P_a + \pi_a) \delta^{(3)}(\mathbf{x} - \mathbf{y}), \quad (5.32)$$

$$\{Q^a(t, \mathbf{x}), C^i(t, \mathbf{y})\} = -\frac{1}{a^3} \tilde{\gamma}^{ik} \partial_k Q^a \delta^{(3)}(\mathbf{x} - \mathbf{y}), \quad (5.33)$$

where $\tilde{\gamma}_{ij} = \gamma_{ij}/a^2$. Since the participating degrees of freedom are the perturbed metric h_{ij} and the field fluctuations Q^a , we can think of four possibilities for the gauge fixing conditions:

1. Both the ψ^0 and the ψ^i conditions come from h_{ij} .
2. Both the ψ^0 and the ψ^i conditions come from Q^a .
3. The ψ^0 condition comes from h_{ij} , whereas the ψ^i condition comes from Q^a .
4. The ψ^0 condition comes from Q^a , whereas the ψ^i condition comes from h_{ij} .

However, the 3-vector condition made up of Q^a , say, $\psi_i = \psi_i(Q^a, \partial_k Q^a)$ satisfies

$$\begin{aligned} \{\psi_j(Q)(t, \mathbf{x}), C^i(t, \mathbf{y})\} &= -\frac{\tilde{\gamma}^{ik}}{a^2} \partial_k Q^a(t, \mathbf{y}) \left[\frac{\partial \psi_j}{\partial Q^a}(t, \mathbf{x}) + \frac{\partial \psi_j}{\partial (\partial_l Q^a)}(t, \mathbf{x}) \partial_l^x \right] \delta^{(3)}(\mathbf{x} - \mathbf{y}) \\ &= -\frac{\tilde{\gamma}^{ik}}{a^2} D_k \psi_j \delta^{(3)}(\mathbf{x} - \mathbf{y}). \end{aligned} \quad (5.34)$$

Thus it vanishes under $\psi_i = 0$, and (5.29) is violated. That is, the spatial vector constraints C^i cannot be satisfied with the field fluctuations Q^a . Hence, plausible gauge fixing conditions should be chosen between the possibilities 1 and 4.

For the first possibility, the gauge fixing is imposed entirely from the metric fluctuation. One simple example is

$$\begin{aligned} \psi_0 &= 0, \\ \psi_i &= \partial^j \left(h_{ij} - \frac{\delta_{ij}}{3} h \right). \end{aligned} \quad (5.35)$$

In terms of the metric decomposition in (4.13) and (4.15) the gauge fixing by $\psi_\mu = 0$ in this choice is to set the scalar components of h_{ij} zero, so that there is no scalar perturbation in the spatial metric as in (5.35). Thus in this gauge condition, the curvature of the three-hypersurface is uniform so it is called uniform-curvature gauge. Since we are considering spatially flat universe, we may call it “flat” gauge condition. In the flat gauge, the scalar degrees of freedom are entirely given to Q^a . Meanwhile, as an example for the fourth possibility, we can make use of the conditions

$$\begin{aligned}\psi_0 &= G_{ab}\dot{\phi}_0^a Q^b, \\ \psi_i &= \partial^j \left(h_{ij} - \frac{\delta_{ij}}{3} h \right).\end{aligned}\tag{5.36}$$

We can note from (3.2) that the condition $\psi_0 = 0$ is equivalent to $T^0_i = 0$. In terms of hydrodynamic fluid, the 3-velocity of the fluid vanishes in this gauge, $u_i = 0$. Thus in this gauge condition we do not observe momentum flux, and we move together with the cosmic fluid. So this gauge condition is known and the “comoving” gauge condition, and one scalar degree of freedom is associated with the metric – the curvature perturbation. But notice that all field contents contribute to ψ_0 . A more convenient way of implementing $\psi_0 = 0$ will be discussed in the following sections.

5.3 Quadratic action in the comoving gauge

Having discussed the possible gauge conditions, now we consider the quadratic action (6.2) for which we have not yet applied any gauge yet. For this, we decompose the metric perturbation h_{ij} using the scalar, vector and tensor components as we already did in (4.13) and (4.15)⁵:

$$h_{ij} = 2\varphi\delta_{ij} + 2H_{T,ij} + 2\partial_{(i}F_{j)} + h_{ij}^{TT}.\tag{5.37}$$

The quadratic action (6.2) looks horribly complex, so we first need to arrange terms for physical clarity. For this, we introduce the Mukhanov-Sasaki variable [35] à la single field inflation as

$$\tilde{Q}^a \equiv Q^a - \frac{\dot{\phi}_0^a}{H}\varphi.\tag{5.38}$$

Then, up to irrelevant auxiliary field terms which can be eliminated from dynamics after appropriate redefinitions, we obtain a surprisingly simple form:

$$S_2 = \int d^4x \frac{a^3}{2} \left\{ G_{ab} D_t \tilde{Q}^a D_t \tilde{Q}^b - \frac{G_{ab}}{a^2} \partial^i \tilde{Q}^a \partial_i \tilde{Q}^b - M_{ab}^2 \tilde{Q}^a \tilde{Q}^b + \frac{m_{\text{Pl}}^2}{4} \left[\left(\dot{h}_{ij}^{TT} \right)^2 - \frac{1}{a^2} \partial^k h_{ij}^{TT} \partial_k h_{ij}^{TT} \right] \right\},\tag{5.39}$$

where M_{ab}^2 is given by (4.20).

Since the pure tensor action in (5.39) is the same as that in single field inflation, we from now on concentrate on the scalar sector. As mentioned at the beginning of this section, it is very conventional to adopt the flat gauge for multi-field inflation in which $\varphi = H_T = 0$ so that simply

⁵We may write the scalar components as $2H_L\delta_{ij} + 2(\partial_i\partial_j - \Delta\delta_{ij}/3)H_T$ so that H_L is the only longitudinal contribution [34]. Note that in this decomposition $\varphi = H_T - \Delta H_T/3$.

$\tilde{Q}^a = Q^a$. This is because the number of the physical degrees of freedom, after eliminating the unphysical ones, is the same as the number of field contents. But in the comoving gauge, where one physical degree of freedom is the curvature perturbation φ , we need to fix a non-trivial temporal gauge condition C^0 for which all the field fluctuations contribute as can be read from ψ_0 in (5.36). Thus, (5.38) is, while it gives a very simple form of the quadratic action (5.39), not most convenient to implement the comoving gauge condition. A more convenient alternative is to decompose Q^a into the directions along and orthogonal to time evolution [36] as

$$Q^a(t, \mathbf{x}) = Q_\perp^a(t, \mathbf{x}) + \dot{\phi}_0^a(t) \tilde{\pi}(t, \mathbf{x}), \quad (5.40)$$

with the orthogonality condition

$$G_{ab} \dot{\phi}_0^a Q_\perp^b = 0. \quad (5.41)$$

We can then rewrite the temporal gauge condition of (5.36) very simply as

$$\psi_0 = \dot{\phi}_0^2 \tilde{\pi}. \quad (5.42)$$

Thus, we can impose the the comoving gauge conveniently by $\tilde{\pi} = 0$. Note that the linear gauge transformation $Q^a \rightarrow Q^a - \dot{\phi}_0^a \xi^0$ tells us

$$\tilde{\pi} \rightarrow \tilde{\pi} - \xi^0 \quad \text{and} \quad Q_\perp^a \rightarrow Q_\perp^a. \quad (5.43)$$

This means that $\tilde{\pi}$ is the fluctuation in the direction of the time translation itself, and is thus interpreted as the Goldstone mode resulting from the spontaneous breaking of the time translation invariance [37]. Meanwhile, the orthogonal fluctuations Q_\perp^a , which are usually called ‘‘isocurvature’’ modes, are gauge invariant⁶. Then, we can write the multi-field version of the Mukhanov-Sasaki variable (5.38) as

$$\tilde{Q}^a(t, \mathbf{x}) = Q_\perp^a(t, \mathbf{x}) - \frac{\dot{\phi}_0^a}{H} (\varphi - H\tilde{\pi})(t, \mathbf{x}) \equiv Q_\perp^a(t, \mathbf{x}) - \frac{\dot{\phi}_0^a}{H} \pi(t, \mathbf{x}). \quad (5.44)$$

Note that due to the orthogonality condition, $n - 1$ out of total n Q_\perp^a 's are independent in the comoving gauge: the remaining single degree of freedom is π , which is also gauge invariant. With this decomposition, the scalar quadratic action is rewritten in terms of gauge invariant variables π and Q_\perp^a as, using the time derivative of (5.41) to eliminate $D_t Q^a$ in the mixing term with π ,

$$S_2 = \int d^4x \frac{a^3}{2} \left[G_{ab} D_t Q_\perp^a D_t Q_\perp^b - \frac{G_{ab}}{a^2} \partial^i Q_\perp^a \partial_i Q_\perp^b - M_{ab}^2 Q_\perp^a Q_\perp^b + 2\epsilon m_{\text{Pl}}^2 \left(\dot{\pi}^2 - \frac{1}{a^2} \partial^i \pi \partial_i \pi \right) - \frac{4}{H} V_a Q_\perp^a \dot{\pi} \right]. \quad (5.45)$$

We close this section by writing the curvature perturbation in a form convenient for computing higher order correlation functions that we will discuss later. We can non-linearly generalize the metric perturbation (5.37) by exponentiating it as [42]

$$\gamma_{ij} = a^2(t) e^{2\varphi(t, \mathbf{x})} \left[e^{h^{TT}} \right]_{ij}, \quad (5.46)$$

⁶Note that in models with non-minimal couplings to gravity [38], for single field case the curvature perturbation remains invariant [39] under the conformal transformation by which the gravitational sector becomes the minimal Einstein-Hilbert one [40], but for multi-field this is not the case [41].

where $\det[e^{h^{TT}}] = 1$, i.e. the matrix h^{TT} is traceless and represents tensor perturbations, so that the scalar perturbation is isolated as φ . Notice that this form also means the scalar perturbation can be interpreted as the ‘‘local expansion rate’’ which can be written as the determinant of γ_{ij} : see (7.39). In the comoving gauge in which $\varphi = \mathcal{R}$, the transformation from π to \mathcal{R} is given by [43, 44]

$$\begin{aligned} \mathcal{R} = & \pi + \left(\epsilon - \frac{\delta}{2} \right) \pi^2 + \frac{1}{H} \pi \dot{\pi} - \frac{1}{4a^2 H^2} [\partial^i \pi \partial_i \pi - \partial^i \partial^j \Delta^{-1} (\partial_i \pi \partial_j \pi)] \\ & + \frac{\epsilon}{H} [\partial^i \pi \partial_i \Delta^{-1} \dot{\pi} - \partial^i \partial^j \Delta^{-1} (\partial_i \pi \partial_j \Delta^{-1} \dot{\pi})] - \frac{1}{4H} \dot{h}_{ij}^{TT} \partial^i \partial^j \pi + \dots \end{aligned} \quad (5.47)$$

In terms of \mathcal{R} , the quadratic action remains the same with π replaced by \mathcal{R} . But the transformation (5.47) does give rise to additional contributions to the higher order action in terms of π as we will see in Section 7.

5.4 Two-field case

Now we consider two-field case explicitly. Due to the orthogonality condition, we know that Q_\perp^a is proportional to the normal vector N^a : $Q_\perp^a \propto N^a$. We denote the amplitude of Q_\perp as \mathcal{F} , and use the (linear) relation (5.47) to replace π simply with \mathcal{R} . That is, the gauge-invariant variable \tilde{Q}^a is written in terms of \mathcal{R} and \mathcal{F} as

$$\tilde{Q}^a = N^a \mathcal{F} - \frac{\dot{\phi}_0^a}{H} \mathcal{R}. \quad (5.48)$$

Then the quadratic action becomes

$$S_2 = \int d^4x \frac{a^3}{2} \left\{ 2\epsilon m_{\text{Pl}}^2 \left[\dot{\mathcal{R}}^2 - \frac{(\nabla \mathcal{R})^2}{a^2} \right] + \dot{\mathcal{F}}^2 - \frac{(\nabla \mathcal{F})^2}{a^2} - M_{\text{eff}}^2 \mathcal{F}^2 + 4\dot{\theta} \frac{\dot{\phi}_0}{H} \dot{\mathcal{R}} \mathcal{F} \right\}. \quad (5.49)$$

where $M_{\text{eff}}^2 = V_{NN} + \epsilon m_{\text{Pl}}^2 H^2 \mathbb{R} - \dot{\theta}^2$. Identifying $u^T = -\dot{\phi}_0 \mathcal{R}/H$ and $u^N = \mathcal{F}$, after some manipulations we can recover (4.32).

5.4.1 Effective single field theory

Note that if a hierarchy of scales is present in the mass matrix, then we can compute a fairly reliable effective single field theory: see [45] for a concrete review on this subject. If we do not have any light mode other than \mathcal{R} , we can integrate out the heavy isocurvature modes \mathcal{F} by performing formally the path integral over \mathcal{F} . Writing (5.49) schematically as

$$S_2 = \frac{1}{2} \int \mathcal{R} \Delta_{\mathcal{R}\mathcal{R}} \mathcal{R} + \frac{1}{2} \int \mathcal{F} \Delta_{\mathcal{F}\mathcal{F}} \mathcal{F} + \int \mathcal{F} \Delta_{\mathcal{R}\mathcal{F}} \mathcal{R}, \quad (5.50)$$

we can evaluate the Gaussian integral over the heavy field \mathcal{F} as

$$\begin{aligned} e^{-S_{\text{eff}}[\mathcal{R}]} &= \exp \left(\frac{1}{2} \int \mathcal{R} \Delta_{\mathcal{R}\mathcal{R}} \mathcal{R} \right) \int \mathcal{D}\mathcal{F} \exp \left(\frac{1}{2} \int \mathcal{F} \Delta_{\mathcal{F}\mathcal{F}} \mathcal{F} + \int \mathcal{F} \Delta_{\mathcal{R}\mathcal{F}} \mathcal{R} \right) \\ &= \exp \left\{ \frac{1}{2} \int \mathcal{R} \Delta_{\mathcal{R}\mathcal{R}} \mathcal{R} - \frac{1}{2} \int \left[(\Delta_{\mathcal{R}\mathcal{F}} \mathcal{R}) \Delta_{\mathcal{F}\mathcal{F}}^{-1} (\Delta_{\mathcal{R}\mathcal{F}} \mathcal{R}) \right] \right\} \det [\Delta_{\mathcal{F}\mathcal{F}}]^{-1/2}. \end{aligned} \quad (5.51)$$

Then the effective action for \mathcal{R} reads

$$S_{\text{eff}}[\mathcal{R}] = \frac{1}{2} \int \mathcal{R} \Delta_{\mathcal{R}\mathcal{R}} \mathcal{R} - \frac{1}{2} \int d^4x d^4y \Delta_{\mathcal{R}\mathcal{F}} \mathcal{R}(x) \Delta_{\mathcal{F}\mathcal{F}}^{-1}(x, y) \Delta_{\mathcal{R}\mathcal{F}} \mathcal{R}(y) + S_{\text{counter}}, \quad (5.52)$$

where S_{counter} denotes the contributions from the functional determinant and does not depend on \mathcal{R} . Note that instead of formally performing the Gaussian integral, we could solve for \mathcal{F} via its equations of motion

$$\Delta_{\mathcal{F}\mathcal{F}} \mathcal{F} = -\Delta_{\mathcal{R}\mathcal{F}} \mathcal{R}, \quad (5.53)$$

which has the formal solution $\mathcal{F} = -\Delta_{\mathcal{F}\mathcal{F}}^{-1} \Delta_{\mathcal{R}\mathcal{F}} \mathcal{R}$. Substituting this solution into (5.50) gives

$$S_{\text{eff}}[\mathcal{R}] = \frac{1}{2} \int \mathcal{R} \Delta_{\mathcal{R}\mathcal{R}} \mathcal{R} - \frac{1}{2} \int \Delta_{\mathcal{R}\mathcal{F}} \mathcal{R} \Delta_{\mathcal{F}\mathcal{F}}^{-1} \Delta_{\mathcal{R}\mathcal{F}} \mathcal{R}, \quad (5.54)$$

with the only difference from the formal result being the absence of S_{counter} . Thus at quadratic order, we can simply substitute the solution for \mathcal{F} back into the original action (5.49) to obtain the desired effective action for \mathcal{R} solely [36].

The effective action obtained in this manner is intrinsically non-local, as the operator $\Delta_{\mathcal{F}\mathcal{F}} = -\square + M_{\text{eff}}^2$ contains derivative operators. However, given that the adiabaticity condition [46]

$$\left| \frac{\ddot{\theta}}{\bar{\theta}} \right| \ll M_{\text{eff}} \quad (5.55)$$

is satisfied, we can expand the solution of (5.53) as a power series of M_{eff}^2 as

$$\mathcal{F} = (-\square + M_{\text{eff}}^2)^{-1} 2\dot{\theta} \frac{\dot{\phi}_0}{H} \dot{\mathcal{R}} = \frac{1}{M_{\text{eff}}^2} \left(1 + \frac{\square}{M_{\text{eff}}^2} + \dots \right) 2\dot{\theta} \frac{\dot{\phi}_0}{H} \dot{\mathcal{R}}, \quad (5.56)$$

making the non-local theory to the one with higher derivatives. Especially, the leading term gives a simple modified kinetic term:

$$S_2 = \int d^4x a^3 \epsilon m_{\text{Pl}}^2 \left[\left(1 + \frac{4\dot{\theta}^2}{M_{\text{eff}}^2} \right) \dot{\mathcal{R}}^2 - \frac{(\nabla \mathcal{R})^2}{a^2} \right], \quad (5.57)$$

which is interpreted as the speed of sound c_s^{-2} [23, 36, 47]. That is, the propagation speed of the adiabatic mode \mathcal{R} is reduced by the interaction with the heavy isocurvature mode \mathcal{F} , as the kinetic energy of \mathcal{R} is extracted to excite \mathcal{F} . Including higher order derivative operators in the expansion (5.56) leads to a modified dispersion relation in such a way that the validity of the effective theory is improved [48].

6 Power spectrum

6.1 Free solutions of mode functions and power spectrum

Having discussed various forms of the quadratic action for perturbations, now we solve the derived equations of motion for those perturbations. What we have seen previously is that in

general there are always interactions between different components. These mixing terms lead to the coupled set of differential equations, for example (4.39) and (4.40). Thus finding exact solutions of these coupled equations is rather non-trivial. But given that the interactions are small enough, we can perturbatively find the solutions and corrections to the power spectrum of the perturbation of our interest. Presuming that the interactions are sufficiently small, we conveniently split the quadratic part of the action as

$$S_2 = S_{2,\text{free}} + S_{2,\text{int}}, \quad (6.1)$$

where $S_{2,\text{free}}$ contains only free, decoupled terms for a certain component Ψ without interaction with other degrees of freedom, while $S_{2,\text{int}}$ includes quadratic interactions among them. The leading solution of a certain perturbation component Ψ is coming from the decoupled free quadratic action for Ψ , which has the following schematic form as a harmonic oscillator:

$$S_{\text{free}} = \int d\tau d^3x \frac{1}{2} \left[\Psi'^2 - (\nabla\Psi)^2 - m^2\Psi^2 \right], \quad (6.2)$$

with m^2 being time-dependent in general. To obtain this form we need to perform further manipulations. For example, for the decoupled free part of the quadratic action for the curvature perturbation \mathcal{R} [or equivalently π at quadratic order: see (5.47)] we rescale \mathcal{R} as

$$\Psi = \frac{a\dot{\phi}_0}{H}\mathcal{R} \equiv z\mathcal{R}, \quad (6.3)$$

which gives $m^2 = z''/z$. The resulting equation of motion we can derive from (6.2) is

$$\Psi'' - \Delta\Psi - m^2\Psi = 0. \quad (6.4)$$

We can write the Fourier mode $\Psi(\tau, \mathbf{k})$, which will be more convenient for the subsequent study, as

$$\Psi(\tau, \mathbf{x}) = \int \frac{d^3k}{(2\pi)^3} e^{i\mathbf{k}\cdot\mathbf{x}} \Psi(\tau, \mathbf{k}), \quad (6.5)$$

then the equation of motion (6.4) becomes, for the Fourier mode,

$$\Psi'' + (k^2 + m^2)\Psi = 0. \quad (6.6)$$

Identifying $k^2 + m^2 \equiv \omega_k^2(\tau)$, we can see that (6.6) describes a harmonic oscillator with time dependent frequency $\omega_k^2(\tau)$. Being a canonically normalized harmonic oscillator, we can follow the standard quantization procedure for Ψ . That is, we promote Ψ and the conjugate momentum $\Pi_\Psi = \Psi'$ to operators $\widehat{\Psi}$ and $\widehat{\Pi}_\Psi$ and imposes the canonical commutation relations between them.

Since Ψ is a free field, we can expand the operator $\widehat{\Psi}$ in terms of the creation and annihilation operators in the Fourier space. The Fourier mode given by (6.5) is promoted to the operator $\widehat{\Psi}(\tau, \mathbf{k})$, which we can expand in terms of the creation and annihilation operators:

$$\widehat{\Psi}(\tau, \mathbf{k}) = a(\mathbf{k})\Psi_k(\tau) + a^\dagger(-\mathbf{k})\Psi_k^*(\tau), \quad (6.7)$$

where the creation and annihilation operators satisfy the standard commutation relations

$$[a(\mathbf{k}), a^\dagger(\mathbf{q})] = (2\pi)^3 \delta^{(3)}(\mathbf{k} - \mathbf{q}), \quad (6.8)$$

otherwise zero. Now we require that the canonical conjugate variables $\widehat{\Psi}$ and $\widehat{\Pi}_\Psi$ satisfy the equal time canonical commutation relation

$$\left[\widehat{\Psi}(\tau, \mathbf{x}), \widehat{\Pi}_\Psi(\tau, \mathbf{y}) \right] = i\delta^{(3)}(\mathbf{x} - \mathbf{y}). \quad (6.9)$$

Using the Fourier mode (6.5) and the expansion (6.7) with the commutation relation (6.8), we can see that the mode function Ψ_k satisfies the normalization condition

$$\Psi_k \frac{d\Psi_k^*}{d\tau} - \frac{d\Psi_k}{d\tau} \Psi_k^* = i. \quad (6.10)$$

To determine the mode function $\Psi_k(\tau)$, which amounts to fix the vacuum state $|0\rangle$ defined by

$$a(\mathbf{k})|0\rangle = 0 \quad \text{for all } k, \quad (6.11)$$

we impose the vacuum boundary condition when the mode is deep inside the horizon $\tau \rightarrow -\infty$, i.e. $k \gg aH$ where the mode function solution is the positive frequency mode with $\omega_k = k$:

$$\Psi_k = \frac{1}{\sqrt{2k}} e^{-ik\tau}. \quad (6.12)$$

With this boundary condition and the normalization condition (6.10), the general solution for (6.6) can be written in terms of the Bessel functions, and for later convenience we use the Hankel function:

$$\Psi_k(\tau) = \sqrt{-\tau} \left[c_1(k) H_\nu^{(1)}(-k\tau) + c_2(k) H_\nu^{(2)}(-k\tau) \right], \quad (6.13)$$

where $c_1(k)$ and $c_2(k)$ are coefficients to be determined, and

$$\nu^2 \equiv \frac{9}{4} - \frac{m^2}{H^2} \quad (6.14)$$

is assumed to be constant. To fix the coefficients, we require that when the mode is deep inside the horizon $\tau \rightarrow -\infty$ we recover the vacuum boundary solution (6.12). This can be found by taking the argument of the Hankel function $-k\tau$ very large,

$$H_\nu^{(1)}(z) \xrightarrow{z \gg 1} \sqrt{\frac{2}{\pi z}} \exp \left[i \left(z - \frac{\pi}{2} \nu - \frac{\pi}{4} \right) \right], \quad (6.15)$$

with $H_\nu^{(2)}$ being the complex conjugate of $H_\nu^{(1)}$. Thus, to match (6.12), $c_1(k)$ and $c_2(k)$ should be

$$c_1(k) = \frac{\sqrt{\pi}}{2} e^{i(\nu+1/2)\pi/2} \quad \text{and} \quad c_2(k) = 0. \quad (6.16)$$

A particularly important and simple case is when $\nu = 3/2$ exactly, which corresponds to the massless limit:

$$\Psi_k = \frac{1}{\sqrt{2k}} \left(1 - \frac{i}{k\tau} \right) e^{-ik\tau}. \quad (6.17)$$

Given the free solution of Ψ_k , now we can compute the power spectrum of Ψ_k , defined by

$$\langle 0 | \Psi(\mathbf{k}) \Psi(\mathbf{q}) | 0 \rangle \equiv (2\pi)^3 \delta^{(3)}(\mathbf{k} + \mathbf{q}) \frac{2\pi^2}{k^3} \mathcal{P}_\Psi(k), \quad (6.18)$$

so that in terms of the mode function Ψ_k

$$\mathcal{P}_\Psi(k) = \frac{k^3}{2\pi^2} |\Psi_k|^2, \quad (6.19)$$

which can be evaluated at a convenient time, e.g. the moment of horizon crossing. For example, the power spectrum of the curvature perturbation can be directly found from its free quadratic action as

$$\mathcal{P}_\mathcal{R} = \frac{k^3}{2\pi^2} \left| \frac{\Psi_k}{z} \right|^2 = \left(\frac{H}{2\dot{\phi}_0} \right)^2 \frac{1}{2\epsilon m_{\text{Pl}}^2}. \quad (6.20)$$

Alternatively, using the kinematic basis discussed in Sections 3 and 4 in which \mathcal{R} is related to the tangential component u^T in the standard manner as

$$\mathcal{R} = -\frac{H}{\dot{\phi}_0} Q^T = -\frac{H}{\dot{\phi}_0} \frac{u^T}{a}, \quad (6.21)$$

we can write

$$\mathcal{P}_\mathcal{R}(k, \tau) = \frac{H^2}{\dot{\phi}_0^2} \mathcal{P}_{TT}(k, \tau) = \frac{k^3}{4\pi^2 a^2 m_{\text{Pl}}^2 \epsilon} |u_k^T|^2. \quad (6.22)$$

where m^2 for u^T is given by, as can be read from (4.39), $m^2 = a^2 M_{TT}^2 - \theta'^2$. Using the mass basis instead gives similar results.

6.2 Corrections to the power spectrum

In the previous section, we have obtained the free power spectrum with the quadratic interactions between different components being ignored. Now we consider the effects of the interactions to the power spectrum perturbatively. Having perturbative interactions, we can compute readily the corrections due to the interaction terms between different fields using the in-in formalism [49], which let us briefly recall here.

One crucial point when interactions exist, viz. the system is evolving, is that any (vacuum) expectation values should be taken with respect to the interaction vacuum state $|\Omega\rangle$, i.e. the *actual* vacuum state of the theory, *not* the free vacuum state $|0\rangle$ defined in the previous section. For this description, we resort to the interaction picture. In quantum mechanics, the interaction picture (or Dirac picture) is an intermediate between the Schrödinger picture and the Heisenberg picture. Whereas in the other two pictures either the state vector or the operators carry time dependence, in the interaction picture both carry part of the time dependence of observables. The purpose of the interaction picture is to shunt all the time dependence due to the free Hamiltonian H_0 onto the operators, leaving only the interaction Hamiltonian H_{int} affecting the time-dependence of the state vectors.

Now, we denote by $\langle \hat{\mathcal{O}}(t) \rangle$ the expectation value evaluated at a time t of a time-dependent operator

$$\hat{\mathcal{O}}(t) = \left(e^{-i \int_{t_{\text{in}}}^t H_0(t') dt'} \right)^\dagger \hat{\mathcal{O}} \left(e^{-i \int_{t_{\text{in}}}^t H_0(t'') dt''} \right), \quad (6.23)$$

where t_{in} is some early “in” time when the interaction is turned on. This expectation value is taken with respect to the vacuum state at that time $|\Omega(t)\rangle$ that has evolved from an “in” state,

which we take as the vacuum $|0\rangle$, according to

$$|\Omega(t)\rangle = e^{-i \int_{t_{\text{in}}}^t H_{\text{int}}(t') dt'} |0\rangle. \quad (6.24)$$

Given appropriate normalizations, now let us consider H_{int} as a small perturbation to the free Hamiltonian H_0 . Expanding the exponential in (6.24), in terms of H_{int} , we obtain⁷

$$\begin{aligned} \langle \widehat{\mathcal{O}}(t) \rangle &= \left\langle 0 \left| \left[1 - i \int_{t_{\text{in}}}^t H_{\text{int}}(t') dt' + \frac{1}{2} \left(-i \int_{t_{\text{in}}}^t H_{\text{int}}(t') dt' \right)^2 + \dots \right]^\dagger \widehat{\mathcal{O}}(t) \right. \right. \\ &\quad \left. \left. \times \left[1 - i \int_{t_{\text{in}}}^t H_{\text{int}}(t') dt' + \frac{1}{2} \left(-i \int_{t_{\text{in}}}^t H_{\text{int}}(t'') dt'' \right)^2 + \dots \right] \right| 0 \right\rangle \\ &= \langle 0 | \widehat{\mathcal{O}}(t) | 0 \rangle + i \int_{t_{\text{in}}}^t dt_1 \langle 0 | [H_{\text{int}}(t_1), \widehat{\mathcal{O}}(t)] | 0 \rangle \\ &\quad - \int_{t_{\text{in}}}^t dt_2 \int_{t_{\text{in}}}^{t_1} dt_2' \langle 0 | [H_{\text{int}}(t_1), [H_{\text{int}}(t_2'), \widehat{\mathcal{O}}(t)]] | 0 \rangle + \dots, \end{aligned} \quad (6.25)$$

where for the last equality we have used the commutator identity

$$AAB - 2ABA + BAA = [A, [A, B]]. \quad (6.26)$$

Having briefly recalled the in-in formalism, now we consider the corrections to the free power spectrum of Ψ due to the interaction terms between different fields, which we collectively denote by Φ . Schematically we write the interaction terms as

$$S_{2,\text{int}} = \int d^4x a^3 c(t) \mathcal{O}^{(\Phi)} \Phi \mathcal{O}^{(\Psi)} \Psi, \quad (6.27)$$

where $c(t)$ is the time-dependent coupling between Ψ and Φ , and $\mathcal{O}^{(X)}$ is the possible derivative operator for the field X . Since the interaction Hamiltonian contains two free field contents, the first non-vanishing contribution to the two-point function comes from the interaction Hamiltonian squared. Correspondingly, the correction terms now involve two free propagators: one for Ψ and the other for the coupled field Φ . Thus, in terms of the free propagator

$$i\Delta_{+-}^X(x, x') \equiv \langle 0 | X(x') X(x) | 0 \rangle, \quad (6.28)$$

where the subscript $+$ ($-$) denotes forward (backward) in time with $i\Delta_{+-} = -i\Delta_{-+}^*$, from (6.25) we can write as the two-point function of Ψ including the leading corrections due to its

⁷Here, we omit for simplicity $\lim_{t_{\text{in}} \rightarrow -\infty(1-i\varepsilon)}$ with $\varepsilon \ll 1$, but t_{in} is evaluated in this limit after all.

interaction with Φ as

$$\begin{aligned}
\langle \Omega | \Psi(t, \mathbf{x}) \Psi(t, \mathbf{y}) | \Omega \rangle &= i\Delta_{+-}^{\Psi} - \int_{t_i}^t dt_1 d^3x_1 \int_{t_i}^t dt_2 d^3x_2 (a^3c)(t_1)(a^3c)(t_2) \\
&\quad \times \mathcal{O}_1^{(\Psi)} i\Delta_{+-}^{\Psi}(t, t_1) \mathcal{O}_2^{(\Psi)} i\Delta_{-+}^{\Psi}(t, t_2) \mathcal{O}_1^{(\Phi)} \mathcal{O}_2^{(\Phi)} i\Delta_{+-}^{\Phi}(t_1, t_2) \\
&\quad - \int_{t_i}^t dt_1 d^3x_1 \int_{t_i}^{t_1} dt_2 d^3x_2 (a^3c)(t_1)(a^3c)(t_2) \\
&\quad \times \left[\mathcal{O}_1^{(\Psi)} i\Delta_{+-}^{\Psi}(t_1, t) \mathcal{O}_2^{(\Psi)} i\Delta_{+-}^{\Psi}(t, t_2) \mathcal{O}_1^{(\Phi)} \mathcal{O}_2^{(\Phi)} i\Delta_{+-}^{\Phi}(t_1, t_2) \right. \\
&\quad \left. + \mathcal{O}_1^{(\Psi)} i\Delta_{-+}^{\Psi}(t, t_1) \mathcal{O}_2^{(\Psi)} i\Delta_{-+}^{\Psi}(t_2, t) \mathcal{O}_1^{(\Phi)} \mathcal{O}_2^{(\Phi)} i\Delta_{+-}^{\Phi}(t_2, t_1) \right].
\end{aligned} \tag{6.29}$$

This is most general expression for the two-point correlation function of Ψ including the leading corrections due to the interaction with Φ . The corresponding power spectrum can be found by taking the Fourier transformation of $\langle \Omega | \Psi(t, \mathbf{x}) \Psi(t, \mathbf{y}) | \Omega \rangle$.

6.2.1 Two-field case

For definiteness, let us consider explicitly the corrections to the power spectrum of \mathcal{R} due to the interaction with Q_{\perp}^a in the simplest two-field case. The total quadratic action including the interaction is (5.49), from which we read the coefficient of the quadratic mixing term and the derivative operators as

$$c(t) = \sqrt{8\epsilon} m_{\text{Pl}} \dot{\theta}, \quad \mathcal{O}^{(\mathcal{R})} = \partial_t, \quad \mathcal{O}^{(\mathcal{F})} = 1. \tag{6.30}$$

Then the two-point correlation of \mathcal{R} is given by

$$\begin{aligned}
\langle \Omega | \mathcal{R}(t, \mathbf{x}) \mathcal{R}(t, \mathbf{y}) | \Omega \rangle &= i\Delta_{+-}^{\mathcal{R}} - \int_{t_i}^t dt_1 d^3x_1 \int_{t_i}^t dt_2 d^3x_2 (a^3c)(t_1)(a^3c)(t_2) \\
&\quad \times \partial_{t_1} i\Delta_{+-}^{\mathcal{R}}(t, t_1) \partial_{t_2} i\Delta_{-+}^{\mathcal{R}}(t, t_2) i\Delta_{+-}^{\mathcal{F}}(t_1, t_2) \\
&\quad - \int_{t_i}^t dt_1 d^3x_1 \int_{t_i}^{t_1} dt_2 d^3x_2 (a^3c)(t_1)(a^3c)(t_2) \\
&\quad \times \left[\partial_{t_1} i\Delta_{+-}^{\mathcal{R}}(t_1, t) \partial_{t_2} i\Delta_{+-}^{\mathcal{R}}(t, t_2) i\Delta_{+-}^{\mathcal{F}}(t_1, t_2) \right. \\
&\quad \left. + \partial_{t_1} i\Delta_{-+}^{\mathcal{R}}(t, t_1) \partial_{t_2} i\Delta_{-+}^{\mathcal{R}}(t_2, t) i\Delta_{+-}^{\mathcal{F}}(t_2, t_1) \right].
\end{aligned} \tag{6.31}$$

The power spectrum is obtained by performing the Fourier transformation of the above result. In the very simple limit of constant $c(t)$, which corresponds to a constant turn, using the massive mode function solution (6.13) with (6.16) for \mathcal{F} , we find

$$\mathcal{P}_{\mathcal{R}} = \left(\frac{H}{2\pi} \right)^2 \frac{1}{2\epsilon m_{\text{Pl}}^2} \left(1 + \frac{4c^2 C}{\epsilon m_{\text{Pl}}^2 H^2} \right), \tag{6.32}$$

where [50]

$$\begin{aligned}
C \equiv \frac{\pi}{4} \Re \left\{ \int_0^{\infty} dx_1 \int_{x_1}^{\infty} dx_2 \left[x_1^{-1/2} H_{\nu}^{(1)}(x_1) e^{ix_1} x_2^{-1/2} H_{\nu}^{(2)}(x_2) e^{-ix_2} \right. \right. \\
\left. \left. - x_1^{-1/2} H_{\nu}^{(1)}(x_1) e^{-ix_1} x_2^{-1/2} H_{\nu}^{(2)}(x_2) e^{-ix_2} \right] \right\}.
\end{aligned} \tag{6.33}$$

Especially, if \mathcal{F} is also very light so that $\nu \rightarrow 3/2$, we can perform the integral analytically using the massless mode function solution (6.17) to find

$$C = \frac{1}{2}(\alpha - \log x)^2 + \frac{\pi^2}{8} - \frac{\gamma^2}{2} - \frac{3}{4} = \frac{\alpha^2}{2} + \frac{\pi^2}{8} - \frac{\gamma^2}{2} - \frac{3}{4} \Big|_{k=aH}, \quad (6.34)$$

where $\alpha \equiv 2 - \log 2 - \gamma \approx 0.729637$ with $\gamma \approx 0.577216$ being the Euler-Mascheroni constant, and for the second equality we have evaluated at the moment of horizon crossing $k = aH^8$ [see also (6.46)]. For the field fluctuations Q^a in the flat gauge, we can find similar results.

6.3 Alternative solutions

In the previous sections, we have considered the free, decoupled and interaction terms separately, and have regarded the latter as perturbations to apply the in-in formalism. This is because fully solving the coupled differential equations is a non-trivial task. We can, however, take a different approach to solve the equations of motion directly. As we will see, we still treat the interaction terms perturbatively, so in this sense it is equivalent to what we have seen in the previous sections. One advantage of the alternative we present here is that all the corrections, including the deviations from perfect de Sitter limit $\tau = -1/(aH)$, are systematically taken into account. We will solve for the field fluctuations Q^a in the flat gauge, but it is straightforwardly extended to the curvature perturbation \mathcal{R} in the comoving gauge.

Our starting point is the equation of motion for Q^a (4.21). Defining $u^a \equiv aQ^a$ and

$$x \equiv -k\tau = \frac{k}{aH} (1 + \epsilon + \dots), \quad (6.35)$$

we can rewrite (4.21) as [52]

$$D_x^2 u^a + \left(1 - \frac{2}{x^2}\right) u^a = \frac{3}{x^2} \zeta^a{}_b u^b, \quad (6.36)$$

where ζ^{ab} is given by

$$\zeta^{ab} = G^{ab}\epsilon + \frac{\dot{\phi}_0^a \dot{\phi}_0^b}{H H} + \frac{\mathbb{R}^a{}_{cd} \dot{\phi}_0^c \dot{\phi}_0^d}{3 H H} - \frac{V^{ab}}{3H^2} + \dots. \quad (6.37)$$

Since u^a canonically normalizes (4.19), we can apply the canonical commutation relations between u^a and its conjugate momentum $D_\tau u^a$:

$$[u^a(\tau, \mathbf{x}), D_\tau u^b(\tau, \mathbf{y})] = i\delta^{(3)}(\mathbf{x} - \mathbf{y})G^{ab}, \quad (6.38)$$

otherwise zero.

What we can see from the rescaled equation of motion (6.36) is that the left hand side describes the evolution of a single component u^a in perfect de Sitter background, while everything else – including the deviations from perfect de Sitter background and mixing with other

⁸In fact, if we integrate the outermost integral of (6.33) from $x_e \equiv -k\tau_e$, we may interpret $-\log(-k\tau_e) = N_k$ as the number of e -folds elapsed between the moment of horizon crossing for the mode of our interest and the end of inflation τ_e [51]. This seems diverging as $x_e \rightarrow 0$, but see the discussion below (6.44).

components – is placed on the right hand side. Thus we can solve the homogeneous part of (6.36),

$$D_x^2 u_0^a + \left(1 - \frac{2}{x^2}\right) u_0^a = 0, \quad (6.39)$$

very easily with the boundary conditions, as in the previous sections, being imposed at $x \rightarrow \infty$ as

$$u_0^a(x, \mathbf{k}) = a^a(\mathbf{k})u_0(x) + a^{a\dagger}(-\mathbf{k})u_0^*(x), \quad (6.40)$$

where the creation and annihilation operators satisfy the commutation relations

$$\left[a^a(\mathbf{k}), a^{b\dagger}(\mathbf{q}) \right] = (2\pi)^3 \delta^{(3)}(\mathbf{k} - \mathbf{q}) G^{ab} \quad (6.41)$$

and the mode function solution $u_0(x)$ is that of a massless field (6.17):

$$u_0(x) = \frac{1}{\sqrt{2k}} \left(1 + \frac{i}{x}\right) e^{ix}. \quad (6.42)$$

With these boundary conditions, along with the same normalization condition as (6.10), we can apply the standard Green's function method to solve the inhomogeneous equation as [51]

$$u^a(x) = u_0^a(x) + \frac{3}{2}i \int_x^\infty \frac{du}{u^2} \zeta^a{}_b u^b(u) [u_0^*(u)u_0(x) - u_0^*(x)u_0(u)]. \quad (6.43)$$

The first correction terms are obtained by computing the integral with $u^b(u)$ being the homogeneous solution u_0 , and the next corrections by plugging the solution with leading correction terms, and so on. We can iterate as many times as we like to find more and more accurate solutions.

To implement the slow-roll approximation evaluated at, say, the moment of horizon crossing $k = aH$ which is different from $x = 1$ by $\mathcal{O}(\epsilon)$ as can be read from (6.35), we take the ansatz $\zeta^{ab} = \zeta^{ab}(\log x)$ and expand ζ^{ab} as a power series in $\log x$:

$$\zeta^{ab} = \sum_{n=0}^{\infty} \frac{\zeta_{n+1}^{ab}}{n!} (\log x)^n, \quad (6.44)$$

with $\zeta_n^{ab} = \mathcal{O}(\epsilon^n)$. This expansion is valid as long as the series converges, which is the case for a wide range of value for x as $e^{-1/\mathcal{O}(\epsilon)} \ll x \ll e^{1/\mathcal{O}(\epsilon)}$. We can substitute the expansion ansatz (6.44) into the Green's function solution (6.43) and integrate iteratively to find the desired solution for φ^a explicitly in terms of the slow-roll parameters. Especially, we can find the asymptotic solution at later time $x \rightarrow 0$ as

$$u^a(x) \xrightarrow{x \rightarrow 0} \frac{i}{\sqrt{2kx}} \left\{ a^a(\mathbf{k}) - a^{a\dagger}(\mathbf{k}) + \left[\left(\alpha + \frac{i\pi}{2} \right) \zeta_{1b}^a - \zeta_{1b}^a \log x + \dots \right] a^b(\mathbf{k}) - \left[\left(\alpha + \frac{i\pi}{2} \right) \zeta_{1b}^a - \zeta_{1b}^a \log x + \dots \right]^* a^{b\dagger}(-\mathbf{k}) + \dots \right\}, \quad (6.45)$$

with α being the same numerical factor as encountered in (6.34). This gives

$$\begin{aligned} \langle Q^a(\mathbf{k})Q^b(\mathbf{q}) \rangle &= \frac{1}{2ka^2x^2} \delta^{(3)}(\mathbf{k} + \mathbf{q}) (G^{ab} + 2\alpha\zeta_1^{ab} - 2\zeta_1^{ab} \log x + \dots) \\ &= \frac{H^2}{2k^3} \delta^{(3)}(\mathbf{k} + \mathbf{q}) \left[(1 - 2\epsilon)G^{ab} + 2\alpha\zeta_1^{ab} + \dots \right] \Big|_{k=aH}, \end{aligned} \quad (6.46)$$

where for the second equality we have evaluated the right hand side at the moment of horizon crossing $k = aH$. While we can proceed as much accurate as we like, what we have obtained is the power spectrum of the field fluctuations in the flat gauge, *not* that of the curvature perturbation \mathcal{R} . Thus we need additional manipulation that relates Q^a and \mathcal{R} . Unlike single field case, in multi-field case it is non-trivial as we have seen from (5.38) and (5.44). Fortunately, there is a very simple geometric identity called the δN formalism that relates Q^a and \mathcal{R} in a very straightforward manner: see (7.44). We postpone more detailed discussions on the δN formalism to the following section.

7 Towards non-linear perturbations

Up to now, we have considered linear cosmological perturbations. They are described by quadratic action presented in many different forms in the previous section like (5.39). The structure of the action is, with appropriate manipulations, essentially that of a harmonic oscillator, so we can directly apply our conventional wisdom in quantum field theory. Being described as quantum harmonic oscillators, linear cosmological perturbations are free, leading to the conclusion that the power spectra are the only non-vanishing correlation functions. The rapid observational advances during the last decade, especially the most recent Planck mission for the CMB temperature anisotropies, have constrained the properties of the power spectra as [12]:

$$\log(10^{10}\mathcal{P}_{\mathcal{R}}) = 3.094 \pm 0.034, \quad (7.1)$$

$$n_{\mathcal{R}} \equiv \frac{d \log \mathcal{P}_{\mathcal{R}}}{d \log k} + 1 = 0.9645 \pm 0.0049, \quad (7.2)$$

$$r \equiv \frac{\mathcal{P}_T}{\mathcal{P}_{\mathcal{R}}} \lesssim 0.12. \quad (7.3)$$

Indeed, these observational constraints have been very powerful in discriminating different models of inflation and in ruling out the models inconsistent with observations. For example, the simple quartic potential model is not favoured well beyond 2σ level, since this model predict too large tensor-to-scalar ratio of $r \gtrsim 0.2$.

With further developments in observations – including the polarization of the CMB, gravitational waves and distribution of galaxies on large scales – that allow us to probe a wide range of scales, we can hope to make use of higher-order correlation functions to constrain more tightly inflation models and to probe inflationary dynamics. The higher-order correlation functions, starting from three-point function or its Fourier transform, the bispectrum, incorporate non-linear perturbations described beyond quadratic action of the cosmological perturbations. Thus non-linear perturbations also tell us the “interactions” among the degrees of freedom

participating in inflation, revealing the important physics underlying inflation that cannot be extracted from the power spectra⁹. The amplitude of the bispectrum, the first non-zero probe of non-linear perturbations, is conveniently parametrized by the so-called non-linear parameter f_{NL} [53], which is roughly speaking the ratio of the bispectrum to the power spectrum squared, and the current observational constraints from the CMB on different configurations of the bispectrum are [54]

$$\begin{aligned} f_{\text{NL}}^{\text{local}} &= 2.5 \pm 5.7 & (0.8 \pm 5.0), \\ f_{\text{NL}}^{\text{equil}} &= -16 \pm 70 & (-4 \pm 43), \\ f_{\text{NL}}^{\text{ortho}} &= -34 \pm 33 & (-26 \pm 21), \end{aligned} \tag{7.4}$$

where the numbers in the parentheses are obtained by combining the polarization data. So currently our universe is largely consistent with free, linear, Gaussian cosmological perturbations. The importance of non-linear perturbations however does not diminish and we can expect to further constrain or even detect the deviation from perfect Gaussian nature of the primordial perturbations, viz. the primordial non-Gaussianity in near future. In this section, we discuss how to describe non-linear perturbations in multi-field inflation.

7.1 Issue of mapping

An important point we have to take care of when we discuss non-linear field fluctuations in multi-field inflation in the flat gauge is that, as emphasized previously, in the field space ϕ^a plays the role of coordinates. Thus, as we do in general relativity, it is of crucial convenience to maintain the covariance of the formulation: the field fluctuations $\delta\phi^a$ around the background trajectory $\phi_0^a = \phi_0^a(t)$ are coordinate dependent – for example, we are free to choose the time slicing condition as $\delta\phi^1 = 0$ – and is thus not covariant. How then to formulate $\delta\phi^a$ covariantly? We begin by noting that the two points, $\phi_0^a(t)$ and $\phi^a = \phi_0^a + \delta\phi^a$, can be connected by a unique geodesic with respect to the field space metric G_{ab} [27, 28]. This geodesic is parametrized by λ that runs from 0 to ε , corresponding to the two endpoints of the geodesic, ϕ_0^a and ϕ^a respectively. Here, ε is a bookkeeping parameter to count the order of perturbation and we will set it to unity in the end. To specify the geodesic, we need the initial point ϕ_0^a and its velocity, which we denote by Q^a . This situation is depicted in Figure 6.

Denoting the covariant derivative with respect to λ by $D_\lambda \equiv D/d\lambda$, we can write the geodesic equation for $\phi^a(\lambda)$ as

$$D_\lambda^2 \phi^a = \frac{d^2 \phi^a}{d\lambda^2} + \Gamma_{bc}^a \frac{d\phi^b}{d\lambda} \frac{d\phi^c}{d\lambda} = 0, \tag{7.5}$$

with the following the initial conditions at $\lambda = 0$:

$$\phi^a|_{\lambda=0} = \phi_0^a, \tag{7.6}$$

$$D_\lambda \phi^a|_{\lambda=0} = \left. \frac{d\phi^a}{d\lambda} \right|_{\lambda=0} \equiv Q^a. \tag{7.7}$$

Expanding $\phi^a(\lambda = \varepsilon)$ as a power series with respect to ε from $\lambda = 0$, we find

$$\phi^a(\lambda = \varepsilon) = \phi^a|_{\lambda=0} + \left. \frac{d\phi^a}{d\lambda} \right|_{\lambda=0} \varepsilon + \frac{1}{2!} \left. \frac{d^2 \phi^a}{d\lambda^2} \right|_{\lambda=0} \varepsilon^2 + \frac{1}{3!} \left. \frac{d^3 \phi^a}{d\lambda^3} \right|_{\lambda=0} \varepsilon^3 + \dots \tag{7.8}$$

⁹Of course higher-order correlation functions may well contribute to the power spectrum as loop corrections.

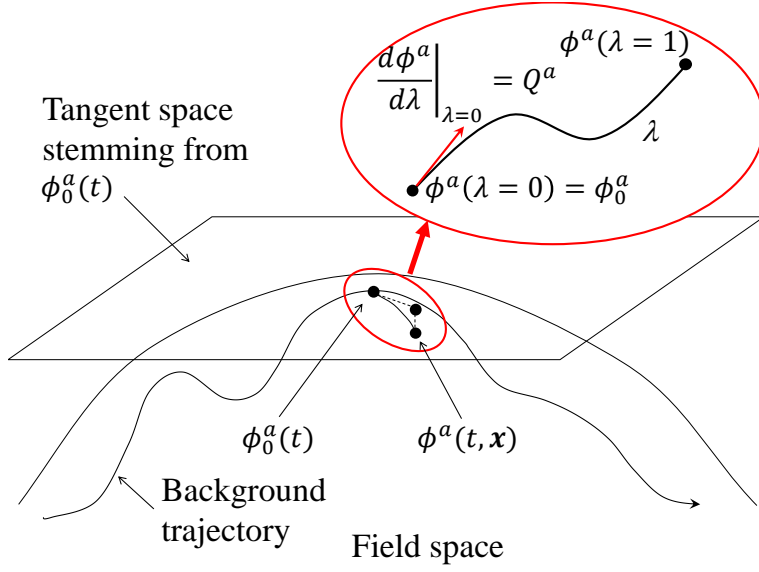


Figure 6: A schematic figure showing a physical field ϕ^a in the field space around the background trajectory $\phi_0^a(t)$. The geodesic connecting ϕ^a and ϕ_0^a is parametrized λ , which runs from 0 to ε .

Since the derivatives with respect to λ here are *not* covariant ones, we can trade quadratic and higher derivatives with single ones by using the geodesic equation (7.5). That is, we can replace the quadratic derivative with

$$\frac{d^2\phi^a}{d\lambda^2} = -\Gamma_{bc}^a \frac{d\phi^b}{d\lambda} \frac{d\phi^c}{d\lambda}, \quad (7.9)$$

and the third order derivative with

$$\frac{d^3\phi^a}{d\lambda^3} = (\Gamma_{de}^a \Gamma_{bc}^e - \Gamma_{bc;d}^a) \frac{d\phi^b}{d\lambda} \frac{d\phi^c}{d\lambda} \frac{d\phi^d}{d\lambda}, \quad (7.10)$$

and so on. Thus, we can write (7.8) as

$$\phi^a(\lambda = \varepsilon) = \phi_0^a + Q^a \varepsilon - \frac{1}{2} \Gamma_{bc}^a Q^b Q^c \varepsilon^2 + \frac{1}{6} (\Gamma_{de}^a \Gamma_{bc}^e - \Gamma_{bc;d}^a) Q^b Q^c Q^d \varepsilon^3 + \dots, \quad (7.11)$$

Setting $\varepsilon = 1$, we finally obtain

$$\phi^a - \phi_0^a \equiv \delta\phi^a = Q^a - \frac{1}{2} \Gamma_{bc}^a Q^b Q^c + \frac{1}{6} (\Gamma_{de}^a \Gamma_{bc}^e - \Gamma_{bc;d}^a) Q^b Q^c Q^d + \dots. \quad (7.12)$$

We can see that at linear order, the field fluctuations $\delta\phi^a$ and the vector Q^a are identical. However, going beyond linear order they are manifestly different as we can see from the above equation. Only when we write the equations in terms of Q^a , they can be expressed in a covariant manner.

7.2 Cubic order action

Having found the covariant description of the field fluctuations, we can straightly calculate the third and higher-order action of the field fluctuations Q^a in the flat gauge. Given the

higher-order action of our interest, we can use the in-in formalism discussed in the previous section to compute the corresponding higher-order correlation functions. For example, to compute the bispectrum, we need the third order action for the first non-zero contributions. It is straightforward after some arrangement to find the third order action as

$$S_3 = \int d^4x a^3 \left\{ 3m_{\text{Pl}}^2 H^2 \alpha^3 + 2m_{\text{Pl}}^2 H \alpha^2 \frac{\Delta}{a^2} \chi - \frac{m_{\text{Pl}}^2}{2a^4} [\chi^{,ij} \chi_{,ij} - (\Delta \chi)^2] \alpha \right. \\ \left. + (g_1)_{abc} Q^a Q^b Q^c + (g_2)_{abc} D_t Q^a Q^b Q^c + (g_3)_{abc} D_t Q^a D_t Q^b Q^c \right. \\ \left. + \alpha G_{ab} \dot{\phi}_0^a \partial_i Q^b \partial^i \chi - G_{ab} D_t Q^a \partial_i Q^b \partial^i \chi - \frac{1}{2} \alpha G_{ab} \frac{\partial^i Q^a \partial_i Q^b}{a^2} \right\}, \quad (7.13)$$

where the coefficients $(g_i)_{abc}$ are given by

$$(g_1)_{abc} = \frac{1}{6} \left(\mathbb{R}_{dabe;c} \dot{\phi}_0^d \dot{\phi}_0^e + V_{abc} \right) + \frac{1}{2} \mathcal{N}_c \left(-\mathbb{R}_{dabe} \dot{\phi}_0^d \dot{\phi}_0^e + V_{ab} \right) - \frac{\dot{\phi}_0^2}{2} \mathcal{N}_a \mathcal{N}_b \mathcal{N}_c, \quad (7.14)$$

$$(g_2)_{abc} = \frac{1}{6} \left(\mathbb{R}_{dbca} + 3\mathbb{R}_{abcd} \right) \dot{\phi}_0^d + \dot{\phi}_0^a \mathcal{N}_b \mathcal{N}_c, \quad (7.15)$$

$$(g_3)_{abc} = -\frac{1}{2} G_{ab} \mathcal{N}_c, \quad (7.16)$$

and α (and \mathcal{N}_a) and χ are the linear solutions of the lapse and shift given by (4.11) and (4.12) respectively. Note that the first three terms of (7.13) are coming from the gravity sector, while the rest from the matter sector. From (7.13) we can compute the bispectrum of the field fluctuations Q^a , evaluated at the moment of the horizon crossing. We can calculate the bispectrum of the curvature perturbation by implementing the δN formalism: see [25, 28].

In the comoving gauge, we have already singled out the curvature perturbation as (5.46) in a non-perturbative manner. So we can straightforwardly expand the action in terms of \mathcal{R} and the orthogonal modes Q_\perp^a . The only caution is that the Goldstone mode π , in terms of which the quadratic action (5.45) is written, is related to \mathcal{R} non-linearly as (5.47) so that we have additional non-linear contributions from the quadratic action. Then we can find the leading action cubic in \mathcal{R} the same as the standard single field case [43]:

$$S_{\mathcal{R}\mathcal{R}\mathcal{R}} = \int d^4x a^3 m_{\text{Pl}}^2 \left[(-\epsilon^2 + 2\epsilon\delta_1) \mathcal{R} \dot{\mathcal{R}}^2 - 2\epsilon^2 \dot{\mathcal{R}} \partial^i \mathcal{R} \Delta^{-1} \partial_i \dot{\mathcal{R}} + (3\epsilon^2 - 2\epsilon\delta_1) \mathcal{R} \frac{(\nabla \mathcal{R})^2}{a^2} \right. \\ \left. + \frac{1}{4} \left(\frac{\epsilon V_{ab} \dot{\phi}_0^a \dot{\phi}_0^b}{H^2} + \frac{V_{abc} \dot{\phi}_0^a \dot{\phi}_0^b \dot{\phi}_0^c}{3H^3} + \frac{G_{ab} \ddot{\phi}_0^a \ddot{\phi}_0^b}{H^2} \right) \mathcal{R}^3 \right] + \mathcal{O}(\epsilon^3). \quad (7.17)$$

Meanwhile, the π - Q_\perp^a mixing term in the quadratic action leads to the following $\mathcal{R}^2 Q_\perp$ mixing in the cubic order action:

$$S_{\text{quadratic mixing}} = \int d^4x a^3 \frac{V_a Q_\perp^a}{H} \left\{ (3\epsilon - \delta_1) \mathcal{R} \dot{\mathcal{R}} + \frac{1}{H} \left(\mathcal{R} \ddot{\mathcal{R}} + \dot{\mathcal{R}}^2 \right) - \frac{1-\epsilon}{4a^2 H} [(\nabla \mathcal{R})^2 - \partial^i \partial^j \Delta^{-1} (\partial_i \mathcal{R} \partial_j \mathcal{R})] \right. \\ \left. + \frac{1}{2a^2 H^2} \left[\partial^i \mathcal{R} \partial_i \dot{\mathcal{R}} - \partial^i \partial^j \Delta^{-1} (\partial_i \mathcal{R} \partial_j \dot{\mathcal{R}}) \right] \right. \\ \left. + \frac{\epsilon}{H} (1 - \partial^i \partial^j \Delta^{-1}) \left(\partial_i \dot{\mathcal{R}} \Delta^{-1} \partial_j \mathcal{R} + \partial_i \mathcal{R} \Delta^{-1} \partial_j \dot{\mathcal{R}} \right) \right\} + \mathcal{O}(\epsilon^2). \quad (7.18)$$

We can also find the other mixing terms $-Q_\perp^3$, $\mathcal{R}Q_\perp^2$ and \mathcal{R}^2Q_\perp – straightly as

$$\begin{aligned}
S_{\text{cubic mixing}} = \int d^4x a^3 \left\{ -\frac{V_{abc}}{6} Q_\perp^a Q_\perp^b Q_\perp^c - \frac{V_{abc}}{6H^2} \left(Q_\perp^a \dot{\phi}_0^b \dot{\phi}_0^c + \dot{\phi}_0^a Q_\perp^b \dot{\phi}_0^c + \dot{\phi}_0^a \dot{\phi}_0^b Q_\perp^c \right) \mathcal{R}^2 \right. \\
+ \frac{V_{abc}}{6H} \left(Q_\perp^a Q_\perp^b \dot{\phi}_0^c + Q_\perp^a \dot{\phi}_0^b Q_\perp^c + \dot{\phi}_0^a Q_\perp^b Q_\perp^c \right) \mathcal{R} \\
+ \epsilon \mathcal{R} \left[-\frac{2}{m_{\text{Pl}}^2 H^2} (V_a Q_\perp^a)^2 + \frac{2}{m_{\text{Pl}}^2 H^2} (\Delta^{-1} \partial_i \partial_j (V_a Q_\perp^a))^2 \right. \\
- 4\epsilon \mathcal{R} (V_a Q_\perp^a) + \frac{2\epsilon}{H} \dot{\mathcal{R}} (V_a Q_\perp^a) - \frac{2\epsilon}{H} \left(\Delta^{-1} \partial^i \partial^j \dot{\mathcal{R}} \right) \Delta^{-1} \partial_i \partial_j (V_a Q_\perp^a) \\
- \frac{1}{2H} \dot{\mathcal{R}} (V_a Q_\perp^a) + \frac{1}{2H} \mathcal{R} \left(3H V_a Q_\perp^a + V_a \dot{Q}_\perp^a \right) \\
+ \frac{1}{2} G_{ab} \left(\dot{Q}_\perp^a \dot{Q}_\perp^b - \partial^i Q_\perp^a \partial_i Q_\perp^b \right) - \frac{1}{2} M_{ab}^2 \left(Q_\perp^a Q_\perp^b - \frac{2\dot{\phi}_0^a}{H} \mathcal{R} Q_\perp^b \right) \left. \right] \\
+ \left[G_{ab} \partial^i Q_\perp^a \dot{Q}_\perp^b - \frac{1}{H} \left(\partial^i \mathcal{R} V_a Q_\perp^a + \mathcal{R} V_a \partial^i Q_\perp^a \right) \right] \Delta^{-1} \partial_i \left(\epsilon \dot{\mathcal{R}} - \frac{V_a Q_\perp^a}{m_{\text{Pl}}^2 H} \right) \\
+ \frac{2}{H} \left(\epsilon \dot{\mathcal{R}} \partial^i \mathcal{R} - \delta_1 H \mathcal{R} \partial^i \mathcal{R} \right) \Delta^{-1} \partial_i (V_a Q_\perp^a) \left. \right\}. \tag{7.19}
\end{aligned}$$

Again, we can straightly compute the bispectrum of the curvature perturbation and the corrections from the interaction with the orthogonal modes Q_\perp^a using the in-in formalism as we did in the previous section.

7.3 Large-scale approximation

Until now, we have considered the standard cosmological perturbation theory. Here, “standard” means that our perturbative expansion is based on the order of perturbations we are interested in. For example, with the typical order of perturbations being denoted by ζ , for linear perturbation theory we are interested in $\mathcal{O}(\zeta)$ and drop higher-order contributions, for second-order theory $\mathcal{O}(\zeta^2)$, and so on. While perfectly legitimate, it is not the only and most convenient approach. Especially, if we are interested in non-linear perturbations, we have to go to the higher-order action or Einstein equations of our interest since the contributions of, say, second-order perturbations are otherwise not captured. Furthermore, as we have seen, finding the higher-order action is very tedious and usually requires further manipulations including integrations by parts and the background equation of motion.

In this section, we briefly discuss an alternative approach based on large-scale approximation. The benefit of this approach is twofold. During inflation, physical scales expand faster than the Hubble horizon. Thus all scales of our observational interest were once far outside horizon during inflation so that taking large-scale approximation for them is making very good sense to track their evolution on super-horizon scales. More importantly, the fully non-linear perturbation equations to be solved become surprisingly simple *irrespective* of the order of perturbations of our interest. This means we can follow the full non-linear evolution of perturbations, not necessarily order by order as we do in the standard manner.

7.3.1 Non-linear equations

First we decompose the energy-momentum tensor as measured by observers moving orthogonally (four-velocity is identical to the normal vector) to the slices set by the 3+1 decomposition of the ADM metric. Naturally, we can interpret the energy density \mathcal{E} as the pure temporal part, i.e. two indices are all contracted with the normal vector, the momentum density \mathcal{J}^i as one index contracted with the normal vector while the other projected, and the spatial energy-momentum tensor \mathcal{S}_{ij} as the pure spatial part, i.e. two indices are all projected. In the ADM decomposition, we can write the unit timelike vector normal to the constant-time hypersurface n^μ has the components

$$n^\mu = \left(\frac{1}{\mathcal{N}}, \frac{\beta^i}{\mathcal{N}} \right), \quad (7.20)$$

$$n_\mu = (-\mathcal{N}, 0). \quad (7.21)$$

With this, if we choose to write in the simplest forms, we have

$$\mathcal{E} = \mathcal{N}^2 T^{00}, \quad (7.22)$$

$$\mathcal{J}_i = \mathcal{N} T^0_i, \quad (7.23)$$

$$\mathcal{S}_{ij} = T_{ij}. \quad (7.24)$$

In the ADM decomposition scheme, we can also write the Einstein equation in terms of the components orthogonal and tangential to the constant-time hypersurface [34]. The two constraint equations follow from those involving temporal components, i.e. both or one of the indices of the Einstein equation is contracted with the normal vector. The geometric parts of them are the well-known Gauss-Codazzi equations, and the corresponding matter parts are precisely the energy and momentum density \mathcal{E} and \mathcal{J}_i found above. Further, these equations do not involve explicit time derivatives. Thus, they are equations of constraints which must be satisfied by the fundamental ADM variables, γ_{ij} and $\dot{\gamma}_{ij}$, at all times. We can obtain the energy and the momentum constraints as

$$R^{(3)} + \frac{2}{3}K^2 - \bar{K}^i_j \bar{K}^j_i = \frac{2}{m_{\text{Pl}}^2} \mathcal{E}, \quad (7.25)$$

$$\bar{K}^j_{i;j} - \frac{2}{3}K_{;i} = \frac{\mathcal{J}_i}{m_{\text{Pl}}^2}, \quad (7.26)$$

respectively, where

$$\bar{K}_{ij} \equiv K_{ij} - \frac{1}{3}\gamma_{ij}K \quad (7.27)$$

is the traceless part of K_{ij} . The dynamical equations thus involve the spatial energy-momentum tensor \mathcal{S}_{ij} . They can be formally written as the Lie derivatives of γ_{ij} and K_{ij} , or more explicitly, the trace and traceless evolution equations can be found as

$$K_{,0} - \beta^i K_{,i} = -\mathcal{N}^{;i}_{;i} + \mathcal{N} \left(R^{(3)} + K^2 + \frac{1}{2m_{\text{Pl}}^2} \mathcal{S} - \frac{3}{2m_{\text{Pl}}^2} \mathcal{E} \right), \quad (7.28)$$

$$\bar{K}^i_{j,0} - \beta^k \bar{K}^i_{j,k} + \beta^i_{,k} \bar{K}^k_j - \beta^k_{,j} \bar{K}^i_k = -\mathcal{N}^{;i}_{;j} + \frac{1}{3} \delta^i_j \mathcal{N}^{;k}_{;k} + \mathcal{N} \left(\bar{R}^{(3)i}_j + K \bar{K}^i_j - \frac{1}{m_{\text{Pl}}^2} \bar{\mathcal{S}}^i_j \right), \quad (7.29)$$

where the trace and traceless parts of \mathcal{S}_{ij} , denoted by \mathcal{S} and $\bar{\mathcal{S}}_{ij}$, are given in the same way as K_{ij} , i.e. $\mathcal{S} \equiv \gamma^{ij}\mathcal{S}_{ij}$ and $\bar{\mathcal{S}}_{ij} \equiv \mathcal{S}_{ij} - \gamma_{ij}\mathcal{S}/3$. The local energy and momentum conservation equations, given by $T^{0\mu}{}_{;\mu} = 0$ and $T^\mu{}_{i;\mu} = 0$ are

$$\mathcal{E}_{,0} - \beta^i \mathcal{E}_{,i} = \mathcal{N}K \left(\mathcal{E} + \frac{\mathcal{S}}{3} \right) + \mathcal{N}\bar{K}^i{}_j \bar{\mathcal{S}}^j{}_i + \frac{1}{\mathcal{N}} (N^2 \mathcal{J}^i)_{;i}, \quad (7.30)$$

$$\mathcal{J}_{i,0} - \beta^j \mathcal{J}_{i,j} - \beta^j{}_{,i} \mathcal{J}_j = \mathcal{N}K \mathcal{J}_i - (\mathcal{E} \delta^j{}_i + \mathcal{S}^j{}_i) \mathcal{N}_{;j} - \mathcal{N} \mathcal{S}^j{}_{i;j}. \quad (7.31)$$

With these non-linear equations with the energy-momentum tensor given by (3.2), we now can apply the large-scale approximation. That is, we assume a certain smoothing scale $1/k$ greater than which we can approximate well the actual observable universe. Then identifying

$$\varepsilon \equiv \frac{k}{aH} \quad (7.32)$$

as the fictitious parameter associated with a spatial gradient ∂_i , we can expand the exact non-linear equations as a power series in ε . For more careful accounts on the conditions for this expansion, see e.g. [42].

Before we proceed further, let us pause for a minute and consider the geometry set by the time slicing. A choice of a time coordinate determines a family of constant-time hypersurfaces Σ in the perturbed space-time, which we refer to as time slicing. With each time slicing there are three important geometrical quantities: namely, the intrinsic scalar curvature of each Σ , the expansion rate θ and the shear σ of the unit vector field normal to Σ . The ‘‘curvature perturbation’’ represents the amplitude of perturbation in the intrinsic curvature of Σ . The remaining two quantities θ and σ are both connected with the behaviour of the vector normal to Σ . Here, we concentrate on the expansion rate θ being given by the divergence of the normal vector (7.20). Then we can identify θ as the extrinsic curvature (4.2):

$$\theta = n^\mu{}_{;\mu} = -K. \quad (7.33)$$

As θ denotes the expansion rate, we can interpret the extrinsic curvature K as the ‘‘local’’ Hubble parameter $H = H(t, \mathbf{x})$:

$$K(t, \mathbf{x}) \equiv -3H(t, \mathbf{x}). \quad (7.34)$$

Then, by neglecting the second-order spatial gradient terms, the *fully non-linear* energy constraints, ADM trace and the scalar field equations of motion become [55]

$$H^2 = \frac{1}{3m_{\text{Pl}}^2} \left(\frac{G_{ab}}{2} \frac{\partial \phi^a}{\mathcal{N} \partial t} \frac{\partial \phi^b}{\mathcal{N} \partial t} + V \right), \quad (7.35)$$

$$\frac{\partial H}{\mathcal{N} \partial t} = - \frac{G_{ab}}{2m_{\text{Pl}}^2} \frac{\partial \phi^a}{\mathcal{N} \partial t} \frac{\partial \phi^b}{\mathcal{N} \partial t}, \quad (7.36)$$

$$\frac{D}{\mathcal{N} \partial t} \left(\frac{\partial \phi^a}{\mathcal{N} \partial t} \right) + 3H \frac{\partial \phi^a}{\mathcal{N} \partial t} + G^{ab} V_b = 0. \quad (7.37)$$

Thus, in terms of the proper time $\mathcal{N}dt$ the above equations are precisely the same as the *background* equations (3.3), (3.4) and (3.5). That is, on sufficiently large scales, the information on the evolution of non-linear perturbations can be described by essentially background equations [56].

Another important point is that from (7.35), (7.35) and (7.37) the local Hubble parameter, which is essentially a geometric quantity, is related to the background-like field evolution. From the non-perturbative form of the spatial metric (5.46), using the identity

$$K = -\frac{d}{Ndt} \log \sqrt{\gamma}, \quad (7.38)$$

from (7.34) we notice

$$H(t, \mathbf{x}) = \frac{d}{Ndt} \log (ae^\varphi). \quad (7.39)$$

Now, for each spatial point \mathbf{x} we can define the integral with respect to the proper time:

$$N \equiv \int HNdt, \quad (7.40)$$

which is, with H being the volume expansion rate of the three-hypersurfaces, the total expansion of the spatial volume, vix. the number of e -folds. From (7.39), given one and another moments t_1 and t_2 , we can trivially obtain

$$N = \int_1^2 HNdt = \log \left[\frac{a(t_2)}{a(t_1)} \right] + \varphi(t_2) - \varphi(t_1) \equiv N_0 + \varphi(t_2) - \varphi(t_1). \quad (7.41)$$

Thus the total number of e -folds is given by the background number of e -folds N_0 given solely by the background scale factor, and the difference between the curvature perturbations evaluated at each moment. That is,

$$\delta N \equiv N - N_0 = \Delta\varphi. \quad (7.42)$$

Notice that we have not specified the gauge conditions on the hypersurfaces at t_1 and t_2 . Thus we are free to choose in such a way that the initial hypersurface at t_1 is flat, on which by definition $\varphi(t_1) = 0$, and the final one at t_2 is comoving, $\varphi(t_2) = \mathcal{R}(t_2)$. Then

$$\delta N(t_1, t_2) = \mathcal{R}(t_2), \quad (7.43)$$

i.e. the perturbation in the number of e -folds is identical to the non-linear final comoving curvature perturbation.

With multi-field inflation in our mind, let us take the initial time t_1 to be some time during inflation when all the modes of our interest are sufficiently outside the horizon so that the large-scale approximation is valid, and the final time t_2 to be some time after inflation when \mathcal{R} has reached its constant value. Then the dependence of δN on t_1 appears through the phase space variable¹⁰ $\phi^a(t_1)$ and $\dot{\phi}^a(t_1)$. But if we further make use of the slow-roll approximation since t_1 is some time during inflation, we can eliminate the dependence on $\dot{\phi}^a(t_1)$ so that δN depends only on $\phi^a(t_1)$. Thus, we can relate the field fluctuations on the initial flat hypersurface $Q^a(t_1)$ and the final comoving curvature perturbation $\mathcal{R}(t_2)$ as

$$\mathcal{R}(t_2) = \delta N = N_a(t_1)Q^a(t_1) + \frac{1}{2}N_{ab}(t_1)Q^a(t_1)Q^b(t_1) + \dots, \quad (7.44)$$

where $N_a \equiv \partial N / \partial \phi^a$ and so on and we have explicitly expanded up to second order. This is the so-called δN formalism [52, 55, 56, 58].

¹⁰As can be read from (5.15) and (5.19), $\dot{\phi}^a$ is to be precise not the conjugate momentum of ϕ^a . See [57] for more careful discussions on the phase space for inflationary dynamics.

7.3.2 Scale dependence in the δN formalism

In the previous section, we have seen that the δN formalism is implemented in the configuration space with two fixed initial and final moments t_1 and t_2 respectively. These moments are common to all modes of observational interest, and hence the momentum dependence which is of crucial observational importance is not explicit. This however becomes manifest by taking into account the moment of horizon crossing for each mode as follows [59]. For a certain mode with k , the horizon crossing happens at $t_0 < t_1$, i.e. $k = (aH)_0$. Then, the number of e -folds elapsed between t_0 and t_1 is obviously k -dependent as

$$\Delta N_k \equiv \log \left(\frac{a_1}{a_0} \right) \approx \log \left[\frac{(aH)_1}{k} \right]. \quad (7.45)$$

Therefore, the field fluctuations at the initial moment $Q^a(t_1)$ in terms of which the δN formalism is written as in (7.44) can be expanded as

$$Q^a(N_1 = N_0 + \Delta N_k) = Q^a(N_0) + \Delta N_k D_N Q^a(N_0) + \frac{1}{2} (\Delta N_k)^2 D_N^2 Q^a(N_0) + \dots, \quad (7.46)$$

where N_0 and N_1 are the numbers of e -folds corresponding to t_0 and t_1 respectively, and D_N is a covariant derivative with respect to N . Thus, we can see that the non-trivial k -dependence is gained between the moment of horizon crossing $t_0(k)$ which is different mode by mode, and the initial moment for the δN formalism t_1 which is common to all modes. In some sense, t_1 is an intermediate reference moment from which the evolution of each mode until t_2 is identical. Figure 7 shows $t_0(k)$, t_1 and t_2 and the evolution of the curvature perturbation.

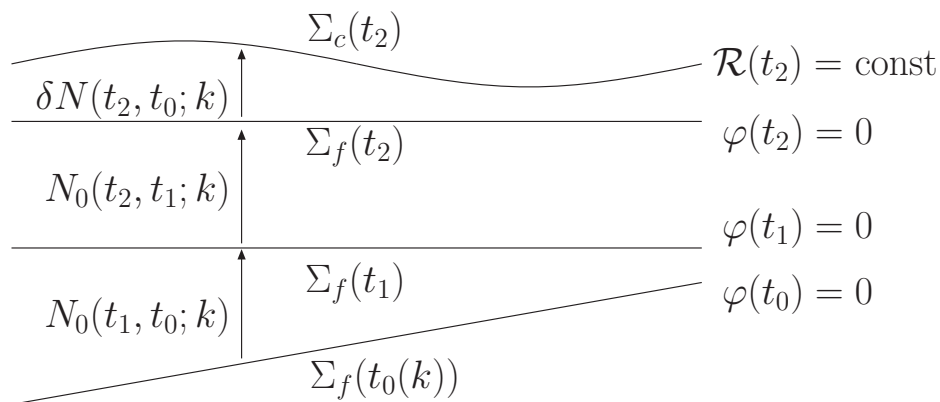


Figure 7: A schematic figure showing different moments described in the main text: $t_0(k)$, t_1 and t_2 . We evaluate the final comoving curvature perturbation $\mathcal{R}(t_2)$ on the hypersurface $\Sigma(t_2)$ with comoving slicing after all isocurvature modes have decayed. We meanwhile choose flat slicing on which the curvature perturbation under this slicing condition, denoted by φ , of course vanishes. Since t_1 and t_2 are common to all modes of observational interest, the evolution of \mathcal{R} is identical between them. But $t_0(k)$ is different for different k -modes so the evolution of \mathcal{R} between t_0 and t_1 is distinctive for each mode.

For practical purpose, we have to find explicitly the derivatives $D_N Q^a$, $D_N^2 Q^a$ and so on. From the equation of motion for Q^a on very large scales with N being the time variable [28],

$$D_N Q^a = w^a{}_b Q^b + \dots, \quad (7.47)$$

$$w_{ab} = u_{(a;b)} + \frac{\mathbb{R}_{c(ab)d} \dot{\phi}_0^c \dot{\phi}_0^d}{3 H H}, \quad (7.48)$$

$$w_{a(bc)} = u_{(a;bc)} + \frac{1}{3} \left[\mathbb{R}_{(a|de|b;c)} \frac{\dot{\phi}_0^d \dot{\phi}_0^e}{H H} - 4 \mathbb{R}_{a(bc)d} \frac{\dot{\phi}_0^d}{H} \right], \quad (7.49)$$

$$u_a = -\frac{V_a}{3H^2}, \quad (7.50)$$

where the indices between vertical bars are excluded from the symmetrization, then (7.46) becomes

$$Q^a(N_1) = Q^a + \Delta N_k \left(w^a{}_b Q^b + \frac{1}{2} w^a{}_{bc} Q^b Q^c + \dots \right) + \frac{1}{2} (\Delta N_k)^2 [(D_N w^a{}_b) Q^b + w^a{}_b w^b{}_c Q^c] + \dots. \quad (7.51)$$

where all terms on the right hand side are evaluated at N_0 , and

$$D_N w^{ab} = w^{ab}{}_{;c} \frac{\dot{\phi}_0^c}{H}. \quad (7.52)$$

Note that ΔN_k spans 5 - 10 for the current observational range of the CMB, but as we can see the coefficients of $(\Delta N_k)^2$ are of second order in slow-roll, so it is safely suppressed compared to ΔN_k terms as long as the expansion (7.46) or (7.51) remains valid.

7.3.3 Correlation functions in the δN formalism

Now using the δN formalism, we compute the correlation functions of the final comoving curvature perturbation $\mathcal{R}(t_2)$. We first move to the Fourier space, where from (7.44) we can write the Fourier component of $\mathcal{R}(t_2)$ as

$$\mathcal{R}_k(t_2) = N_a(t_1) Q_k^a(t_1) + \frac{1}{2} N_{ab}(t_1) [Q^a(t_1) \star Q^b(t_1)]_k + \dots, \quad (7.53)$$

where star denotes a convolution. We first calculate the power spectrum $\mathcal{P}_{\mathcal{R}}$ and its momentum dependence, viz. the spectral index $n_{\mathcal{R}}$. For the power spectrum it is sufficient to work to linear order in Q^a , in which case it is identical to $\delta\phi^a$. The two-point correlation function of $\mathcal{R}(t_2)$ is, from (7.53), related to that of Q^a as

$$\langle \mathcal{R}_k(t_2) \mathcal{R}_q(t_2) \rangle = N_a(t_1) N_b(t_1) \langle Q_k^a(t_1) Q_q^b(t_1) \rangle. \quad (7.54)$$

Using (7.51), the two-point correlation function of Q^a at t_1 can be written in terms of that at the moment of horizon crossing t_0 as

$$\langle Q_k^a(t_1) Q_q^b(t_1) \rangle = \langle Q_k^a(t_0) Q_q^b(t_0) \rangle + 2\Delta N_k w^a{}_c \langle Q_k^b(t_0) Q_q^c(t_0) \rangle, \quad (7.55)$$

where *assuming* that Q^a is light and follows slow-roll dynamics, $\langle Q_k^a Q_q^b \rangle$ is given by (6.46). Thus, from the definition of the power spectrum (6.18), we can read the leading power spectrum of the curvature perturbation in the δN formalism as

$$\mathcal{P}_{\mathcal{R}} = \left(\frac{H}{2\pi} \right)^2 N^a N_a. \quad (7.56)$$

To compute the scale dependence to leading order in slow-roll, there are obvious two leading contributions: the scale dependence of $H^2(t_0)$ which gives rise to -2ϵ , and $w^a{}_c$ which is multiplied by ΔN_k . The other terms, e.g. the derivative of ζ^{ab} , are further slow-roll suppressed so are sub-leading. Hence we can straightforwardly calculate the spectral index as

$$n_{\mathcal{R}} - 1 = \frac{D \log \mathcal{P}_{\mathcal{R}}}{d \log k} = -2\epsilon - 2 \frac{N_a N_b w^{ab}}{N_c N^c}. \quad (7.57)$$

An obvious new contribution is the field space curvature \mathbb{R}_{abcd} contained in w^{ab} . This term must be $\mathcal{O}(\epsilon)$ to be consistent with the observational constraint (7.2), barring an accidental cancellation with other terms.

The bispectrum of \mathcal{R} is defined in a similar manner to the power spectrum by

$$\langle \mathcal{R}_{\mathbf{k}_1}(t_2) \mathcal{R}_{\mathbf{k}_2}(t_2) \mathcal{R}_{\mathbf{k}_3}(t_2) \rangle \equiv (2\pi)^3 \delta^{(3)}(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) B_{\mathcal{R}}(k_1, k_2, k_3). \quad (7.58)$$

As we have seen, the δN formalism concerns the evolution on super-horizon scales thus can describe the non-Gaussianity generated by non-linear evolution on very large scales. It is however blind to what happens on smaller scales, including the moment of horizon crossing. We thus usually assume when we compute the bispectrum using the δN formalism that the intrinsic non-Gaussianity of the fields at the moment of horizon crossing is negligible,

$$\langle Q_{\mathbf{k}_1}^a(t_0) Q_{\mathbf{k}_2}^b(t_0) Q_{\mathbf{k}_3}^c(t_0) \rangle = 0. \quad (7.59)$$

This is however not necessarily the case always, since there are models of inflation where the intrinsic non-Gaussianity at the horizon crossing is not negligible. Therefore to make a proper account of non-Gaussianity using the δN formalism, we should make use of the form that gives the intrinsic non-Gaussianity as small as possible, by e.g. appropriate field redefinition [60].

From (7.53), we can see that the three-point function consists of two terms,

$$\langle \mathcal{R}_{\mathbf{k}_1}(t_2) \mathcal{R}_{\mathbf{k}_2}(t_2) \mathcal{R}_{\mathbf{k}_3}(t_2) \rangle = N_a N_b N_c \langle Q_{\mathbf{k}_1}^a Q_{\mathbf{k}_2}^b Q_{\mathbf{k}_3}^c \rangle + \frac{1}{2} \left\{ N_{ab} N_c N_d \langle [Q^a \star Q^b]_{\mathbf{k}_1} Q_{\mathbf{k}_2}^c Q_{\mathbf{k}_3}^d \rangle + 2 \text{ perm} \right\}. \quad (7.60)$$

We can compute these terms very easily: for the first term, using (7.51) and (6.46), we can easily find

$$\begin{aligned} & N_a(t_1) N_b(t_1) N_c(t_1) \langle Q_{\mathbf{k}_1}^a(t_1) Q_{\mathbf{k}_2}^b(t_1) Q_{\mathbf{k}_3}^c(t_1) \rangle \\ &= N_a(t_1) N_b(t_1) N_c(t_1) \left\{ \frac{1}{2} \Delta N_{k_1} w^a{}_{de} \langle [Q^d(t_0) \star Q^e(t_0)]_{\mathbf{k}_1} Q_{\mathbf{k}_2}^b(t_0) Q_{\mathbf{k}_3}^c(t_0) \rangle + 2 \text{ perm} \right\} \\ &= (2\pi)^3 \delta^{(3)}(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) N_a(t_1) N_b(t_1) N_c(t_1) \frac{H^4(t_0)}{4k_1^3 k_2^3 k_3^3} w^{abc} (k_1^3 \Delta N_{k_1} + 2 \text{ perm}). \end{aligned} \quad (7.61)$$

Likewise, for the second term using (7.55) trivially gives

$$\begin{aligned} & \frac{1}{2} N_{ab}(t_1) N_c(t_1) N_d(t_1) \left\langle [Q^a(t_1) \star Q^b(t_1)]_{\mathbf{k}_1} Q_{\mathbf{k}_2}^c(t_1) Q_{\mathbf{k}_3}^d(t_1) \right\rangle \\ &= (2\pi)^3 \delta^{(3)}(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) N_{ab}(t_1) N_c(t_1) N_d(t_1) \frac{H^4(t_0)}{4k_1^3 k_2^3} (\gamma^{ac} \gamma^{bd} + 2\Delta N_{k_1} w^{ac} \gamma^{bd} + 2\Delta N_{k_2} \gamma^{ac} w^{bd}) . \end{aligned} \quad (7.62)$$

Thus putting both terms together, the leading bispectrum reads

$$B_{\mathcal{R}}(k_1, k_2, k_3) = \frac{N^{ab} N_a N_b}{(N^c N_c)^2} [P_{\mathcal{R}}(k_1) P_{\mathcal{R}}(k_2) + 2 \text{ perm}] , \quad (7.63)$$

where we have defined the dimensionfull power spectrum

$$P_{\mathcal{R}} \equiv \frac{2\pi^2}{k^3} \mathcal{P}_{\mathcal{R}} , \quad (7.64)$$

so that the definition of the power spectrum (6.18) reads similar to (7.58):

$$\langle \mathcal{R}(\mathbf{k}) \mathcal{R}(\mathbf{q}) \rangle \equiv (2\pi)^3 \delta^{(3)}(\mathbf{k} + \mathbf{q}) P_{\mathcal{R}}(k) . \quad (7.65)$$

A convenient way of parametrizing non-Gaussianity is to introduce a set of so-called non-linear parameters. It is defined by the local expansion of the curvature perturbation around its linear, Gaussian component [53]:

$$\mathcal{R} = \mathcal{R}_g + \frac{3}{5} f_{\text{NL}} \mathcal{R}_g^2 + \dots . \quad (7.66)$$

Using the δN formalism, we are interested in the form of the non-linearity generated during the super-horizon evolution of the curvature perturbation. On super-horizon scales, perturbations at different locations cannot communicate with each other, and in this sense the non-linearity produced during this stage is local. Conversely, in the Fourier space, the momenta that constitute the momentum triangle of the Fourier transformation of the three-point correlation function, i.e. the bispectrum, are very different. Typically, one of three momenta is very small in magnitude, and the remaining two are pointing almost the opposite directions with nearly equal magnitude. Because of the form of the triangle these momenta constitute, it is frequently referred to as the ‘‘squeezed limit’’. From the expansion (7.66), we can easily compute the bispectrum as

$$B_{\mathcal{R}}(k_1, k_2, k_3) = \frac{6}{5} f_{\text{NL}} [P_{\mathcal{R}}(k_1) P_{\mathcal{R}}(k_2) + 2 \text{ perm}] , \quad (7.67)$$

where we have assumed for simplicity that f_{NL} is a constant. Comparing (7.63) and (7.67), we can read the leading, scale-independent f_{NL} as [61]

$$\frac{6}{5} f_{\text{NL}} = \frac{N^{ab} N_a N_b}{(N^c N_c)^2} . \quad (7.68)$$

Furthermore, as we did for the power spectrum, we can also compute the scale dependence of f_{NL} by incorporating ΔN_k term explicitly. In the equilateral configuration in which we are sensibly calculate $n_{f_{\text{NL}}}$, the running of f_{NL} [62], we can trivially find

$$n_{f_{\text{NL}}} \equiv \frac{D \log f_{\text{NL}}}{d \log k} = -\frac{N_a N_b N_c w^{abc}}{N^{de} N_d N_e} + 4w^{ab} \left(\frac{N_a N_b}{N^d N_d} - \frac{N_{ac} N_b N^c}{N^{de} N_d N_e} \right) . \quad (7.69)$$

The first term in (7.69) comes from (7.61), i.e. the non-linearity generated during the evolution of the field fluctuations between t_0 and t_1 , while the other terms in (7.69) arises from the non-linear couplings.

8 Conclusions

Inflation offers a simple and consistent framework in which both the homogeneity and isotropy of the current observable universe as well as the tiny initial perturbations as can be observed from the temperature anisotropies in the CMB. Indeed, with the recent observations on the CMB we have entered the era of data-driven cosmology and the concordance cosmological model, Λ CDM, provide a very good fit to the power spectrum of the CMB anisotropies. The constraints on the properties of the primordial perturbations, (7.1) - (7.4), tell us that the primordial adiabatic perturbation has nearly scale-invariant power spectrum and follows almost perfect Gaussian statistics, with the contribution of the tensor perturbation, i.e. the primordial gravitational waves, occupying less than a few percent of power on large scales relevant for the CMB observations. These are mostly consistent with the predictions of simple single field inflation models.

While single field inflation provides the best fit to the data without free parameters, then what should be the merit to consider multi-field inflation? As mentioned at the beginning of this article, multi-field inflation can open a rich possibilities beyond the predictions in single field models that could be constrained and even detected by observations. Some of them include the isocurvature perturbations, correlated or anti-correlated with the curvature perturbation, detectable level of non-Gaussianity that may have non-trivial scale dependence, and possibly residual signatures that survive elusive post-inflationary dynamics such as reheating. Beyond these exciting observational possibilities there are a number of theoretical motivations to study multi-field inflation. The early universe is a huge particle accelerator, and the trail of yet unknown parent physics, whose low-energy effective theory is the standard model of particle physics, should be left on cosmic scales with a multiple number of inflaton fields. In this article we have considered a few basic bricks to understand the dynamics of multi-field inflation.

With advanced observation programs in operation and in plan on the CMB as well as large-scale structure, we are entering the precision era for cosmological observations with unprecedented quality and quantity of data in the coming decade. By that time we should a good understanding about the inflationary dynamics with strong observational supports and/or constraints, along with further lessons about the nature of the physics underlying inflation. At the same time we may well witness the interesting possibilities of multi-field inflation, with further observational and theoretical opportunities widely open.

Acknowledgements

I thank Ana Achúcarro, Christian Byrnes, Ki-Young Choi, Xian Gao, Jai-chan Hwang, Donghui Jeong, Gonzalo Palma, Subodh Patil, Misao Sasaki, David Seery, Min-Seok Seo, Gary Shiu, Ewan Stewart, Spyros Sypsas and Takahiro Tanaka for sharing their insights with me on the subjects discussed in this article. I am also grateful to Misao Sasaki and Takahiro Terada for helpful comments on earlier drafts. I acknowledge the support from the Korea Ministry

of Education, Science and Technology, Gyeongsangbuk-Do and Pohang City for Independent Junior Research Groups at the Asia Pacific Center for Theoretical Physics. I am also supported in part by a Starting Grant through the Basic Science Research Program of the National Research Foundation of Korea (2013R1A1A1006701) and by a TJ Park Science Fellowship of POSCO TJ Park Foundation.

A Gauge transformations

In this appendix section, we address the issue of gauge transformation which was not discussed in the main text. For that, we first consider the issue of “background” and “perturbation”. In the background universe U , there is no ambiguity in choosing the time coordinate on the homogeneous and isotropic spatial hypersurfaces in such a way that time is constant: $t = t_1$ corresponds to the moment when the homogeneous scalar field has a specific value of $\phi(t = t_1)$, and so on. However, in a perturbed universe \hat{U} , our choice of time is arbitrary in the sense that we can choose arbitrary coordinate system where the deviation from homogeneity and isotropy is small. In different coordinate systems, the notion of perturbations is different too. For example, we can choose spatial hypersurfaces on which the density perturbation vanishes. Thus, just saying that the density perturbation is such and such is not enough. We have to also specify the coordinate system in describing the density perturbation.

Let us consider in a more detail. How can we define the perturbation in a scalar quantity $\hat{\phi}$ at a point p in the perturbed universe \hat{U} ? To define the perturbation, we need to specify the corresponding background value ϕ_0 : the difference between $\hat{\phi}$ and ϕ_0 is the perturbation $\delta\phi(p)$. But what is the corresponding background ϕ_0 ? For this, we have to specify a coordinate system, or *mapping* in such a way that each point in the perturbed universe \hat{U} is associated with the corresponding point x^μ in the background universe U . Once this mapping is specified, the perturbation

$$\delta\phi(p) = \hat{\phi}(p) - \phi_0(x^\mu) \tag{A.1}$$

is meaningful.

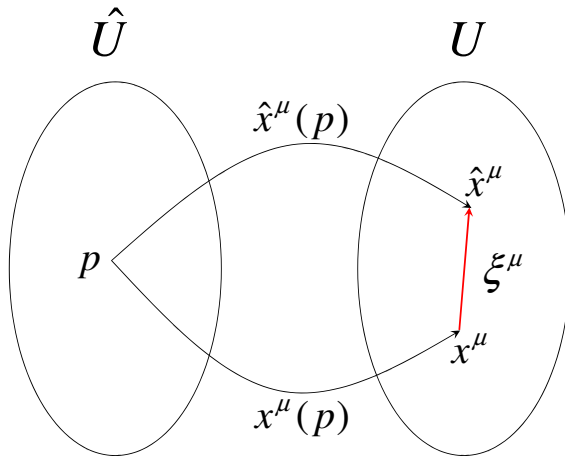


Figure 8: A schematic image of gauge transformation.

So to specify perturbations we only need to specify the coordinate system, or the mapping between \hat{U} and U . The problem is, as stated before, there is no natural choice of this mapping and one is as good (or bad) as the others. Thus we need to know how one mapping is related to another. It is very important to note that any change induced by a change in the mapping is *not* physical: it is simply a transformation because we have changed the coordinate, or “gauge”, to

describe the same thing. In this sense, this non-physical change is called *gauge transformation*. Suppose two coordinate systems, x^μ and \hat{x}^μ , which map p in \widehat{U} to the corresponding *different* points in U , are related by

$$\hat{x}^\mu(p) = x^\mu(p) + \xi^\mu(x^\nu(p)). \quad (\text{A.2})$$

If this transformation is infinitesimal, we have

$$\begin{aligned} \widehat{\delta\phi}(p) &= \widehat{\phi}(p) - \phi_0(\hat{x}^\mu(p)) \\ &= \delta\phi(p) - [\phi_0(\hat{x}^\mu(p)) - \phi_0(x^\mu(p))] \\ &= \delta\phi(p) - \xi^\nu \frac{\partial \phi_0}{\partial x^\nu}(x^\mu(p)). \end{aligned} \quad (\text{A.3})$$

Since the background universe U is spatially homogeneous and isotropic, $\phi_0 = \phi_0(x^0)$ so we simply have

$$\widehat{\delta\phi}(p) = \delta\phi(p) - \dot{\phi}_0(t(p))\xi^0(x^\mu(p)), \quad (\text{A.4})$$

where we have taken $x^0 = t$. It is very important to note that we are comparing two different mappings, $x^\mu(p)$ and $\hat{x}^\mu(p)$, from the *same* point p in \widehat{U} . It is schematically shown in Figure 8.

Note that we can extract the gauge transformations of the metric perturbations by requiring that ds^2 be invariant under the gauge transformation¹¹: with the coordinate transformations

$$t \rightarrow \widehat{t} = t + \xi^0(t, \mathbf{x}), \quad (\text{A.5})$$

$$x^i \rightarrow \widehat{x}^i = x^i + \xi^i(t, \mathbf{x}), \quad (\text{A.6})$$

we can easily see that in the linear order

$$a(\widehat{t}) \equiv \widehat{a} = (1 + H\xi^0) a(t), \quad (\text{A.7})$$

$$\widehat{dt} = \left(1 + \dot{\xi}^0\right) dt + \xi_{,i}^0 dx^i, \quad (\text{A.8})$$

$$\widehat{dx}^i = dx^i + \dot{\xi}^i dt + \xi^i_{,j} dx^j. \quad (\text{A.9})$$

From the fact that the line element in space-time is the same irrespective of the coordinate

¹¹Generic gauge transformation law for an arbitrary tensor quantity Q under the coordinate transformation $x^\mu \rightarrow \hat{x}^\mu = x^\mu + \xi^\mu$ can be written in terms of the Lie derivatives as

$$\widehat{\delta Q} - \delta Q = -\mathcal{L}_\xi Q,$$

where \mathcal{L}_ξ is the Lie derivative in the direction of the vector ξ . By component, a Lie derivative on a tensor $T^{\alpha_1 \dots}_{\beta_1 \dots}$ is given by

$$\mathcal{L}_\xi (T^{\alpha_1 \dots}_{\beta_1 \dots}) = \xi^\gamma \partial_\gamma T^{\alpha_1 \dots}_{\beta_1 \dots} - (\partial_\gamma \xi^{\alpha_1}) T^{\gamma \alpha_2 \dots}_{\beta_1 \dots} + (\partial_{\beta_1} \xi^\gamma) T^{\alpha_1 \dots}_{\gamma \beta_2 \dots} + \dots.$$

For example, the perturbations in the metric tensor transform as

$$\widehat{\delta g}_\mu - \delta g_{\mu\nu} = -g_{\mu\nu,\rho} \xi^\rho - g_{\rho\nu} \xi^{\rho}_{,\mu} - g_{\mu\rho} \xi^{\rho}_{,\mu}.$$

transformation, we can write using the perturbed metric (4.13)

$$\begin{aligned}
\widehat{ds}^2 &= - \left(1 + 2\widehat{A}\right) \widehat{dt}^2 + 2\widehat{a}\widehat{\mathcal{B}}_i \widehat{dt} \widehat{dx}^i + \widehat{a}^2 \left[(1 + 2\widehat{\varphi}) \delta_{ij} + 2\widehat{\mathcal{E}}_{ij} \right] \widehat{dx}^i \widehat{dx}^j \\
&= - \left[1 + 2 \left(\widehat{A} + \dot{\xi}^0 \right) \right] dt^2 + 2a \left(\widehat{\mathcal{B}}_i - \frac{\xi_{,i}^0}{a} + a\dot{\xi}_i \right) dt dx^i \\
&\quad + a^2 \left\{ \left[1 + 2 \left(\widehat{\varphi} + H\xi^0 \right) \right] \delta_{ij} + 2 \left(\widehat{\mathcal{E}}_{ij} + \frac{\xi_{i,j} + \xi_{j,i}}{2} \right) \right\} dx^i dx^j, \tag{A.10}
\end{aligned}$$

where $\xi_i = \delta_{ij}\xi^j$. Equating this expression with (4.13), we can find that under the coordinate transformation given by (A.5) and (A.6), the new metric perturbations are given by

$$\widehat{A} = A - \dot{\xi}^0, \tag{A.11}$$

$$\widehat{\mathcal{B}}_i = \mathcal{B}_i + \frac{\xi_{,i}^0}{a} - a\dot{\xi}_i, \tag{A.12}$$

$$\widehat{\varphi} = \varphi - H\xi^0, \tag{A.13}$$

$$\widehat{\mathcal{E}}_{ij} = \mathcal{E}_{ij} - \frac{\xi_{i,j} + \xi_{j,i}}{2}. \tag{A.14}$$

Further, we can decompose the spatial gauge transformation vector ξ^i into the scalar and transverse vector components as we did for the metric perturbation,

$$\xi^i = \delta^{ij}\xi_{,j} + \xi^{(\text{tr})i}, \tag{A.15}$$

where $\xi^{(\text{tr})i}_{,i} = 0$. Then, we can find trivially that the scalar, vector and tensor components of \mathcal{B}_i and \mathcal{E}_{ij} given by (4.14) and (4.15) transform as

$$\widehat{B} = B + \frac{\xi^0}{a} - a\dot{\xi}, \tag{A.16}$$

$$\widehat{S}_i = S_i - a\dot{\xi}_i^{(\text{tr})}, \tag{A.17}$$

$$\widehat{H}_T = H_T - \xi, \tag{A.18}$$

$$\widehat{F}_i = F_i - \xi_i^{(\text{tr})}, \tag{A.19}$$

$$\widehat{h}_{ij}^{TT} = h_{ij}^{TT}. \tag{A.20}$$

Notice that the tensor perturbation h_{ij}^{TT} as well as the combination $S_i - a\dot{F}_i$ remain the same under the gauge transformation, i.e. it is *gauge-invariant*. Thus, when we consider the vector and tensor perturbations, we need not worry about the gauge ambiguity because the variables we are dealing with are from the beginning gauge invariant. Gauge ambiguity only matters for scalar perturbations, and we will explicitly discuss this issue in the following section. Also, it is fruitful to notice that for the scalar components of the Einstein equation, B and H_T only appear in the specific combination $aB - a^2\dot{H}_T$ [44]. From (A.16) and (A.18), we can see that

$$\widehat{a}\widehat{B} - \widehat{a}^2\widehat{\dot{H}}_T = a \left(B + \frac{\xi^0}{a} - a\dot{\xi} \right) - a^2 \frac{d}{dt} (H_T - \xi) = aB - a^2\dot{H}_T + \xi^0, \tag{A.21}$$

so that although the transformations of B and H_T include the spatial component of the gauge transformation ξ , in practice only ξ^0 , the time translation matters.

References

- [1] A. Einstein, *Annalen Phys.* **49**, 769 (1916) [*Annalen Phys.* **14**, 517 (2005)].
- [2] G. Lemaitre, *Annales Soc. Sci. Brux. Ser. I Sci. Math. Astron. Phys. A* **47**, 49 (1927) ; *Gen. Rel. Grav.* **29**, 641 (1997) [*Annales Soc. Sci. Brux. Ser. I Sci. Math. Astron. Phys. A* **53**, 51 (1933)].
- [3] E. Hubble, *Proc. Nat. Acad. Sci.* **15**, 168 (1929).
- [4] A. A. Penzias and R. W. Wilson, *Astrophys. J.* **142**, 419 (1965).
- [5] R. H. Dicke, P. J. E. Peebles, P. G. Roll and D. T. Wilkinson, *Astrophys. J.* **142**, 414 (1965).
- [6] C. W. Misner, *Astrophys. J.* **151**, 431 (1968).
- [7] A. H. Guth, *Phys. Rev. D* **23**, 347 (1981) ; A. D. Linde, *Phys. Lett. B* **108**, 389 (1982) ; A. Albrecht and P. J. Steinhardt, *Phys. Rev. Lett.* **48**, 1220 (1982).
- [8] E. B. Gliner, *Sov. Phys. JETP* **22**, 378 (1966) [*Zh. Eksp. Teor. Fiz.* **49**, 542 (1966)].
- [9] K. Sato, *Mon. Not. Roy. Astron. Soc.* **195**, 467 (1981) ;
- [10] A. A. Starobinsky, *Phys. Lett. B* **91**, 99 (1980).
- [11] V. F. Mukhanov and G. V. Chibisov, *JETP Lett.* **33**, 532 (1981) [*Pisma Zh. Eksp. Teor. Fiz.* **33**, 549 (1981)].
- [12] P. A. R. Ade *et al.* [Planck Collaboration], arXiv:1502.02114 [astro-ph.CO].
- [13] See e.g. D. H. Lyth and A. Riotto, *Phys. Rept.* **314**, 1 (1999) [hep-ph/9807278] ; D. Baumann and L. McAllister, arXiv:1404.2601 [hep-th].
- [14] H. P. Nilles, *Phys. Rept.* **110**, 1 (1984).
- [15] For an incomplete collection of articles, see e.g. H. Kodama and M. Sasaki, *Prog. Theor. Phys. Suppl.* **78**, 1 (1984) ; R. H. Brandenberger, *Rev. Mod. Phys.* **57**, 1 (1985) ; V. F. Mukhanov, H. A. Feldman and R. H. Brandenberger, *Phys. Rept.* **215**, 203 (1992) ; J. E. Lidsey, A. R. Liddle, E. W. Kolb, E. J. Copeland, T. Barreiro and M. Abney, *Rev. Mod. Phys.* **69**, 373 (1997) [astro-ph/9508078] ; A. Riotto, hep-ph/0210162 ; B. A. Bassett, S. Tsujikawa and D. Wands, *Rev. Mod. Phys.* **78**, 537 (2006) [astro-ph/0507632] ; D. Wands, *Lect. Notes Phys.* **738**, 275 (2008) [astro-ph/0702187 [ASTRO-PH]] ; K. A. Malik and D. Wands, *Phys. Rept.* **475**, 1 (2009) [arXiv:0809.4944 [astro-ph]] ; D. Langlois, *Lect. Notes Phys.* **800**, 1 (2010) [arXiv:1001.5259 [astro-ph.CO]] ; Y. Wang, *Commun. Theor. Phys.* **62**, 109 (2014) [arXiv:1303.1523 [hep-th]] ; J. Martin, C. Ringeval and V. Venin, *Phys. Dark Univ.* **5-6**, 75 (2014) [arXiv:1303.3787 [astro-ph.CO]].

- [16] E. W. Kolb and M. S. Turner, “The Early Universe,” *Front. Phys.* **69**, 1 (1990) ; A. D. Linde, “Particle physics and inflationary cosmology,” *Contemp. Concepts Phys.* **5**, 1 (1990) [hep-th/0503203] ; A. R. Liddle and D. H. Lyth, “Cosmological inflation and large scale structure,” Cambridge, UK: Univ. Pr. (2000) 400 p ; V. Mukhanov, “Physical Foundations of Cosmology,” Cambridge, UK: Univ. Pr. (2005) 421 p ; S. Weinberg, “Cosmology,” Oxford, UK: Oxford Univ. Pr. (2008) 593 p ; D. H. Lyth and A. R. Liddle, “The primordial density perturbation: Cosmology, inflation and the origin of structure,” Cambridge, UK: Cambridge Univ. Pr. (2009) 497 p.
- [17] R. Adam *et al.* [Planck Collaboration], arXiv:1502.01582 [astro-ph.CO].
- [18] J. c. Hwang and H. Noh, *Phys. Lett. B* **495**, 277 (2000) [astro-ph/0009268] ; J. c. Hwang and H. Noh, *Class. Quant. Grav.* **19**, 527 (2002) [astro-ph/0103244] ; J. c. Hwang and H. Noh, *Phys. Rev. D* **76**, 103527 (2007) [arXiv:0704.1927 [astro-ph]] ; J. c. Hwang, H. Noh and C. G. Park, arXiv:1511.01360 [gr-qc].
- [19] See e.g. T. P. Sotiriou and V. Faraoni, *Rev. Mod. Phys.* **82**, 451 (2010) [arXiv:0805.1726 [gr-qc]] ; A. De Felice and S. Tsujikawa, *Living Rev. Rel.* **13**, 3 (2010) [arXiv:1002.4928 [gr-qc]] ; S. Nojiri and S. D. Odintsov, *Phys. Rept.* **505**, 59 (2011) [arXiv:1011.0544 [gr-qc]].
- [20] C. Armendariz-Picon, T. Damour and V. F. Mukhanov, *Phys. Lett. B* **458**, 209 (1999) [hep-th/9904075] ; J. Garriga and V. F. Mukhanov, *Phys. Lett. B* **458**, 219 (1999) [hep-th/9904176].
- [21] S. Groot Nibbelink and B. J. W. van Tent, hep-ph/0011325 ; S. Groot Nibbelink and B. J. W. van Tent, *Class. Quant. Grav.* **19**, 613 (2002) [hep-ph/0107272] ; C. M. Peterson and M. Tegmark, *Phys. Rev. D* **83**, 023522 (2011) [arXiv:1005.4056 [astro-ph.CO]] ; C. M. Peterson and M. Tegmark, *Phys. Rev. D* **87**, no. 10, 103507 (2013) [arXiv:1111.0927 [astro-ph.CO]].
- [22] C. Gordon, D. Wands, B. A. Bassett and R. Maartens, *Phys. Rev. D* **63**, 023506 (2001) [astro-ph/0009131] ; G. I. Rigopoulos, E. P. S. Shellard and B. J. W. van Tent, *Phys. Rev. D* **73**, 083521 (2006) [astro-ph/0504508] ; D. Langlois and F. Vernizzi, *Class. Quant. Grav.* **27**, 124007 (2010) [arXiv:1003.3270 [astro-ph.CO]] ; X. Gao, *JCAP* **1310**, 039 (2013) [arXiv:1307.2564 [hep-th]].
- [23] A. Achúcarro, J. O. Gong, S. Hardeman, G. A. Palma and S. P. Patil, *JCAP* **1101**, 030 (2011) [arXiv:1010.3693 [hep-ph]].
- [24] R. L. Arnowitt, S. Deser and C. W. Misner, *Gen. Rel. Grav.* **40**, 1997 (2008) [gr-qc/0405109].
- [25] D. Seery and J. E. Lidsey, *JCAP* **0509**, 011 (2005) [astro-ph/0506056].
- [26] D. Langlois and S. Renaux-Petel, *JCAP* **0804**, 017 (2008) [arXiv:0801.1085 [hep-th]].
- [27] J. O. Gong and T. Tanaka, *JCAP* **1103**, 015 (2011) Erratum: [*JCAP* **1202**, E01 (2012)] [arXiv:1101.4809 [astro-ph.CO]].

- [28] J. Elliston, D. Seery and R. Tavakol, JCAP **1211**, 060 (2012) [arXiv:1208.6011 [astro-ph.CO]].
- [29] X. Gao, D. Langlois and S. Mizuno, JCAP **1210**, 040 (2012) [arXiv:1205.5275 [hep-th]] ; X. Gao, D. Langlois and S. Mizuno, JCAP **1310**, 023 (2013) [arXiv:1306.5680 [hep-th]].
- [30] C. P. Burgess, M. W. Horbatsch and S. P. Patil, JHEP **1301**, 133 (2013) [arXiv:1209.5701 [hep-th]].
- [31] X. Gao and J. O. Gong, JHEP **1508**, 115 (2015) [arXiv:1506.08894 [astro-ph.CO]].
- [32] S. Anderegge and V. F. Mukhanov, Phys. Lett. B **331**, 30 (1994) [hep-th/9403091] ; T. Prokopec and G. Rigopoulos, Phys. Rev. D **82**, 023529 (2010) [arXiv:1004.0882 [gr-qc]] ; J. O. Gong, M. S. Seo and G. Shiu, JHEP **1607**, 099 (2016) [arXiv:1603.03689 [hep-th]]. See also J. Garriga, X. Montes, M. Sasaki and T. Tanaka, Nucl. Phys. B **513**, 343 (1998) [astro-ph/9706229].
- [33] See e.g. M. Henneaux and C. Teitelboim, “Quantization of gauge systems,” Princeton, USA: Univ. Pr. (1992) 520 p ; S. Weinberg, “The Quantum theory of fields. Vol. 1: Foundations,” Cambridge, UK: Univ. Pr. (1995), 640 p.
- [34] J. M. Bardeen, Phys. Rev. D **22**, 1882 (1980).
- [35] V. F. Mukhanov, JETP Lett. **41**, 493 (1985) [Pisma Zh. Eksp. Teor. Fiz. **41**, 402 (1985)] ; M. Sasaki, Prog. Theor. Phys. **76**, 1036 (1986).
- [36] A. Achúcarro, J. O. Gong, S. Hardeman, G. A. Palma and S. P. Patil, JHEP **1205**, 066 (2012) [arXiv:1201.6342 [hep-th]].
- [37] C. Cheung, P. Creminelli, A. L. Fitzpatrick, J. Kaplan and L. Senatore, JHEP **0803**, 014 (2008) [arXiv:0709.0293 [hep-th]].
- [38] See e.g. D. I. Kaiser and E. I. Sfakianakis, Phys. Rev. Lett. **112**, no. 1, 011302 (2014) [arXiv:1304.0363 [astro-ph.CO]] ; K. Schutz, E. I. Sfakianakis and D. I. Kaiser, Phys. Rev. D **89**, no. 6, 064044 (2014) [arXiv:1310.8285 [astro-ph.CO]].
- [39] T. Chiba and M. Yamaguchi, JCAP **0810**, 021 (2008) [arXiv:0807.4965 [astro-ph]] ; J. O. Gong, J. c. Hwang, W. I. Park, M. Sasaki and Y. S. Song, JCAP **1109**, 023 (2011) [arXiv:1107.1840 [gr-qc]] ; T. Kubota, N. Misumi, W. Naylor and N. Okuda, JCAP **1202**, 034 (2012) [arXiv:1112.5233 [gr-qc]].
- [40] D. S. Salopek, J. R. Bond and J. M. Bardeen, Phys. Rev. D **40**, 1753 (1989) ; J. c. Hwang, Class. Quant. Grav. **14**, 1981 (1997) [gr-qc/9605024].
- [41] J. White, M. Minamitsuji and M. Sasaki, JCAP **1207**, 039 (2012) [arXiv:1205.0656 [astro-ph.CO]] ; J. White, M. Minamitsuji and M. Sasaki, JCAP **1309**, 015 (2013) [arXiv:1306.6186 [astro-ph.CO]].
- [42] D. H. Lyth, K. A. Malik and M. Sasaki, JCAP **0505**, 004 (2005) [astro-ph/0411220].

- [43] J. M. Maldacena, JHEP **0305**, 013 (2003) [astro-ph/0210603].
- [44] H. Noh and J. c. Hwang, Phys. Rev. D **69**, 104011 (2004) [astro-ph/0305123].
- [45] J. Chluba, J. Hamann and S. P. Patil, Int. J. Mod. Phys. D **24**, no. 10, 1530023 (2015) [arXiv:1505.01834 [astro-ph.CO]].
- [46] S. Cespedes, V. Atal and G. A. Palma, JCAP **1205**, 008 (2012) [arXiv:1201.4848 [hep-th]] ; A. Achucarro, V. Atal, S. Cespedes, J. O. Gong, G. A. Palma and S. P. Patil, Phys. Rev. D **86**, 121301 (2012) [arXiv:1205.0710 [hep-th]].
- [47] A. J. Tolley and M. Wyman, Phys. Rev. D **81**, 043502 (2010) [arXiv:0910.1853 [hep-th]] ; A. Achucarro, J. O. Gong, S. Hardeman, G. A. Palma and S. P. Patil, Phys. Rev. D **84**, 043502 (2011) [arXiv:1005.3848 [hep-th]].
- [48] R. Gwyn, G. A. Palma, M. Sakellariadou and S. Sypsas, JCAP **1304**, 004 (2013) [arXiv:1210.3020 [hep-th]] ; J. O. Gong, M. S. Seo and S. Sypsas, JCAP **1503**, no. 03, 009 (2015) [arXiv:1407.8268 [hep-th]].
- [49] J. S. Schwinger, J. Math. Phys. **2**, 407 (1961) ; L. V. Keldysh, Zh. Eksp. Teor. Fiz. **47**, 1515 (1964) [Sov. Phys. JETP **20**, 1018 (1965)] ; for extensive discussions on the application to inflationary cosmology, see e.g. S. Weinberg, Phys. Rev. D **72**, 043514 (2005) [hep-th/0506236].
- [50] X. Chen and Y. Wang, JCAP **1004**, 027 (2010) [arXiv:0911.3380 [hep-th]]. See also X. Chen and Y. Wang, JCAP **1209**, 021 (2012) [arXiv:1205.0160 [hep-th]].
- [51] J. O. Gong and E. D. Stewart, Phys. Lett. B **510**, 1 (2001) [astro-ph/0101225].
- [52] T. T. Nakamura and E. D. Stewart, Phys. Lett. B **381**, 413 (1996) [astro-ph/9604103] ; J. O. Gong and E. D. Stewart, Phys. Lett. B **538**, 213 (2002) [astro-ph/0202098].
- [53] E. Komatsu and D. N. Spergel, Phys. Rev. D **63**, 063002 (2001) [astro-ph/0005036].
- [54] P. A. R. Ade *et al.* [Planck Collaboration], arXiv:1502.01592 [astro-ph.CO].
- [55] D. S. Salopek and J. R. Bond, Phys. Rev. D **42**, 3936 (1990).
- [56] M. Sasaki and T. Tanaka, Prog. Theor. Phys. **99**, 763 (1998) [gr-qc/9801017].
- [57] G. N. Remmen and S. M. Carroll, Phys. Rev. D **88**, 083518 (2013) [arXiv:1309.2611 [gr-qc]].
- [58] A. A. Starobinsky, JETP Lett. **42**, 152 (1985) [Pisma Zh. Eksp. Teor. Fiz. **42**, 124 (1985)] ; M. Sasaki and E. D. Stewart, Prog. Theor. Phys. **95**, 71 (1996) [astro-ph/9507001].
- [59] C. T. Byrnes and J. O. Gong, Phys. Lett. B **718**, 718 (2013) [arXiv:1210.1851 [astro-ph.CO]] ; J. O. Gong, JCAP **1505**, no. 05, 041 (2015) [arXiv:1409.8151 [astro-ph.CO]].
- [60] G. Domenech, J. O. Gong and M. Sasaki, arXiv:1606.03343 [astro-ph.CO].

- [61] D. H. Lyth and Y. Rodriguez, Phys. Rev. Lett. **95**, 121302 (2005) [astro-ph/0504045].
- [62] X. Chen, Phys. Rev. D **72**, 123518 (2005) [astro-ph/0507053] ; C. T. Byrnes, K. Y. Choi and L. M. H. Hall, JCAP **0902**, 017 (2009) [arXiv:0812.0807 [astro-ph]] ; C. T. Byrnes, S. Nurmi, G. Tasinato and D. Wands, JCAP **1002**, 034 (2010) [arXiv:0911.2780 [astro-ph.CO]] ; C. T. Byrnes, M. Gerstenlauer, S. Nurmi, G. Tasinato and D. Wands, JCAP **1010**, 004 (2010) [arXiv:1007.4277 [astro-ph.CO]]. See also Z. Kenton and D. J. Mulryne, JCAP **1510**, no. 10, 018 (2015) [arXiv:1507.08629 [astro-ph.CO]].