Field Theory in Cosmology

Prof. Enrico Pajer[*]

Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Wilberforce Road, Cambridge, CB3 0WA, UK

These lecture notes discuss quantum field theory in curved spacetime and inflation as well as statistical field theory for large scale structures and the cosmic microwave background. They cover the material taught during the course *Field theory in cosmology* for Part III of the Cambridge Mathematical Tripos. The course is aimed at students that have already had a first course in Quantum Field Theory (QFT) and General Relativity (GR). Some knowledge of cosmology is useful, but not necessary, as the relevant material is reviewed when needed.

[*] enrico.pajer@gmail.com

Contents

Introduction

Field theory has proven to be an extremely useful language to describe nature. We are very familiar with the role fields play at the classical level from thinking about the gravitational field in Newton's theory. In Maxwell's description of electromagnetic phenomena, we learn that the electric and magnetic fields are not just a crutch to compute the forces acting on changed particles. The electromagnetic field is an entity on its own: it carries energy and momentum, it can be set up in a variety of initial conditions and it evolves according to precise laws.

At the classical level, fields are often contrasted with particles: fields permeate space while particles are localized; fields are continuous functions while particles are discrete. This distinction between particles and field already becomes blurred when we enter the realm of statistical physics. Even though a given system might be in a specific microstate, where all particles occupy precise positions, if we can only observe macroscopic properties of the system this point of view is useless. We are then forced to ascribe to the system a macroscopic state, i.e. an ensemble of many distinct microscopic configurations. In a macroscopic state, particles don't occupy precise positions anymore. Rather they can be characterized by a probability of being here or there. So, in statistical physics, the concept of a field emerges again. For example, density, temperature or pressure are statistical fields that describe macroscopic properties of a system or properties of an average particle.

In quantum mechanics, all particles have an associated "probability" wave (actually an amplitude, with $|\text{amplitude}|^2 \sim$ probability) and fields emerge again as the right language to describe the dynamics. Indeed, it is not a coincidence that the very idea of quantization emerged form the study of statistical systems, as in Planck's solution of the ultraviolet catastrophe. The particle-wave duality completely shatters the distinction between particles and fields. Finally, when we face the daunting task of marrying quantum mechanics with special relativity and its demands of causality and locality, we are forced once again to resort to fields. *Quantum field theory* (QFT) is the framework within which quantum mechanics can describe interactions that respect the observed locality of natural phenomena. Fields in QFT are not probability waves nor amplitudes. Fields are the tool by which locality and oftentimes Lorentz invariance is imposed onto the Hilbert space of quantum mechanics.

In cosmology, namely the study of the evolution of the universe, we encounter fields playing all the roles discussed above. In trying to describe the first fraction of a second of the big bang, we are forced to use QFT because both relativity and quantum mechanics induce large deviations from a classical behavior. As time goes on and the universe expands, the quantum perturbations generated during the primordial universe become classical in the sense that all practically measurable observables commute with each other to a large degree of accuracy. But our theory is still not deterministic because the initial conditions are only provided now as classical probabilities. We then need *statistical field theory* to describe observations. This is particularly important in the description of the formation of structures in the universe, where our inability to predict initial conditions in a deterministic way changes the dynamics qualitatively. Finally, in the quasi-linear regime that is relevant to describe how light moves in the universe and creates the Cosmic Microwave Background, we revert back to classical, deterministic field theory, with statistical effects playing only a marginal role.

Field theory and cosmology are therefore a match made in heaven. For those interested in cosmology, this course will attempt to provide a solid theoretical foundation and a field theory toolkit to tackle the hardest problems. For those interested in field theory, this course will provide a point of view complementary to that of courses that focus on particle physics (such as QFT and advanced QFT in Part III). In the expanding spacetime provided by cosmology, fields behave differently from we observe in the flat spacetime of particle physics. New phenomena take place such as the ambiguity to define a vacuum state and particle creation; the fundamental observables change too: we set aside scattering amplitudes and focus on cosmological correlators.

This course discusses applications of classical, statistical and quantum field theory to cosmology. The course comprises of three interconnected topics:

- Inflation and primordial quantum perturbations (QFT in curved spacetime)
- The matter and galaxy distribution in the Large Scale Structure of the Universe (statistical field theory)
- The Cosmic Microwave Background (classical and statistical field theory)

The goals of the course are: to discuss open problems in cosmology and describe their intimate relation to fundamental high energy physics; to provide the basic knowledge to understand modern research literature in cosmology; to explore how field theory provides a unifying formalism to describe disparate physical processes from the birth of the Universe to the highly non-linear cosmic web.

If one had to condense the whole course in a single equation it would be this one

$$\langle\prod_a^n Y(\mathbf{x}_a)\rangle = \left[\prod_a^n \int_{\mathbf{k}_a} e^{i\mathbf{k}_a\mathbf{x}_a}\Delta_Y(\mathbf{k}_a)\right]\langle\prod^n \mathcal{R}(\mathbf{k}_a)\rangle + \mathcal{O}(\mathcal{R}^{n+1})\,, \tag{0.1}$$

$$\text{Cosmo observations} \sim \text{QFT in curved spacetime}\,.$$

Let's discuss what all these quantities are: on the left-hand side, $Y(\mathbf{x})$ can be the distribution on cosmological scales of anything you can get a hold of. For example $Y(\mathbf{k})$ could be the density of dark matter, galaxies, photons, or the curvature of spacetime itself. On the right-hand side, $\mathcal{R}$ represents the fluctuations of spacetime induced by quantum mechanical fluctuations during the primordial universe, in the first fraction of a second of the big bang. The transfer functions $\Delta_Y(\mathbf{k})$ tells us how the primordial fluctuations $\mathcal{R}$ are converted into fluctuations of the distribution of staff by the time evolution after the primordial universe, during the hot big bang. Notice that the relation is just linear, up to higher order terms that become very small when the points $\mathbf{x}_a$ are separated by cosmological distances, e.g. $\gg$ Mpc today. This relation tells us that by the spatial correlations of $\langle Y^n\rangle$ measured by cosmological surveys are in fact measurements of what we believe are quantum mechanical correlators $\langle\zeta^n\rangle$ in a curved spacetime. The three parts of this course focus on three parts of the above equation: in the first part we learn how to compute $\langle\mathcal{R}\rangle$ using quantum field in curved spacetime, in the second part we learn how to compute the higher order terms $\mathcal{O}(\mathcal{R}^2)$ from gravitational collapse in the late universe using statistical field theory and finally, in the third part we learn how derive the transfer functions for a given model of the history of the universe.

Part I

Quantum Field Theory and Inflation

Part I of these notes will cover the application of quantum field theory techniques to inflation in the early Universe. We will touch on many of the conceptual and computation novelties when doing QFT on curved spacetime, and organise the narrative around the following central question:

Given a particular action, $S[g_{\mu\nu}, \phi]$, for matter fields ϕ coupled to a spacetime metric $g_{\mu\nu}$, how can we predict the statistical correlations between the curvature perturbations produced during inflation (and imprinted in the CMB)?

Answering this question is crucial if we are to extract information about the early Universe from our cosmological measurements today. We will split the answer to this question into four steps, which are summarised in Figure 1. The idea is to start by treating both matter and metric classically, then treat matter as a quantum field on a classical spacetime background, then treat both matter and metric perturbations quantum mechanically, and finally compare our results with the correlations we have or are trying to observe in the cosmological suveys. So the plan is:

Step 1: Find classical backgrounds (Sec. 1).

Step 2: Quantize matter perturbations (Sec. 2, Sec. 3 and Sec. 4).

Step 3: Quantize metric perturbations (Sec. 5 and 7).

Step 4: Compare to observations (Sec. 8).

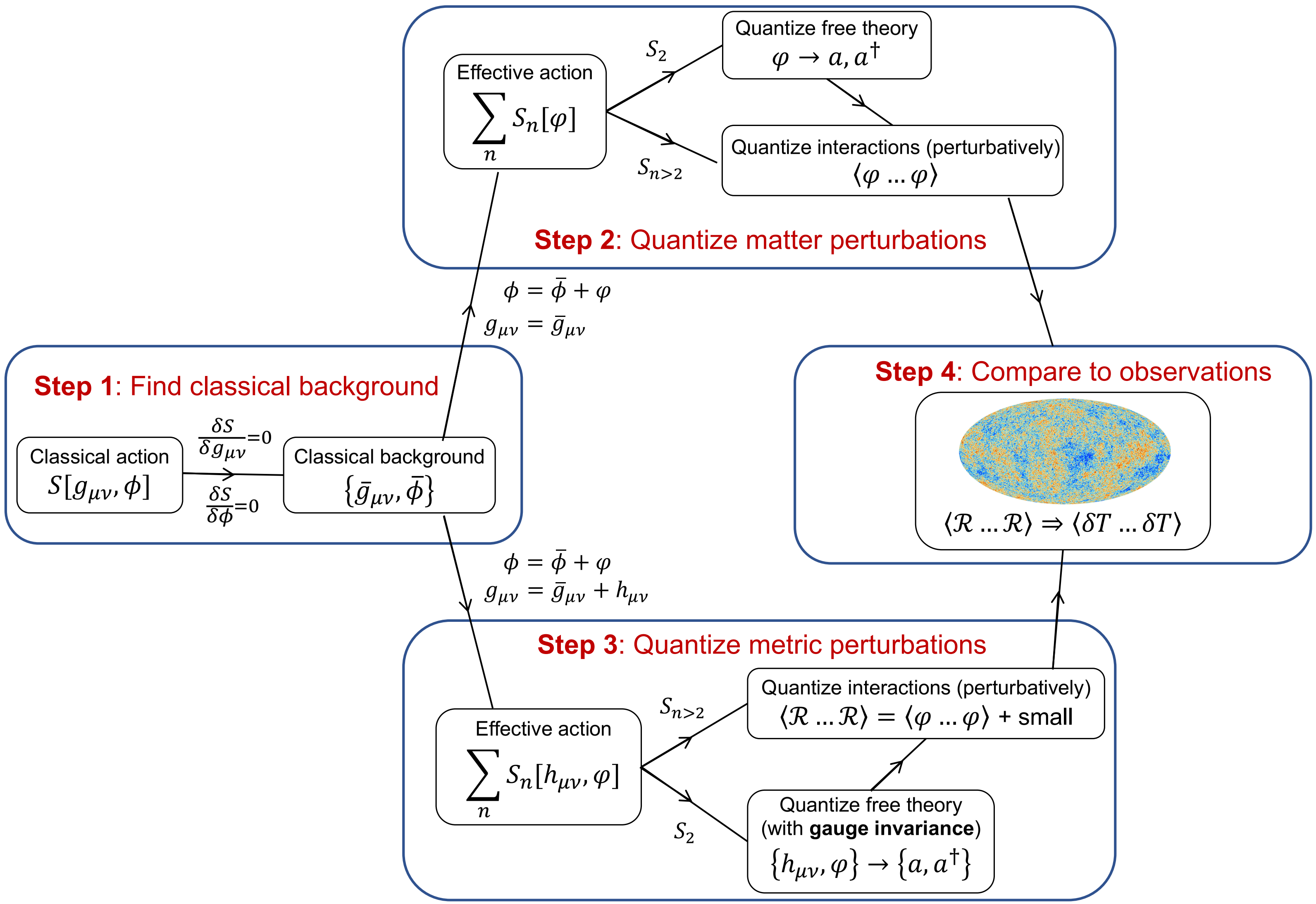


Figure 1: Roadmap for lectures 1–12, showing how each topic is related to one another and to the overall goal of computing, from first principles, the primordial initial condition for the CMB and hot Big Bang cosmology.

1 A brief review of background cosmology ref

In this section, we start by reviewing some relevant facts about cosmological backgrounds. Later on, this will simplify our task of describing quantum fields on these backgrounds and will allow us to make contact with observations. Most of the material in this Sec. is also covered in any introductory cosmology course (e.g. Part III Cosmology).

In general relativity (GR), the spacetime metric $g_{\mu\nu}(\mathbf{x}, t)$ is a dynamical degree of freedom that obeys a set of ten, coupled, second order, non-linear partial differential equations known as the Einstein Equations,

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = \frac{1}{M_{\rm Pl}^2} T_{\mu\nu} \,, \tag{1.1}$$

which follow from varying the action,

$$S = \int d^4x \sqrt{-g} \left[\frac{M_P^2}{2} R + \mathcal{L}_M \right] \,. \tag{1.2}$$

A few definitions are in order. $M_{\rm Pl} \equiv (8\pi G_N)^{-1/2}$ is the reduced Planck mass (in units such that $\hbar = c = k_B = 1$), $R_{\mu\nu}$ is the Ricci tensor given in terms of the Christoffel symbols by[2]

$$R_{\mu\nu} \equiv 2\Gamma^{\rho}_{\mu[\nu,\rho]} + 2\Gamma^{\rho}_{\lambda[\rho}\Gamma^{\Lambda}_{\beta]\alpha} \,, \tag{1.3}$$

$$\Gamma^{\mu}_{\alpha\beta} \equiv \frac{1}{2} g^{\mu\gamma} \left(g_{\alpha\gamma,\beta} + g_{\beta\gamma,\alpha} - g_{\alpha\beta,\gamma} \right) \,. \tag{1.4}$$

Finally, $T_{\mu\nu}$ is the energy-momentum tensor, which can be extracted from a given matter action $S_M = \int d^4x \sqrt{-g}\, \mathcal{L}_M$ by

$$T^{\mu\nu} = \frac{2}{\sqrt{-g}} \frac{\delta S_M}{\delta g_{\mu\nu}} \,, \tag{1.5}$$

Notice that it is symmetric by definition. It is painfully clear from (1.1) that exact solutions of GR are in general hard to come by. But a handful of highly symmetric solutions are relatively easy to find. The simplest possible solution of the Einstein Equations is Minkowski spacetime,

$$ds^2 = -dt^2 + dx^i \delta_{ij} dx^j \,, \tag{1.6}$$

which requires $T_{\mu\nu} = 0$. This solution of GR is closest to our intuition of space and time. It is also *maximally symmetric*, i.e. it possesses the largest amount of symmetry: in $(3+1)$-dimensions this is the ten isometries that combine to form the Poicaré group ($ISO(3,1)$): three rotations and three boosts, forming the Lorentz group ($SO(3,1)$) plus one time and three spatial translations. Minkowski is so symmetric that any point can be related to any other point by a symmetry transformation. There are actually two other spacetimes that have this property, *de Sitter* (dS) and *Anti-de Sitter* (AdS) spacetime, and they will be discussed at the end of this section.

[2] Indices right after a comma indicate partial derivative, as in $g_{\mu\nu,\rho} = \partial_\rho g_{\mu\nu}$, while indices right after a semi-colon indicate covariant derivatives, as in $A_{\mu;\nu} = \nabla_\nu A_\mu$.

1.1 Cosmological spacetimes and cosmic dynamics

In a maximally symmetric universe such as Minkowski, there cannot be any beginning or end of time; there cannot even be a history because every time is equivalent to any other time. While such an eternal universe might be appealing from a philosophical or aesthetical point of view, it is in contradiction with the last century of cosmological observations. In particular, the observation of the expansion of the universe and of the Cosmic Microwave Background radiation show that the universe in the past was much denser and hotter than it is now. In such a universe time translations should not be a symmetry, or more precisely there should be no time-like Killing vectors.
The simplest possibility is the Friedmann-Lemaitre-Robertson-Walken (FLRW) metric

$$ds^2 = -dt^2 + a(t)^2 \frac{dx^i dx^j \delta_{ij}}{(1+K\mathbf{x}^2/4)^2} \tag{1.7}$$

where the metric is written in "quasi-cartesian" coordinates, the parameter K is known as the *spatial curvature* and $a(t)$ is the *scale factor*, whose dynamic will be dictated by the Einstein equations. If we so wish, by rescaling x and a we can always normalize curvature to one of three values: $K = -1$ for an open universe, $K = 0$ for a flat universe and $K = 1$ for a closed universe. The spatial coordinates $\mathbf{x}^i$ are called *comoving coordinates.* Notice that the physical distance between two points at fixed comoving coordinates changes with time because of the scale factor $a(t)$. The time t corresponds to the *proper time* of observers that are at rest with respect to the comoving coordinates. It is often convenient to use a different time coordinate, $ad\tau = dt$, known as *conformal time*, so that

$$ds^2 = a^2(\tau)\left[-d\tau^2 + \frac{dx^i dx^j \delta_{ij}}{(1+K\mathbf{x}^2/4)^2}\right] . \tag{1.8}$$

This FLRW metric has six isometries, which locally can be thought of as three space translations and three rotations. It describes a *homogeneous and isotropic* universe that looks the same at every point in space in every direction. This metric is simple enough that it provides a class of exact, fully non-linear solutions of the Einstein equations. A most remarkable fact that should blow your mind is that this most simple spacetime for $K = 0$ is in fact a very good description of our own universe on distances much larger than average distance between galaxies, about a few Megaparsec[3] (Mpc). There are small deviations from perfect homogeneity in our universe: you, me, the galaxy we live in and the other billion of galaxies out there. We will come to discuss those in the second part of this course.
As compared with Minkowski spacetime, in an FLRW spacetime time translation and the three Lorentz boosts fail to be isometries because of the time dependence of $a(t)$, which is conveniently captured by the so-called *Hubble parameter*,

$$H(t) \equiv \frac{\dot{a}(t)}{a(t)} . \tag{1.9}$$

All of the new and surprising cosmological phenomena that you have not yet encountered in particle physics will have an H appearing somewhere. The absence of time

[3] One Megaparsec, or 10^6 parsec is a useful unit of distance in cosmology. One parsec is the distance at which one astronomical unit (Au), the average distance Earth-Sun, subtends and angle of one arcsecond, which is 1/3600 of a degree. A parcec is about 3.26 lightyears.

translations has profound implications for constructing a QFT. First, energy is not conserved and particles can be created or destroyed as the universe expands. We believe that the structures we observe in the universe were created precisely by this process. Second, unlike in Minkowski, there is in general no unique choice for a vacuum state. It will be only under certain specific conditions and assumptions that we will be able to choose a preferred vacuum. Third, the expansion of space changes the energy of a given state hence mixing different scales, which requires care when discussing Effective Field Theories (EFTs).

Continuity equation. To have a chance at solving the Einstein Equations we need to specify an energy-momentum tensor that has the same symmetries as the spacetime metric. The most generic $T_{\mu\nu}$ that is invariant under rotations and translations is[4]

$$T^{\mu}{}_{\nu} = \mathrm{Diag}\left\{-\rho, p, p, p\right\} , \tag{1.11}$$

where the *energy density* ρ (with units $[\rho] = E/L^3$) and the *pressure* p (with units $[p] = M/(T^2L) = E/L^3$) are only functions of time $\rho = \rho(t)$, $p = p(t)$. We can interpret this as the energy-momentum tensor of a homogeneous perfect fluid in its rest frame

$$T_{\mu\nu} = (\rho + p)\, u_\mu u_\nu + g_{\mu\nu} p \,, \tag{1.12}$$

where u_μ is the normalized fluid velocity ($u_\mu u^\mu = -1$), which in the rest frame would be $u_\mu = \{1, 0, 0, 0\}$. The Einstein Equations imply that the energy-momentum tensor is covariantly conserved, namely

$$T^{\mu\nu}{}_{;\nu} = \partial_\nu T^{\mu\nu} + \Gamma^{\mu}_{\alpha\nu} T^{\alpha\nu} + \Gamma^{\nu}_{\alpha\nu} T^{\mu\alpha} = 0\,. \tag{1.13}$$

Of these four equations, the only non-vanishing one is $\nu = 0$ (a consequence of rotation invariance). Using the FLRW metric to compute the Christoffel symbols and (1.11) one finds the *continuity equation*

$$\boxed{\dot\rho + 3H\,(\rho + p) = 0}\,. \tag{1.14}$$

In words, this tells us that the energy density changes in time only if the universe expands or contracts, $H \neq 0$. The Einstein Equations will not tell us what type of matter permeates the universe. To specify that we need to specify an *equation of state*, namely how the pressure is related to the density and possibly other thermodynamical variables. Most systems of interest in cosmology can be described to good approximation by a simple, linear equation of state

$$p = w\rho\,, \tag{1.15}$$

where the constant w is the "equation-of-state parameter". For these linear equations of state, it is easy to solve the continuity equation for any give expansion history $a(t)$:

$$\dot\rho + 3\frac{\dot a}{a}(1+w)\rho = 0 \quad \Rightarrow \quad \rho(t) = \rho(t_0)\left[\frac{a(t)}{a(t_0)}\right]^{-3(1+w)} . \tag{1.16}$$

For example:

[4]The indices are raised and lowered with the FLRW metric. So for example, for $K = 0$

$$T_{\mu\nu} = \mathrm{Diag}\left\{\rho, a^2p, a^2p, a^2p\right\} . \tag{1.10}$$

- For non-relativistic matter, also known as *dust*, the velocity is much smaller than the speed of light, $v \ll c$ and so the energy $E = m\sqrt{c^2 + v^2} \simeq mc$ is much larger than the pressure $E \sim mc \gg mv$. Then the pressure, which is a measure of the average momentum of particles is negligible compared with the energy density, which is proportional to the mass density, $p \ll \rho$ or $w \ll 1$. In an expanding universe dust dilutes as $\rho \propto a^{-3}$.

- For relativistic matter, also known as *radiation*, momentum and energy are equal and therefore the pressure is similar to the energy density. From statistical mechanics we find out that the precise proportionality constant is[5] $p = \rho/3$ and so $w = 1/3$. In an expanding universe radiation dilutes as $\rho \propto a^{-4}$, which is faster than dust.

- For a *cosmological constant*[6], also known as vacuum energy, $T_{\mu\nu}/M_{\rm Pl}^2 = -\Lambda g_{\mu\nu}$ and therefore $p = -\rho = -M_{\rm Pl}^2 \Lambda$ or $w = -1$. Since now $\rho + p = 0$, from (9.29) we learn that a cosmological constant does not dilute as the universe expands, $\rho \propto a^0 \sim const$. We could have expected this from its name.

A simple interpretation of the above scalings is that of an expanding box of linear size $a(t)$. Non-relativistic matter density dilutes with the volume, i.e. a^{-3}. Relativistic matter, a.k.a. radiation, also dilutes with the volume as a^{-3}, but it has an extra a^{-1} suppression due to the redshift of the momentum of each particle (and the mass is negligible).

Friedmann equation. Let's now solve the Einstein Equations for an FLRW metric. Using the definition of the Ricci tensors in (1.3), and the FLRW metric (1.7), a lengthy but straightforward computation shows

$$R^0{}_0 = 3\frac{\ddot{a}}{a}\,, \qquad R^i{}_j = \delta_{ij}\,\frac{2K + 2\dot{a}^2 + a\ddot{a}}{a^2}\,, \qquad R = 6\left[\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 + \frac{K}{a^2}\right]. \tag{1.17}$$

The 00-component of the Einstein Equation (most conveniently computed with one upper and one lower index) in (1.1) is then easily derived

$$\boxed{3M_{\rm Pl}^2\left(H^2 + \frac{K}{a^2}\right) = \sum_a \rho_a\,,} \tag{1.18}$$

where i runs over all constituents of the universe (for example radiation, Dark Matter, neutrinos and baryons). This is the *Friedmann equation*. Notice that since an FLRW metric has only one free function $a(t)$, we need only one of the ten Einstein Equations. It is sometimes convenient to make the Friedman equation dimensionless by dividing it by the *critical density* (a function of time)

$$\rho_c \equiv 3M_{\rm Pl}^2 H^2\,. \tag{1.19}$$

This leads to,

$$1 + \Omega_K = \sum_a \Omega_a\,, \quad \text{with} \quad \Omega_K \equiv \frac{K}{H^2 a^2}\,, \quad \Omega_a \equiv \frac{\rho_a}{\rho_c}\,. \tag{1.20}$$

[5] A good mnemonic for this is to recall that the theory of electromagnetism does not have any scale and is therefore conformal invariant. This in turns demands that the energy-momentum tensor is traceless, $T^\mu_\mu = 0$, from which $w = 1/3$ follows immediately.

[6] Our conventions for the definition of Λ are as in (1.42).

The Ω's are called *fractional energy densities* and are manifestly dimensionless. From this form of the equation we see that curvature parameter K tells us whether the energy density of the constituents of the universe is smaller or larger than the critical one.

Single-component flat universe. To develop some intuition, let's focus on a simple universe that has only one type of stuff, i.e. with a single component, and zero curvature $K = 0$. Using a to parameterize time we can solve the Friedman equation as follows:

$$H = \frac{\dot{a}}{a} = \sqrt{\frac{\rho}{3M_{\rm Pl}^2}} = H_0 \left(\frac{a_0}{a}\right)^{3(1+w)/2} \Rightarrow a(t) = \left[\frac{3}{2}\left(1+w\right) H_0 t\right]^{\frac{2}{3(1+w)}} , \tag{1.21}$$

where $w \neq 1$ was assumed and I fixed the integration constant requiring that a vanishes at past infinity.
Important solutions for the scale factor are then

- For non-relativistic matter (*dust*), $w \simeq 0$ so $a \propto t^{2/3}$.
- For relativistic matter (*radiation*), $w = 1/3$ so $a \propto t^{1/2}$.
- For a cosmological constant (*vacuum energy*), $w = -1$ this expressions is singular. Solving this particular case separately one finds $a \propto e^{H_0 t}$.

Notice that if a is a monomial in t one finds always $H \propto t^{-1}$, or more precisely

$$H(t) = \frac{2}{3(1+w)} \frac{1}{t} \, . \tag{1.22}$$

This gives the *age of the universe* for a single-component universe

$$t_{age} = \frac{2}{3(1+w)} \frac{1}{H(t_{age})} \, . \tag{1.23}$$

Acceleration Equation. There are two other combinations of the Einstein Equations that come in handy. First, subtracting the 00-component from the trace (i.e. the sum over the ii components), one finds the *acceleration equation*,

$$M_{\rm Pl}^2 \frac{\ddot{a}}{a} = -\frac{1}{6}\left(\rho + 3p\right) \, . \tag{1.24}$$

Second, by taking the time derivative of the Friedmann equation and using the continuity equation to get rid of $\dot{\rho}$, we can find the variation of the Hubble parameter,

$$-\dot{H} M_{\rm Pl}^2 = \frac{1}{2}\left(\rho + p\right) \, . \tag{1.25}$$

Most cosmological "stuff" obeys the so-called Null Energy Condition, namely $\rho + p \geq 0$, and so H decreases during the expansion of the universe.
As an aside: we say that an energy-momentum tensor $T_{\mu\nu}$ satisfies the Null Energy Condition iff for every null vector N^μ (i.e. $N^\mu N_\mu = 0$) one has,

$$T_{\mu\nu} N^\mu N^\nu \geq 0 \quad \text{(NEC)} \, . \tag{1.26}$$

For a perfect fluid we can choose $N^\mu = \{1, -1, 0, 0\}$ and find $\rho + p \geq 0$. Violations of the NEC are often associated with pathologies such as ghosts instabilities (i.e. field with the wrong-sing kinetic term that can be created at will by decreasing the energy of the system) or tachyon instabilities [40]. Yet, more exotic theories with non-standard kinetic terms, such as the ghost condensate, are known to safely violate the NEC, see e.g. [29, 80].

1.2 Motivations for inflation

There are several problems with any cosmological model in which the universe undergoes decelerated expansion[7] all the way until the Big Bang, as is the case for example when radiation or matter dominated. I will refer to this class of models collectively as "Hot Big Bang" model, where "hot" refers to the temperature of radiation. I will first discuss two old "background" problems, namely the horizon and curvature problems, which can be stated already for the unperturbed FLRW universe. These problems were originally formulated in the 80's and have not changed much since. Then I will mention two new "perturbation" problems, namely scale invariance and phase-coherence problems, which have to do with the large amount of new data we have collected in the past 30 years, especially from the Cosmic Microwave Background (CMB). There are many additional details here (beyond what was described in the lectures), and so this section 1.2 is non-examinable.

The curvature problem. The curvature problem is the fact that we do not observe any spatial curvature in our universe, $K \simeq 0$, despite the fact that curvature dilutes more slowly than radiation and matter and so grows with time relatively to them. Let us see this in formulae.
Current observational bounds tell us that the energy density in curvature today, $\Omega_{K,0}$, is very small [1],

$$\Omega_{K,0} = 0.000 \pm 0.005\,. \tag{1.27}$$

On the other hand, as we saw in around (1.7), the most general homogeneous and isotropic space can have spatial curvature, i.e. $K \neq 0$. From Eq. (1.27) we see that Ω_K grows with time in an decelerated ($\ddot{a} < 0$) expanding ($\dot{a} > 0$) universe

$$\dot{\Omega}_K = -\ddot{a}\frac{2K}{\dot{a}^3} \propto -\ddot{a} \propto (\rho + 3p) \propto (1 + 3w)\,, \tag{1.28}$$

where in the second step I used the acceleration equation (1.24) to show that in an expanding universe ($\dot{a} > 0$) the fact that $\rho + 3p > 0$ implies deceleration (this is also known as the Strong Energy Condition (SEC)). Since at early times in ΛCDM the universe is dominated by radiation, $w = 1/3$, we conclude that Ω_K must have been even smaller in the past. In other words, extrapolating closer and closer to the Big Bang singularity at $a \to$ and $\rho \to \infty$, we are forced to assume that the initial curvature was tiny, $\Omega_K(a_i) \to 0$, or equivalently the initial total density of the universe was extremely close to the critical one, $\sum_i \rho_i \to \rho_c$ (defined in 1.19). There are only two logical possibilities:

1. The initial conditions of the universe, as it emerged from some yet unknown non-perturbative theory of quantum gravity[8], were extremely fined tuned close to $\Omega_K =$

[7] In the current standard cosmological model known as ΛCDM, an accelerated expansion is induced by the cosmological constant Λ at late times, $z \simeq 0.5$. At any earlier time the expansion is decelerated and so these problems also affect ΛCDM.

[8] Strictly within GR, K is just a parameter, not a dynamical variable, and so there in no physical perturbation that can make $\Omega_K = 0$ unstable. On the other hand, GR is most likely just a low-energy (subPlanckian) effective description of some UV-complete theory of quantum gravity, and it is at least plausible that $\Omega_K = 0$ might be unstable within that larger, yet unknown theory. Perhaps a more concrete example is bubble nulceation. instanton solutions are known in which a new universe nucleates from a single point. To respect the isometries of the system the new universe must have some negative curvature. It is not known whether bubble nucleation and the ensuing ideas about the multiverse play a role in the history of our own universe, and the discussion among experts continues.

0. In this scenario, the existence of our universe is a very rare fluctuation, since any larger initial value of $\Omega_K(t_i)$ would have grown to dominate the energy density and would have prevented the formation of galaxies and therefore of life as we know it. Not a great option, in the opinion of many.

2. The early expansion history of our universe is modified to stop Ω_K from growing as we move back in time. From (1.28) we see that this requires either $\ddot{a}, \dot{a} < 0$, i.e. an early phase of decelerated contraction, or $\ddot{a}, \dot{a} > 0$, i.e. an early phase of accelerated expansion. Since we know the current universe is expanding (recall Hubble's law), the first of these options requires to *bounce* i.e. to transition from $\dot{a} \propto H < 0$ to $\dot{a} \propto H > 0$. Achieving the bounce in a controlled construction is still an open problem and the many proposed models have a series of pathologies. Therefore we will henceforth assume an early phase of accelerated expansion, a.k.a. *inflation*.

Summarizing, to explain the spatial flatness of the observed universe we postulate the existence of a primordial phase of accelerated expansion, $\ddot{a}, \dot{a} > 0$, called inflation.

Horizon problem. In general, we would like to have

$$\left(\begin{array}{c} \text{Distance between regions} \\ \text{of space that look similar} \end{array} \right) \ll \left(\begin{array}{c} \text{Distance travelled by light} \\ \text{since the beginning of time} \end{array} \right), \tag{1.29}$$

so that we can explain why two regions look similar in a way that is compatible with causality. However, in any Hot Big Bang model this inequality is dramatically violated. More precisely, cosmological observations of far away objects allow us to see regions in the past that are separated by much more that than the particle horizon at the time, which is the furthest a signal can travel. Any mechanism attempting to explain homogeneity across these regions then necessarily violates causality and/or locality, leading to the horizon problem.

To see this quantitatively, recall that the *comoving distance* between two generic times t_1 and t_2 with $a_1 = a(t_1) < a(t_2) = a_2$ is found to be

$$\chi(a_1, a_2) \equiv \int_{a_1}^{a_2} \frac{da}{a^2 H} = \frac{1}{a_1 H_1} \frac{2}{3w+1} \left[\left(\frac{a_2}{a_1} \right)^{(3w+1)/2} - 1 \right], \tag{1.30}$$

where I assumed $w \neq -1/3$. Then, the distance of an object at redshift $1+z = a^{-1}$ from us at $a = a_0 = 1$ is given by

$$\chi(a, 1) \equiv \int_{a}^{a_0} \frac{d\log a}{aH} = \frac{1}{H_0} \frac{2}{3w+1} \left[1 - a^{(3w+1)/2} \right], \tag{1.31}$$

Imagine now to look out in the night sky in opposite directions and detect a pair of antipodal objects, each sending us radiation with the same redshift z. The relative comoving distance $\Delta\chi$ between the objects is just $2\chi(a, 1)$. To simplify the algebra, let us neglect Dark Energy and so $w > -1/3$ and assume $a \ll 1$. Then

$$\Delta\chi(a, 1) \simeq 2 \times \frac{1}{H_0} \frac{2}{3w+1} \simeq \frac{\mathcal{O}(1)}{H_0}, \tag{1.32}$$

Recall that the redshift of these objects is $1 + z = 1/a$, and so we conclude that high redshift objects $z \gg 1$ are at a distance of order the Hubble radius today H_0^{-1}, almost

independently of z. Since this is a comoving distance between objects at fixed comoving position (i.e. far away object are in the Hubble flow), it does not depend on time.
Let us compare now this distance with the *comoving particle horizon* in a Hot Big Bang model, i.e. extrapolating radiation domination all the way to $a_i = 0$. Recall that the comoving particle horizon $x_{\text{p.h.}}$ is the comoving distance traveled by light since the beginning of time τ_i, namely $x_{\text{p.h.}}(a) \equiv \chi(a_i, a)$. Recall also that for $w > -1/3$, or equivalently decelerated expansion $\ddot{a} < 0$ (as it is the case for radiation and dust), one can safely take $a_i \to 0$ and so $x_{\text{p.h.}}(a)$ equals the comoving Hubble radius[9] times an order one number

$$x_{\text{p.h.}}(a) = \frac{1}{aH}\frac{2}{3w+1} \simeq \frac{1}{aH}\mathcal{O}(1) \simeq r_H(a)\mathcal{O}(1) \quad (\text{decelerated})\,. \tag{1.33}$$

Assuming decelerated expansion since the Big Bang, and combining (1.33) with (1.32) one finds

$$\frac{\Delta\chi(a)}{x_{\text{p.h.}}(a)} \simeq 2\,\frac{aH}{a_0 H_0} \simeq 2\left(\frac{1}{a}\right)^{(3w+1)/2} \gg 1 \quad (\text{decelerated})\,. \tag{1.34}$$

This means that, in an ever decelerating universe, by observing far away objects ($1/a = 1 + z \gg 1$) we are actually probing scales much larger than the particle horizon at that time. In practice, one can reach $a = (1+z)^{-1} \sim 0.1$ with quasar and $a \sim z^{-1} \sim 10^{-3}$ with Cosmic Microwave Background (CMB) photons. In both cases, the observed physical properties (e.g. density of quasars, temperature and polarization of the CMB) are the same in average in all directions. These observables said to be "statistically isotropic". We conclude that, in the absence of accelerated expansion in our past, the mechanism responsible for this observed statistical isotropy must violate causality. This is the *particle horizon problem.*
Conversely, for a phase of accelerated expansion, $\ddot{a} > 0$ or $w < -1/3$ (such as during Dark Energy domination or inflation) during a period $a \in \{a_i, a_f\}$, the result is divergent as $a_i \to 0$:

$$x_{\text{p.h.}}(a_f) = \frac{1}{a_f H_f}\frac{2}{|3w+1|}\left[\left(\frac{a_f}{a_i}\right)^{|3w+1|/2} - 1\right] \tag{1.35}$$

$$\simeq \frac{1}{a_f H_f}\frac{2}{|3w+1|}\left(\frac{a_f}{a_i}\right)^{|3w+1|/2} \gg r_H \quad (\text{accelerated})\,. \tag{1.36}$$

In the extreme case $w \simeq -1$ (inflation), H is approximately constant and $x_{\text{p.h.}}$ asymptotes to

$$x_{\text{p.h.}} \to \frac{1}{a_i H_i} \quad (\text{inflation})\,. \tag{1.37}$$

Combining this with (1.32) we see that we can make $\Delta\chi(a)/x_{\text{p.h.}}(a)$ as small as we want by taking a_i sufficiently small. This allows to try and find an explanation for the observed statistical isotropy that respects causality. Yet again, we are lead to postulate a phase of accelerated expansion $\ddot{a}, \dot{a} > 0$ in the early universe.

[9] In the literature, r_H is often referred to as *Hubble "horizon"*. This is a misnomer since neither $(aH)^{-1}$ nor its physical cousin H^{-1} represent a horizon in the usual sense of GR. This nomenclature is widely spread and not harmful as long as one is aware of the subtleties.

Phase coherence problem. Our universe has perturbations on all observed scales. A remarkable fact is that these perturbations are observed to oscillate in exact synchronicity. This occurs even on very large scales. In any Hot Big Bang model, distant regions oscillate with precisely the same phase even thought they lie outside of each other's particle horizon. This is the *phase coherence* of cosmological perturbations. This is a problem because on these super-horizon scales no causal mechanism can be devised to "synchronize" the phases and so their coherence becomes a very unlikely coincidence.
This strongly suggests that there was a primordial phase, before the hot Big Bang, during which perturbations were generated and synchronized. For more detail see e.g. the lecture notes [71]. This is a crucial point. It tells us that the seeds of the distribution of everything we see in the night sky today were actually sown during the first fraction of a second of the Big Bang. Cosmological observations are then a probe of *primordial perturbations* and we can use them to learn about the laws of physics during this primordial time.

Scale invariance problem. The last problem with the Hot Big Bang is the surprising fact that the amplitude of perturbations observed in our universe is approximately the same (within 4%) on all cosmological scales (about 3 orders of magnitudes, between 10^{-4} and 10^{-1}Mpc^{-1}). This remarkable feature of primordial perturbations goes under the name of (approximate) *scale invariance*. The mathematical statement is that for every $\lambda \in \mathbb{R}$ and $n \in \mathbb{N}^+$, a field ϕ obeys scale invariance iff

$$\langle \phi(\mathbf{x}_1)\phi(\mathbf{x}_2)\dots\phi(\mathbf{x}_n)\rangle = \langle \phi(\lambda\mathbf{x}_1)\phi(\lambda\mathbf{x}_2)\dots\phi(\lambda\mathbf{x}_3)\rangle\,, \tag{1.38}$$

where all the fields are evaluated at the same time[10]. Scale invariance is most evident in the large scales ($l \lesssim 50$) of the CMB temperature angular power spectrum, shown in Figure 2, and a detailed analysis shows that the initial conditions for the CMB were scale invariant on all scales.
One would like to see scale invariance emerging from some (scaling) symmetry of the primordial physics that generated perturbations. A very simple and elegant solution is found by assuming that, during some primordial era, the background spacetime was well approximated by *de Sitter spacetime* (dS) (in so-called flat slicing, a.k.a. Poincaré coordinates)

$$ds^2 = \frac{-d\tau^2 + dx^i dx^j \delta_{ij}}{\tau^2 H^2} = -dt^2 + e^{2Ht} dx^i dx^j \delta_{ij}\,, \tag{1.39}$$

for some constant Hubble parameter H. This is a flat FLRW spacetime with scale factor

$$a = -\frac{1}{H\tau} = e^{Ht} \quad \text{(de Sitter)}\,. \tag{1.40}$$

For time in the interval $-\infty < t < +\infty$ (equivalently $-\infty < \tau < 0$), this scale factor describes an expanding ($\dot{a} = Ha > 0$) accelerated ($\ddot{a} = H^2 a > 0$) universe. Like Minkowski spacetime, de Sitter is also maximally symmetric. One of the ten isometries is the *dilation* symmetry

$$\tau \to \lambda\tau\,, \quad \mathbf{x} \to \lambda\mathbf{x}\,. \tag{1.41}$$

[10] Beware that this is Cosmology lingo. In other fields, such as Conformal Field Theory, the term scale invariance is refers to the rescaling of time as well as space in the correlators. Also, ϕ could have a non-vanishing conformal dimension Δ, so that $\phi(x) \to \lambda^{\Delta}\phi(x)$. In cosmology, scale invariance usually refers to $\Delta = 0$, as in (1.38).

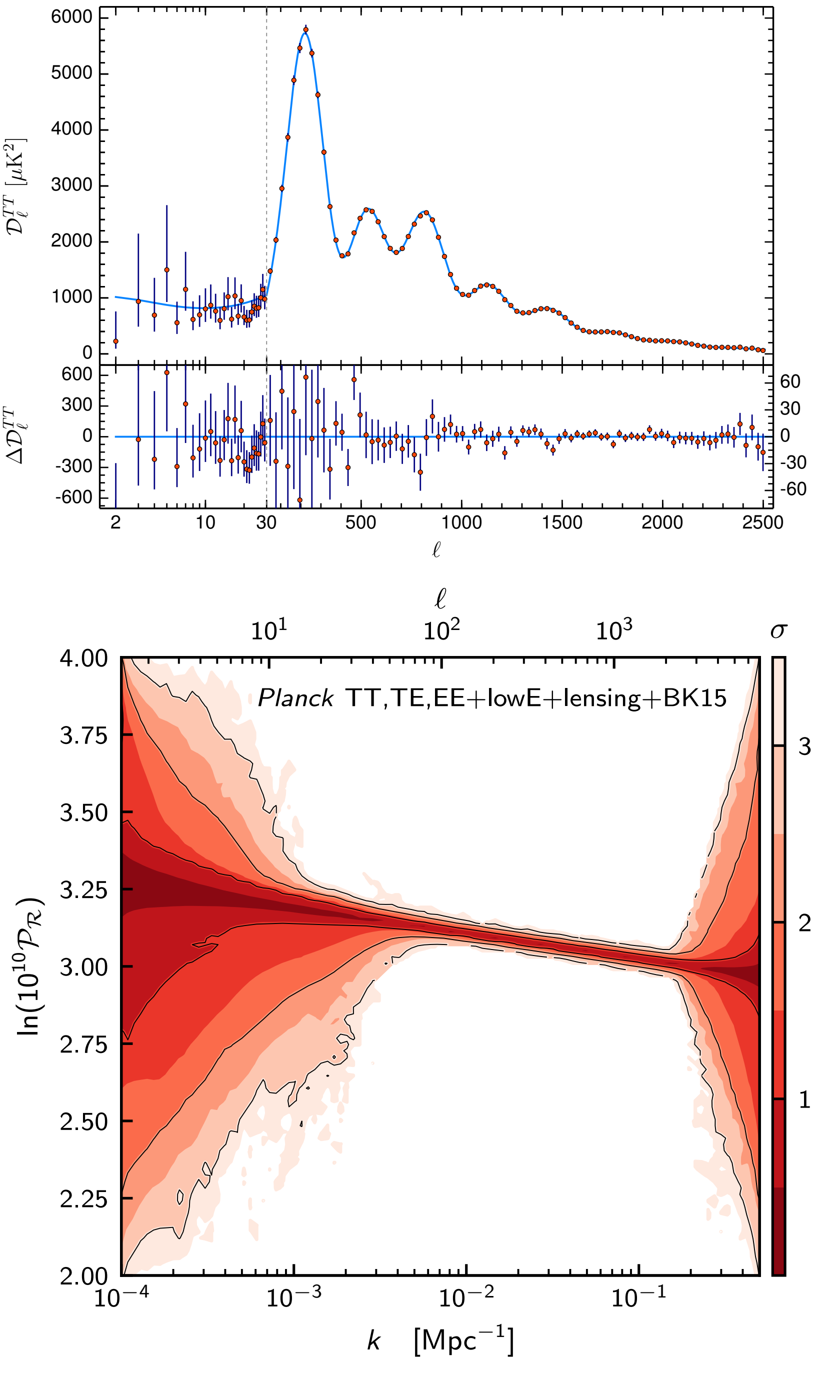


Figure 2: Observations from Planck 2018 [3]. The top plot shows the power spectrum of the CMB temperature anisotropy measured today, as a function of multipole moment l (or equivalently inverse angular scale). For $l \lesssim 50$ the angular power spectrum approximately coincides with the power spectrum of the primordial initial conditions, showing that it is indeed approximately scale invariant, i.e. constant as function of l. The bottom plot shows the same measurements as the top plot but evolved backwards in time using the ΛCDM evolution equations, i.e. a reconstructed power spectrum of the primordial initial condition. The primordial power spectrum is scale invariant up to a 4% “red tilt” (which means there is slightly more power on large scales).

If all other non-gravitational background quantities depend very weakly on time, then Eq. (1.41) is an approximate symmetry of the full theory and primordial correlators must be invariant under it. This in turn implies scale invariance. We will come back to this property after having learned to compute correlators in curved spacetime.

1.3 A prolonged phase of quasi-de Sitter expansion

The problems encountered in the previous Sec. suggest that we need a prolonged phase of accelerated expansion (curvature, horizon and phase coherence problem), with a background close to dS (scale invariance), which we will call *inflation* [50, 65, 91]. Let's see this in detail.

De Sitter spacetime is a solution of Einstein equations in the presence of a cosmological constant

$$S = \int d^4x\sqrt{-g}\frac{M_{\rm Pl}^2}{2}\left(R - 2\Lambda\right) \quad \Rightarrow \quad R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R + g_{\mu\nu}\Lambda = 0\,. \tag{1.42}$$

The trace of this expression in $D = d+1$ spacetime dimensions for $D > 2$ tells us $R = \Lambda 2D/(D-2)$ and therefore dS is an Einstein manifold, namely the Ricci tensor is proportional to the metric[11]

$$R_{\mu\nu} = \frac{2\Lambda}{D-2}g_{\mu\nu}\,. \tag{1.44}$$

However, as the name suggest, the cosmological *constant* does not change with time and an exact dS spacetime is eternal, and cannot be connected to the universe as we know it. There is an easy fix: we introduce a "clock" ϕ that "turns off" the cosmological constant Λ after some time so that the dS phase can indeed stop when desired. I will call this clock-dependent cosmological non-constant $V(\phi)$, to avoid confusing it with the cosmological constant Λ. We will describe the dynamics of ϕ shortly.
The horizon, curvature and phase coherence problems taught us that we should postulate the existence of an early phase of accelerated expansion $\ddot{a}, \dot{a} > 0$. Let's reformulate this as

$$\frac{\ddot{a}}{a} = \dot{H} + H^2 = H^2\left(1 - \epsilon\right) > 0\,, \tag{1.45}$$

where I have introduced the *first Hubble slow-roll parameter*

$$\boxed{\epsilon \equiv -\frac{\dot{H}}{H^2}}\,, \tag{1.46}$$

which is a dimensionless measure of the time variation of H. From (1.45), we recognise that acceleration requires $\epsilon < 1$. Also, as long as the matter sector satisfies the Null Energy Condition, which we will assume in the following, then $\epsilon > 0$. Observations of the Cosmic Microwave Background (CMB) and of Large Scale Structures (LSS) probe

[11] Actually, the full Riemann tensor is also given in terms of the metric

$$R_{\mu\nu\rho\sigma} = \frac{R}{12}\left(g_{\mu\rho}g_{\nu\sigma} - g_{\mu\sigma}g_{\nu\rho}\right)\,. \tag{1.43}$$

cosmological scales over roughly three orders of magnitudes, and we observe approximate scale invariance up to percent corrections. We gave a heuristic argument that scale invariance follows when the background is close to de Sitter spacetime, i.e. H is approximately constant. Quantitatively, we will therefore be interested in small deviations from dS, namely

$$0 < \epsilon \ll 1 \,. \tag{1.47}$$

To address the horizon and curvature problems, we need the phase of quasi de Sitter expansion to last for some "time". The requirement of generating a large and flat universe gives us a lower bound on how much the scale factor has to grow during inflation. This is best expressed in terms of the *number of e-foldings* N, defined as

$$dN \equiv d \log a = H dt \quad \Rightarrow \quad N - N_0 = \log\left(\frac{a}{a_0}\right) \,. \tag{1.48}$$

For inflation to solve the Hot Big Bang problems we require

$$\Delta N_{\text{infl}} \equiv N_i - N_e > 50 + \log\left(\frac{T_{\text{reh}}}{10^{10}\text{GeV}}\right) \,, \tag{1.49}$$

where N_i denotes the beginning and N_e the end of inflation and T_{reh} is the reheating temperature, namely the temperature of radiation at the beginning of the Hot Big Bang that followed inflation. For phenomenologically viable reheating temperatures $\Delta N_{\text{infl}} \in \{25 - 60\}$. I'll often use $\Delta N_{\text{infl}} \sim 50$ for numerical estimates.

We observe approximate scale invariance for about 7 of the total ΔN_{infl} efoldings of expansion, but it is natural to assume that $\epsilon \ll 1$ remains valid during the whole of inflation. To quantify this, let us re-write the definition of ϵ and generalise it to the second and higher order Hubble slow-roll parameters (all dimensionless)

$$\epsilon \equiv -\frac{\dot{H}}{H^2} = -\partial_N \ln H \,, \tag{1.50}$$

$$\eta \equiv \frac{\dot{\epsilon}}{H\epsilon} = \partial_N \ln(\epsilon) \,, \tag{1.51}$$

$$\xi_{n\geq 3} \equiv \partial_N \ln \xi_{n-1} \,, \tag{1.52}$$

with $\xi_2 \equiv \eta$ and where I used $dN = H dt$ from (1.48). Then, the Taylor expansion of ϵ around some reference time N_* is

$$\epsilon(N) - \epsilon(N_*) = \left.\frac{\partial \epsilon}{\partial N}\right|_{N_*} (N - N_*) + \left.\frac{\partial^2 \epsilon}{\partial N^2}\right|_{N_*} \frac{(N - N_*)^2}{2} + \mathcal{O}\left(\partial_N^3 \epsilon\right) \tag{1.53}$$

$$= \epsilon \left[\eta (N - N_*) + \eta \xi_3 \frac{(N - N_*)^2}{2} + \mathcal{O}\left(\eta^3, \eta^2 \xi_3, \eta \xi_3 \xi_4, \epsilon\right)\right] \,, \tag{1.54}$$

where all the slow-roll parameters are evaluated at N_*. The requirement that ϵ does not change much during inflation is then $\eta \Delta N_{\text{infl}} < 1$, $\xi_3 \eta \Delta N_{\text{infl}} < 1$, and so on. Altogether,

$$\epsilon, \eta, \xi_n \ll 1 \quad \Rightarrow \quad \text{Slow-roll inflation} \,. \tag{1.55}$$

1.4 An Example: Single-field inflation in $P(X,\phi)$ theories

In the previous subsection, we have characterised the expansion history during inflation. We now want to ask how such an expansion history can emerge dynamically, from solving the equations of motion.

Since spatial curvature decays quickly during inflation, from now on we will assume $K = 0$. To try to mimic a cosmological constant, we were led to consider the action of scalar field coupled to gravity. We will assume the rather general action, which I will call[12] a $P(X,\phi)$ (read "P of X and ϕ" or simply "P of X") theory,

$$S = \int \sqrt{-g}\left[\frac{1}{2}M_{\rm Pl}^2 R + P(X,\phi)\right] , \tag{1.56}$$

where X represents the standard canonical kinetic term,

$$X \equiv -\frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi = \frac{1}{2}\left(\dot\phi^2 - \frac{\partial_i\phi\,\delta^{ij}\,\partial_j\phi}{a^2}\right) , \tag{1.57}$$

and $P(X,\phi)$ is a generic function of its two arguments. For example, a so-called *canonical* scalar field corresponds to the simple choice

$$P = X - V(\phi) \quad \text{(canonical scalar)} , \tag{1.58}$$

where $V(\phi)$ is some potential. The coupling between the scalar field and gravity is called *minimal*, because it simply arises from writing down a Lorentz-invariant action in Minkowski spacetime and substituting $d^4x \to d^4x\sqrt{-g}$ and $\eta_{\mu\nu} \to g_{\mu\nu}$. I will not discuss here non-minimal couplings, such as for example $R\phi^2$ or $R^{\mu\nu\rho\sigma}\partial_\mu\phi\partial_\nu\phi\partial_\rho\phi\partial_\sigma\phi$. The energy-momentum tensor (1.5) is then

$$T_{\mu\nu} = P_{,X}\partial_\mu\phi\partial_\nu\phi + g_{\mu\nu}P(X,\phi) . \tag{1.59}$$

This takes the same form as the energy-momentum tensor of a perfect fluid, (1.12), under the following identifications

$$\rho = 2XP_{,X} - P , \qquad p = P , \qquad u_\mu = -\frac{\partial_\mu\phi}{\sqrt{2X}} . \tag{1.60}$$

Let us start by focussing on the homogeneous background dynamics, $\phi = \bar\phi(t)$. The homogeneous equation of motion is

$$\ddot{\bar\phi}\left(P_{,X} + 2XP_{,XX}\right) + 3H\dot{\bar\phi}P_{,X} + \left(2XP_{,X\phi} - P_{,\phi}\right) = 0 , \tag{1.61}$$

while the Friedmann and acceleration equations read

$$3M_P^2H^2 = 2XP_{,X} - P , \qquad -M_P^2\dot H = XP_{,X} . \tag{1.62}$$

The specific choice of P is irrelevant for us as we will not be solving any of these equations. It suffices to notice that

$$\epsilon = \frac{3XP_{,X}}{2XP_{,X} - P} , \tag{1.63}$$

and so there are (many) choices of the function P that support a prolonged phase of slow-roll inflation, i.e. $\epsilon, \eta \ll 1$ (a necessary condition is $P_{,\phi} \neq 0$ [42]).

[12]This class of theories goes under many different names in the literature depending on the context: P-of-X theory, k-inflation [7], k-essence [8] and, specifically when $P(X,\phi) = P(X)$, superfluid [90].

2 Free fields on curved backgrounds

ref

According to our current leading paradigm, the *quantum* fluctuations on top of the classical inflationary background are the seeds of the structures that we see in our universe today. Hence we would like to quantize the inflationary model discussed above. To this end, we start from the homogeneous background $\bar{g}_{\mu\nu}$ and $\bar{\phi}$ that we discussed in the previous section, where $\bar{g}_{\mu\nu}$ is the FLRW spacetime whose scale factor $a(t)$ is given by the solution of the Friedmann and acceleration equations (1.62), and $\bar{\phi}$ obeys (1.61). Then we want to promote fluctuations to quantum operators

$$g_{\mu\nu}(t,\mathbf{x}) = \bar{g}_{\mu\nu}(t) + \hat{h}_{\mu\nu}(t,\mathbf{x})\,, \quad \phi(t,\mathbf{x}) = \bar{\phi}(t) + \hat{\varphi}(t,\mathbf{x})\,. \tag{2.1}$$

We will achieve this in steps. First, in this Sec. we will quantize free theories. Second, in Sec. 3 we will learn the formalism to quantize general interacting theories, perturbatively. Third, we will apply this formalism to a scalar field theory in Sec. 4 and finally to gravity in Sec. 5.

2.1 Massless scalar in de Sitter

A good starting point to understand more realistic models of inflation is a massless[13] scalar field ϕ in de Sitter spacetime without any classical background $\bar{\phi}(t) = 0$. This can arise for example by simply taking $P = X - V$ with $\phi = 0$ a minimum of V. Consider the action

$$S = -\int \sqrt{-g}\frac{1}{2}\partial_\mu\varphi\partial^\mu\varphi = \int d^3x dt\, a^3 \frac{1}{2}\left(\dot{\varphi}^2 - \frac{1}{a^2}\partial_i\varphi\partial_i\varphi\right)\,, \tag{2.2}$$

In Fourier space, this free theory reduces to an infinite sum of decoupled harmonic oscillators

$$S = \int \frac{d^3k}{(2\pi)^3} dt\, a^3 \frac{1}{2}\left[\dot{\varphi}(\mathbf{k})\dot{\varphi}(-\mathbf{k}) - \frac{k^2}{a^2}\varphi(\mathbf{k})\varphi(-\mathbf{k})\right]\,, \tag{2.3}$$

where

$$\varphi(\mathbf{x}) = \int_{\mathbf{k}} \varphi(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}}, \quad \varphi(\mathbf{k}) = \int_{\mathbf{x}} \varphi(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}}\,, \tag{2.4}$$

with

$$\int_{\mathbf{k}} \equiv \int \frac{d^3\mathbf{k}}{(2\pi)^3}\,, \quad \int_{\mathbf{x}} \equiv \int d^3\mathbf{x}\,. \tag{2.5}$$

To quantize the theory, we promote φ to an operator (but we will omit the hat). As for the harmonic oscillator, we write φ in terms of creation and annihilation operators[14]

$$\varphi(\mathbf{k},t) = f_k(t) a_{\mathbf{k}} + f_k^*(t) a^\dagger_{-\mathbf{k}}\,, \tag{2.6}$$

which satisfy

$$[a_{\mathbf{k}}, a_{\mathbf{k}'}] = 0\,, \qquad [a_{\mathbf{k}}, a^\dagger_{\mathbf{k}'}] = (2\pi)^3 \delta^3_D\left(\mathbf{k} - \mathbf{k}'\right)\,. \tag{2.7}$$

[13] See Footnote 19 for more details on the meaning of "mass" in de Sitter.

[14] Notice that $\varphi^*(\mathbf{k}) = \varphi(-\mathbf{k})$, as required by the fact that $\varphi(\mathbf{x})$ is a real field.

Here we have chosen to work in the Heisenberg picture, in which operators depend on time while the states do not, so $\varphi = \varphi(\mathbf{k}, t)$ (but I'll omit the time argument when no ambiguity arises). All the time dependence of $\varphi(t, \mathbf{k})$ has been collected in $f_k(t)$ and $f_k^*(t)$, which are known as *mode functions*. They are determined by requiring that φ solves the equations of motion[15] derived from (2.3)

$$\ddot{f}_k + 3H\dot{f}_k + \frac{k^2}{a^2} f_k = 0\,. \tag{2.9}$$

Because of the isotropy of the background f_k depends only on the norm of $\mathbf{k}$, as suggested by the notation. This equation becomes more familiar if we use conformal time ($' \equiv \partial_\tau$)

$$(af_k)'' + \left(k^2 - \frac{a''}{a}\right)(af_k) = 0\,, \tag{2.10}$$

where it looks like a harmonic oscillator with a time dependent mass $a''/a = 2/\tau^2$. The two linearly independent solutions are the complex conjugate of each other

$$f_k = \alpha(1 + ik\tau)e^{-ik\tau} + \beta(1 - ik\tau)e^{ik\tau}\,. \tag{2.11}$$

The integration constants α and β must be compatible with the canonical commutation relations. Let Π be the momentum conjugate of φ,

$$\Pi(\mathbf{x}) \equiv \frac{\partial\mathcal{L}}{\partial_\tau\varphi(\mathbf{x})} = a^2(\tau)\partial_\tau\varphi(\mathbf{x})\,. \tag{2.12}$$

Then, canonical quantization imposes the constraint

$$[\hat{\varphi}_{\mathbf{k}}, \hat{\Pi}_{\mathbf{k}'}] = a^2\left(f_k f_k'^* - f_k^* f_k'\right)(2\pi)^3\delta^3\left(\mathbf{k} + \mathbf{k}'\right) \stackrel{!}{=} i(2\pi)^3\delta^3\left(\mathbf{k} + \mathbf{k}'\right)\,, \tag{2.13}$$

which in turn implies

$$|\alpha|^2 - |\beta|^2 = \frac{H^2}{2k^3}\,. \tag{2.14}$$

Initial conditions To complete determine the integration constants[16] α and β we notice that in the far past, i.e. for $k\tau \gg 1$, (2.10) reduces to the Klein-Gordon equation for the field (af_k), since $k^2 \gg a''/a$. In other words, the field $(a\varphi(\mathbf{k}))$ at early times lives effectively in Minkowski spacetime. In this limit we expect to recover the (Heisenberg picture) free scalar field that we learn about in introductory QFT,

$$\varphi(\mathbf{x}, t) = \int \frac{d^3k_p}{(2\pi)^3}\frac{e^{i\mathbf{k}_p\cdot\mathbf{x}}}{\sqrt{2k_p}}\left[e^{-ik_pt}a_{\mathbf{k}_p} + e^{ik_pt}a^\dagger_{-\mathbf{k}_p}\right] \quad \text{(Minkowski)}\,, \tag{2.16}$$

[15] Or equivalently the Heisenberg equation

$$\dot{\varphi}(\mathbf{k}) = i[H, \varphi(\mathbf{k})]\,. \tag{2.8}$$

where H is the Hamiltonian derived from the action (2.3).

[16] An alternative derivations goes as follows. Inverting (2.6) for $\hat{a}_{\mathbf{k}}$,

$$\hat{a}_{\mathbf{k}} = af_k^*\left(\frac{i\hat{\Pi}_{\mathbf{k}}}{a} - \frac{i\partial_\tau f_k^*}{f_k^*}a\hat{\varphi}_{\mathbf{k}}\right) \tag{2.15}$$

we find that $\hat{a}_{\mathbf{k}}$ annhilates the vacuum in the infinite past iff $-i\partial_\tau f_k^* \to f_k^* k$ as $\tau \to -\infty$. This implies that $f_k^* \sim e^{+ik\tau}$ asymptotically, which can be recognised as the mode function on Minkowski space for particles with positive energy $+k$. This requirement fixes $\beta = 0$.

where k_p is the physical wave number, related to the comoving one at some time by $k_p = k/a$. Recall that this choice of time dependence means that $\varphi(\mathbf{k}_p) \supset e^{+ik_p t} a^\dagger_{\mathbf{k}_p}$ *creates particles of positive energy*, as can be check from

$$\hat{H}\varphi(\mathbf{k})\left|0\right\rangle = [\hat{H}, \varphi(\mathbf{k})]\left|0\right\rangle + \varphi(\mathbf{k})\hat{H}\left|0\right\rangle = -i\dot{\varphi}(\mathbf{k})\left|0\right\rangle = +k\varphi(\mathbf{k})\left|0\right\rangle , \tag{2.17}$$

where the hat distinguishes the Hamiltonian $\hat{H}$ from the Hubble parameter H, and in the second step we used the Heisenberg equation and $\hat{H}\left|0\right\rangle = 0$.
To find α and β we thefore match the solution (2.11) of the de Sitter equation of motion and its time derivative to the Minkowski one, (2.16), at some early time τ_* such that $k\tau_* \gg 1$

$$af_k\Big|_{\tau=\tau_*} = \frac{e^{ik_p t}}{\sqrt{2k_p}}\Big|_{t=t_*} , \tag{2.18}$$

$$\partial_t(af_k)\Big|_{\tau=\tau_*} = \partial_t\left(\frac{e^{ik_p t_*}}{\sqrt{2k_p}}\right)\Big|_{t=t_*} . \tag{2.19}$$

Solving the linear system for α and β one finds

$$\alpha = ie^{ik\tau_*(1+Ht_*)}\frac{H}{\sqrt{2k^3}}\left[1 + \frac{i}{k\tau_*} - \frac{1}{2(k\tau_*)^2}\right] , \tag{2.20}$$

$$\beta = ie^{-ik\tau_*(1-Ht_*)}\frac{H}{\sqrt{2k^3}}\frac{1}{2(k\tau_*)^2} . \tag{2.21}$$

If the matching is done in the infinite past, $k\tau_* \to -\infty$, this reduces simply to

$$\lim_{\tau\to-\infty}|\alpha| = \frac{H}{\sqrt{2k^3}} , \qquad \lim_{\tau\to-\infty}\beta = 0 . \tag{2.22}$$

The normalization of α is fixed only up to an overall phase because one can always shift t in Minkowski and so the value of t_* is arbitrary. The dS mode functions that create positive-energy particles in the infinite past therefore are

$$\boxed{f_k = \frac{H}{\sqrt{2k^3}}(1 + ik\tau)e^{-ik\tau}} \quad \text{(dS mode functions)} . \tag{2.23}$$

The dS mode functions (2.23) are very different from the Minkowski counterpart when the physical wavenumber becomes smaller than the comoving Hubble parameter

$$k < k_{\text{H.c.}} = aH = \frac{1}{|\tau|} \quad \text{(Hubble crossing)} , \tag{2.24}$$

where "H.c." stands for *Hubble crossing*, sometime also called horizon crossing. In physical length scales, this means the physical wavelength $\lambda_p = a/k$ is stretched by the expansion to be become larger than the Hubble radius $1/H$. Since k and H are (approximately) constant, while $a = e^{Ht}$ grows with time, all modes cross the Hubble radius as time proceeds and become "superHubble" modes. Unlike "subHubble" modes, $k \gg aH$, which oscillate, superHubble modes *freeze out* and asymptotes a constant.

More rigorous ways to derive (2.23) include Hamiltonian minimization, i.e. choosing as vacuum the lowest energy state in the asymptotic past and matching to the uniquely defined Euclidean vacuum of the Wick rotated Euclidean field theory. Now that we have related φ to creation and annihilation operators, we can specify the "vacuum state" $|0\rangle$ by the usual condition defining the "Fock vacuum" $a_{\mathbf{k}}\,|0\rangle = 0$ for all $\mathbf{k}$'s. In this context, the state $|0\rangle$ is called the Bunch-Davies vacuum[17]. Excited states are then obtained by acting with creation operators on this vacuum.

Expectation values and the power spectrum What observables can we compute for this theory? As familiar from Quantum Mechanics, observables are given by the expectation value of operators. In cosmology, we have observational access only to these expectation values in the infinite future $k\tau \to 0$. In this limit, observables become approximately constant and so we will only be interested in the expectation value of product of correlators at equal time, or simply cosmological correlators for short

$$\lim_{\tau\to 0}\langle \mathcal{O}(\mathbf{k}_1,\tau)\dots\mathcal{O}(\mathbf{k}_n,\tau)\rangle\,, \tag{2.25}$$

for some local operators $\mathcal{O}$. Because we are studying a *free* theory, all information is contained in the two-point correlators of φ and its conjugate momentum π. All odd-point correlators vanish by the symmetry $\varphi \to -\varphi$ and all higher even-point correlators can be reduced to the two-point one using Wick's theorem.

Box 2.1 Free theories and Gaussianity All free theories can be understood by analogy with the most famous free theory, the quantum harmonic oscillator. In quantum mechanics you learn that the probability of finding a particle at position x is a Gaussian distribution

$$\mathrm{Prob}(x) \sim |\psi(x)|^2 \propto e^{-x^2/(2\sigma^2)}\,, \tag{2.26}$$

where $\psi(x)$ is the position space wavefunction and $\sigma^2 = \langle x^2\rangle$ describes the width of the Gaussian and is fixed by the parameters of the theory ($\sigma^2 = (2m\omega)^{-1}$). It is clear by parity that

$$\langle x^{2n+1}\rangle \propto \int_{-\infty}^{\infty} dx\,|\psi(x)|^2 x^{2n+1} \sim \int dx\, e^{-x^2/(2\sigma^2)} x^{2n+1} = 0\,. \tag{2.27}$$

While with repeated integration by parts we can always rewrite $\langle x^{2n}\rangle$ in terms of $\langle x^2\rangle^n$. Because of this, the expression free theory and Gaussian theory are often used interchangeably. In the next Sec. we will study interacting theories, where we will compute *non-Gaussianities*, i.e. deviations from a Gaussian wavefunction. This discussion applies to quantum*field* theory (as opposed to quantum mechanics) by thinking of $\phi(\mathbf{k})$ as an infinite collection of decoupled harmonic oscillators, as depicted in Fig. 3

Let's compute the two-point correlation function of φ:

$$\lim_{\tau\to 0}\langle \varphi(\mathbf{k},\tau)\varphi(\mathbf{k}',\tau)\rangle = |f_k(,\tau)|^2\langle a_{\mathbf{k}} a^\dagger_{-\mathbf{k}'}\rangle \tag{2.28}$$

$$= (2\pi)^3\delta_D^3\left(\mathbf{k}+\mathbf{k}'\right)P(k,\tau) \tag{2.29}$$

with

$$P(k,\tau) = \frac{H^2}{2k^3}\left[1+(k\tau)^2\right]\,. \tag{2.30}$$

[17] It also coincides with the Hartle-Hawking state encountered when solving the Wheeler de Witt equation

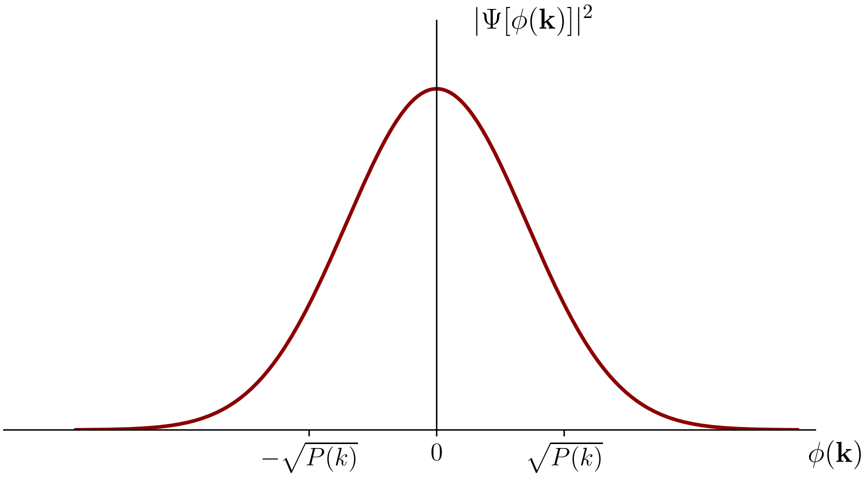


Figure 3: The figure show that the probability distribution function of a free field theory, which is proportional to the norm square of the wavefunction, is a multivariate Gaussian in the infinitely many decoupled harmonic oscillators $\phi(\mathbf{k})$, each with variance $P(k)$.

In the late time limit, or equivalently in the limit in which we are looking at a perturbation well outside the Hubble radius $k|\tau| \ll 1$, this reduced to the iconic k^{-3} power spectrum

$$\lim_{\tau\to 0} P(k,\tau) \equiv P(k) = \frac{H^2}{2k^3}\,. \tag{2.31}$$

Here, we have introduced the *power spectrum* $P(k)$, which is just the two-point correlator stripped of the Dirac delta and its accompanying factor of $(2\pi)^3$. A few comments are in order:

- The Dirac delta reminds us that momentum is conserved as a consequence of the homogeneity of the background. In Example Sheet 1 you will show that this delta appears in all correlators. Pictorially we can imagin that perturbations in this state must exist in pairs of opposite wavenumber $\mathbf{k}$ and $-\mathbf{k}$.
- $P(k)$ does not depend on the direction of $\mathbf{k}$ as consequence of the isotropy of the background.
- As we will see shortly, the fact that the power spectrum asymptotes some (non-vanishing) constant value as $\tau \to 0$ is related to the absence of a mass.
- The k-dependence $P \propto k^{-3}$ is the one corresponding to *scale invariance*, as defined in (1.38). To see this, we can Fourier transform to the real-space *correlation*

function,[18]

$$\langle\varphi(\mathbf{x})\varphi(\mathbf{0})\rangle = \int_{\mathbf{k}\mathbf{k}'}\langle\varphi(\mathbf{k})\varphi(\mathbf{k}')\rangle \sim H^2\,, \tag{2.33}$$

and notice that the correlation does not depend on distance. In particular, it doesn't change if we rescale $\mathbf{x} \to \lambda\mathbf{x}$.

It is worth comparing the power spectrum in dS, (2.31), with that in Minkowski

$$\langle\varphi(\mathbf{k})\varphi(\mathbf{k}')\rangle = (2\pi)^3\delta_D^3\left(\mathbf{k}+\mathbf{k}'\right)\frac{1}{2k} \quad \text{(Minkowski)}\,. \tag{2.34}$$

and the associated real-space correlation function

$$\langle\varphi(\mathbf{x})\varphi(\mathbf{0})\rangle = \int_{\mathbf{k}\mathbf{k}'}\langle\varphi(\mathbf{k})\varphi(\mathbf{k}')\rangle \sim \frac{1}{x^2} \quad \text{(Minkowksi)}\,. \tag{2.35}$$

So in dS the correlation function is independent of the distance, while in Minkowski it decays as $1/x^2$, as expected for a massless particle. In the Example Sheet 1, you will derive the correlators involving the momentum conjugate.

2.2 Massive scalar in de Sitter

It is interesting to ask what changes if the scalar field has a mass[19] m,

$$S = -\int\sqrt{-g}\frac{1}{2}\left[\partial_\mu\varphi\partial^\mu\varphi + m^2\varphi^2\right]\,. \tag{2.36}$$

You will quantize this theory in Example Sheet 1. The relation to creation and annihilation is the same

$$\varphi(\mathbf{k},t) = f_k(t)a_{\mathbf{k}} + f_k^*(t)a^\dagger_{-\mathbf{k}}\,, \tag{2.37}$$

but now the mode functions are modified to

$$f_k(\tau) = i\frac{\sqrt{\pi}H}{2}(-\tau)^{3/2}H_\nu^{(1)}(-k\tau)\,,\quad \nu = \sqrt{\frac{9}{4}-\frac{m^2}{H^2}}\,, \tag{2.38}$$

where $H^{(1)}$ is the Hankel function of the first kind and we are free to multiply this by any complex phase $e^{i\eta}$. Hankel functions are solutions of the Bessel's differential equation and are linear combination of Bessel functions and in general they cannot be

[18] Actually the dS correlation function at separated points, $\mathbf{x} \neq 0$, is IR divergent. The physical reason is that dS is eternal. This divergence can be regularized either with a small tilt of the power spectrum $k^{-(3+\delta)}$, with $0 < \delta \ll 1$ or with an IR cutoff $k_{\rm IR}$ of the integral ($\tilde{k} = xk$ is dimensionless)

$$\langle\varphi(\mathbf{x})\varphi(\mathbf{0})\rangle = \frac{H^2}{(2\pi)^2}\int_{xk_{\rm IR}}^{\infty} d\tilde{k}\,\frac{\sin\tilde{k}}{\tilde{k}^2} \xrightarrow{xk_{\rm IR}\to 0} H^2\left[\gamma_E - 1 + \log(xk_{\rm IR}) + \mathcal{O}((xk_{\rm IR})^2)\right]\,. \tag{2.32}$$

.

[19] The insightful reader might wonder about the true meaning of mass in dS. Indeed, in Minkowski, mass can be defined very generally as one of the two parameters (eigenvalues of the Casimir operators) classifying the unitary irreps of the Poincaré group (the other parameter being spin for $m^2 > 0$ or helicity for $m = 0$). In dS this is not possible (see [46] for an in depth discussion). Here by mass we simply mean the parameter m appearing in (2.36). In particular we assume that the scalar field is minimally coupled to gravity, e.g. there is no coupling to the Ricci scalar of the form $\xi R\phi$, and scale factor appears only because of $\sqrt{-g}$ and $g^{\mu\nu}$.

expressed in terms of elementary functions. However when ν is half integer, $H_\nu^{(1,2)}(x)$ reduces to an oscillating phase $e^{\pm ix}$ times $x^{-\nu}$ times a polynomial in x. An interesting case is $m = \sqrt{2}H$ so that $\nu = 1/2$ and

$$f_k(\tau) = \frac{H}{\sqrt{2k}} \tau e^{-ik\tau} . \tag{2.39}$$

This value of the mass corresponds to a conformally coupled scalar, namely a scalar non-minimally coupled to gravity with a $\varphi^2 R$ term in such a way that the full theory is invariant under a Weyl rescaling of the metric $g_{\mu\nu} \to \Omega(x) g_{\mu\nu}$. Since the de Sitter metric is proportional to that of Minkowski, where the Ricci scalar vanishes, this theory is equivalent to a massless scalar in Minkowski. This explains why the conformally coupled mode function in (2.39) is just the rescaled Minkowski mode function in (2.18). Back to general ν. For $\tau \to 0$ this becomes

$$f_k(\tau \to 0) = H\tau^{3/2} \left[\frac{\sqrt{\pi}(-k\tau)^\nu}{2^{1+\nu}\Gamma[\nu+1]} (1 + i\cot(\pi\nu)) - \frac{i(-k\tau)^{-\nu}}{\sqrt{\pi} 2^{1-\nu} \Gamma[\nu]} \right] + \dots . \tag{2.40}$$

Now it is useful to distinguish two cases. The first case is when the mass square is small or negative, $m^2 < 9H^2/4$. Then ν is real and positive. In this case, the first term in brackets approaches zero faster than the second and can be neglected. So the power spectrum now becomes

$$P(k) = |f_k|^2 = \frac{H^2}{\pi 2^{2(\nu-1)} \Gamma(\nu)^2} \frac{(-k\tau)^{3-2\nu}}{k^3} \quad (\text{for } m^2 < \frac{9}{4} H^2) . \tag{2.41}$$

Because of the mass, the power spectrum is not scale invariant anymore, $P \propto k^{-2\nu}$. Also, P has acquired a time dependence. For positive $m^2 > 0$, one finds $3 - 2\nu > 0$ and the power spectrum decays with time and vanish at future infinity. This is to be expected because the quadratic potential pushes the field towards $\varphi = 0$. For negative m^2 we would expect an instability and indeed the power spectrum grows with time and diverges at future infinity.
The second case is when the mass square is large and positive, $m^2 > 9H^2/4$, then ν becomes complex and the two terms in the brackets of (2.40) are of the same order. The power spectrum oscillates while decaying as τ^3. In cosmology, we are mostly interested in massless or almost massless fields, which do not create large instability and whose perturbations survive long enough to be observable at late times.

2.3 Particle creation*

For QFT in Minkowski, we can think of excitations generated by the creation operators as particles. However, in curved spacetime particle and particle number are more subtle concepts. Let's see this in detail. We found field excitations in dS oscillate as $f_k \sim e^{-ik\tau}$, where the comoving wavenumber k is related to the Minkowski energy by

$$E = \sqrt{k_i k_j g^{ij}} = \frac{k}{a} . \tag{2.42}$$

Let us now Taylor expand the time-dependent phase of f_k in time around some time t_*:

$$\begin{aligned} -ik\tau &= i\frac{k}{aH} = i\frac{k}{a_*}\frac{a_*}{aH} \\ &= i\frac{E}{H} e^{-H(t-t_*)} \simeq iE \left[\frac{1}{H} - (t - t_*) + \frac{1}{2} H (t - t_*)^2 + \dots \right] . \end{aligned} \tag{2.43}$$

The first term in brackets is an irrelevant phase. The second term is precisely the time dependence of particles in Minkowski, namely e^{-iEt}. So field excitations in dS have a chance to look like particles only for a time interval $(t - t_*) \ll 1/H$, during which we can neglect the higher order terms in brackets. Moreover, we must demand that during this interval, the wavefunction oscillates many times, so $E(t - t_*) \gg 1$. Using again (2.43) this requires $E/H = -k\tau \gg 1$. When this condition is not satisfied, the energy and momentum of the state are redshifted by the expansion before a single oscillation of the wavefunction.

Even when particles can be defined, the expansion of the universe can create particles (unless there is a conserved quantum number) at a rate controlled again by H. This is to be expected as the expansion of the universe breaks time translations and so energy is not conserved. Let's see this in more detail.

In the previous section, we found the mode functions by demanding that φ creates positive-energy particles at $k\tau \to -\infty$. Let's instead require that φ creates positive-energy particles at some finite τ_*, still satisfying $|k\tau_*| \gg 1$. By matching to the Minkowski vacuum at τ_*, we find α and β as given in (2.20). The quantized field then takes the form

$$\varphi(\mathbf{k}) = g_k b_{\mathbf{k}} + g_k^* b_{-\mathbf{k}}^\dagger \tag{2.44}$$

with[20]

$$\begin{aligned} g_k &= \alpha f_k(\tau) + \beta f_k^*(\tau) \\ &= \frac{H}{\sqrt{2k^3}} \left[1 + \frac{i}{k\tau_*} - \frac{1}{2(k\tau_*)^2}\right] f_k(\tau) + e^{-2ik\tau_*} \frac{H}{\sqrt{2k^3}} \frac{1}{2(k\tau_*)^2} f_k^*(\tau)\,, \end{aligned} \tag{2.45}$$

and $\{b_{\mathbf{k}}, b_{\mathbf{k}}^\dagger\}$ a new set of creation and annihilation operators, which define a new vacuum state by $b_{\mathbf{k}} |\tilde{0}\rangle = 0$. By matching the two expressions for $\varphi(\mathbf{k})$, (2.6) and (2.44), we see that the two sets of ladder operators are related,

$$a_{\mathbf{k}} = \frac{\sqrt{2k^3}}{H} \left(\alpha b_{\mathbf{k}} + \beta^* b_{-\mathbf{k}}^\dagger\right), \qquad a_{\mathbf{k}}^\dagger = \frac{\sqrt{2k^3}}{H} \left(\beta b_{-\mathbf{k}} + \alpha^* b_{\mathbf{k}}^\dagger\right), \tag{2.46}$$

This relation is called a *Bogoliubov transformation*. It can be inverted to give

$$b_{\mathbf{k}} = \frac{\sqrt{2k^3}}{H} \left(\alpha^* a_{\mathbf{k}} + \beta^* a_{-\mathbf{k}}^\dagger\right), \qquad b_{\mathbf{k}}^\dagger = \frac{\sqrt{2k^3}}{H} \left(\beta a_{-\mathbf{k}} + \alpha a_{\mathbf{k}}^\dagger\right), \tag{2.47}$$

Now we want to ask what a detector that measures $b_k^\dagger$ excitations would measure in the Bunch Davies vacuum. To this end, we define the "b-particle" number operator

$$N_b(\mathbf{k}) = b_{\mathbf{k}}^\dagger b_{\mathbf{k}}\,. \tag{2.48}$$

As expected, this operator has a vanishing expectation value in the $|\tilde{0}\rangle$ state. But if we compute its expectation value in the Bunch-Davies vacuum $|0\rangle$ we find

$$\langle 0| N_b(\mathbf{k}) |0\rangle = \frac{2k^3}{H^2} |\beta_k|^2 (2\pi)^3 \delta_D^3(\mathbf{0}) = \frac{1}{4(k\tau)^4} (2\pi)^3 \delta_D^3(\mathbf{0})\,. \tag{2.49}$$

[20] Again, we used the arbitrariness in t_* to multiply g_k by a convenient phase.

The singular factor $\delta_D^3(\mathbf{0})$ is a reminder that we are working with an infinite volume

$$(2\pi)^3\delta_D^3(\mathbf{0}) = \lim_{V\to\infty}\int_V d^3x e^{-i\mathbf{0}\cdot\mathbf{x}} = \lim_{V\to\infty} V\,. \tag{2.50}$$

It is therefore wise to compute the number density of particles, $n_b(\mathbf{k}) \equiv N_b(\mathbf{k})/V$, instead of the total number $N_b(\mathbf{k})$. We find

$$\langle 0|\, n_b(\mathbf{k})\, |0\rangle = \frac{1}{4(k\tau)^4} \neq 0\,. \tag{2.51}$$

In words, the Bunch-Davies "vacuum" state is found to contain some b-particles, defined with respect to the "vacuum" state at some large but finite $|\tau_*|$. As we take $\tau_* \to -\infty$, $|\tilde{0}\rangle$ approaches $|0\rangle$ and indeed the number density of particles vanishes as expected. We can say that the expansion of the universe creates particles. These particles are always created in pairs of opposite wavenumber, to conserve momentum, $n(-\mathbf{k}) = n(\mathbf{k})$.

3 Interacting fields and the in-in formalism ref

Now that we understand free fields, we can describe weakly interacting fields as we do in particle physics, i.e. in a perturbative expansion. Indeed, we already know that gravity is a non-linear theory and so at least we should see gravitational interactions at play in the early universe. Moreover, we don't know the laws of physics at very high energies, so it is possible that the inflaton had other, non-gravitational interactions as well. While free theories are fully characterized by their power spectra, interacting theories have an enormously richer phenomenology, which we will start exploring here. In particular interactions leads to correlators that are not fixed by the power spectrum. Since the wavefunction of is non-Gaussian (see Box 1), these correlators are often called *non-Gaussianities.* In this section, we'll set up the general formalism to compute correlators in interacting theories in Sec. 3.1 and 3.2 and provide two explicit examples in Sec. B.

3.1 Particle physics and scattering amplitudes

A highly effective way to study an object is to throw things at it and see how they bounce off. This describes mundane activities such as looking at things by scattering photons. But it also applies to more advanced "imaging" techniques such as X-ray radiography, electron microscopes and particle accelerators, just to name a few. In the quantum mechanical context the main object of study are scattering amplitudes, namely quantum mechanical amplitudes for the schematic process

$$S_{\alpha\beta} \equiv \langle \alpha, out|\beta, in\rangle \equiv \langle \alpha; +\infty|\beta; -\infty\rangle_S = \langle\alpha|\, S\, |\beta\rangle_H \tag{3.1}$$

where $|\alpha\rangle$ and $|\beta\rangle$ are eigenstates of the free Hamiltonian, and the subscripts S and H refer to the Schrödinger and Heisenberg pictures, respectively[21]. S is a unitary operator, $SS^\dagger = \mathbb{I}$, while $S_{\alpha\beta}$ is a unitary matrix with complex entries, $S_{\alpha\beta}S^\dagger_{\beta\gamma} = \delta_{\alpha\gamma}$.

[21] Recall that in the Schrödinger (S) and Heisenberg (H) pictures

$$|\psi, t\rangle_S = e^{-i\int \hat{H}dt}\,|\psi, t_i\rangle_S\,, \qquad \mathcal{O}_S(t) = \mathcal{O}_S(t_i) \equiv \mathcal{O}_S\,, \tag{3.2}$$

$$|\psi\rangle_H = |\psi, t_i\rangle_S = e^{+i\int \hat{H}dt}|\psi, t\rangle_S\,, \qquad \mathcal{O}_H(t) = e^{i\int \hat{H}dt}\mathcal{O}_S\, e^{-i\int \hat{H}dt}\,. \tag{3.3}$$

for some reference initial time t_i.

Given a Hamiltonian $\hat{H}$, one finds the largest number of operators that commute with $\hat{H}$, i.e. a subset of the symmetries of theory, and uses their eigenvalues to label the α and β states. For example, in particle physics single particles states are irreducible representations of the Poincaré group and are classified by their four-momentum p^μ and their *spin* if massive or *helicity* if massless (i.e. the irreps of the associated little group, see e.g. Chapter 3 of [103]). For example, the scattering of 2 into $(n-2)$ particles has amplitude

$$S_{\alpha\beta} = 2^{n/2}\sqrt{E_1 E_2 \dots E_n}\, \langle\Omega|\, a_{\mathbf{p}_3}(\infty)\dots a_{vp_n}(\infty) a^\dagger_{\mathbf{p}_1}(-\infty) a^\dagger_{\mathbf{p}_2}(-\infty)\, |\Omega\rangle \ , \tag{3.4}$$

where $|\Omega\rangle$ is the Minkowski vacuum, $E(\mathbf{p}) = \sqrt{\mathbf{p}^2 + m^2}$, and we used the relativistic normalization of states. Probabilities are obtained by squaring amplitudes

$$\text{Prob} \sim |\, \langle\alpha|\, S\, |\beta\rangle\, |^2 \,. \tag{3.5}$$

For most systems of physical interest, the S-matrix can only be computed in perturbation theory. Let's assume that the Hamiltonian of the theory can be divided into a free Hamiltonian $\hat{H}_0$ and an interaction Hamiltonian

$$\hat{H} = \hat{H}_0 + \hat{H}_{int} \,, \tag{3.6}$$

which induces a small perturbation. To study the effect of $\hat{H}_{int}$ it is convenient to introduce the *interaction picture*, labelled by I, where operators evolve with the free Hamiltonian $\hat{H}_0$ and states evolve with the interaction Hamiltonian:

$$|\psi, t\rangle_I = e^{+i\int \hat{H}_0 dt} |\psi, t\rangle_S \,, \tag{3.7}$$

$$\mathcal{O}_I(t) = e^{+i\int \hat{H}_0 dt} \mathcal{O}_S e^{-i\int \hat{H}_0 dt} \,, \tag{3.8}$$

The interaction picture is related to the Heisenberg picture by introducing the interaction picture evolution operator $U_I(t, t_i)$ between some initial time t_i and some time t:

$$|\psi, t\rangle_I = U_I(t, t_i)\, |\psi\rangle_H \ , \tag{3.9}$$

$$\mathcal{O}_I(t) = U_I(t, t_i)\mathcal{O}_H(t) U_I^\dagger(t, t_i) \,. \tag{3.10}$$

From the Heisenberg equation for $\mathcal{O}_H$ or the Schrödinger equation for $|\psi\rangle_S$, one finds that the evolution operator in the interaction picture U_I obeys

$$\frac{d}{dt_2} U_I(t_2, t_1) = -i\hat{H}_{int}(t_2) U_I(t_2, t_1) \,, \tag{3.11}$$

$$\frac{d}{dt_1} U_I(t_2, t_1) = i U_I(t_2, t_1)\hat{H}_{int}(t_1) \,, \tag{3.12}$$

as well as $U_I(t, t) = 1$. For $t_2 \geq t_1$, the solution of these equation is concisely given by Dyson's formula

$$U_I(t_2, t_1) = T \exp\left(-i\int_{t_1}^{t_2} dt' \hat{H}_{int}(t')\right) \ , \tag{3.13}$$

$$U_I^\dagger(t_2, t_1) = \bar{T} \exp\left(i\int_{t_1}^{t_2} dt' \hat{H}_{int}(t')\right) \ , \tag{3.14}$$

where the (anti) time-ordered operator ($\bar{T}$) T arranges the operators from left to right in order of (increasing) decreasing time. In the perturbative expansion, this takes the form

$$U(t_2, t_1) = 1 - i \int_{t_1}^{t_2} dt' \hat{H}_{int}(t') - \int_{t_1}^{t_2} dt' \int_{t_1}^{t'} dt'' \hat{H}_{int}(t') \hat{H}_{int}(t'') + \dots . \tag{3.15}$$

When the arguments are not in the right order, the solution of (3.11) and (3.12) is instead given by (using the shorthand $U_{21} = U_I(t_2, t_1)$, etc.)

$$U_{12} \equiv U_{21}^\dagger = U_{21}^{-1} \qquad\qquad U_{12}^\dagger \equiv U_{21} \,, \tag{3.16}$$

in such a way that

$$U_{12}U_{21} = U_{21}U_{12} = 1 \,, \qquad\qquad U_{32}U_{21} = U_{31} \,, \tag{3.17}$$

for any ordering of $t_{1,2,3}$. Dyson's formula then gives us the useful representation

$$S = U_I(\infty, -\infty) = T \exp \left[-i \int_{-\infty}^{+\infty} dt' \hat{H}_{int}(t') \right] . \tag{3.18}$$

Notice that $\hat{H}_{int}$ and hence U_I in the interaction picture are written in terms of *free fields* (2.16).

It is worth discussion a piece of jargon that is often used in the literature.

- **in-out:** A matrix element $\langle\phi| U |\chi\rangle$, which contains a *single* time-evolution operator U, is called an *in-out* object, because it has the interpretation of taking some initial "in" state $|\chi\rangle$ in the past, evolving it for some time and then projecting it onto some final "out" state $\langle\phi|$ in the future. The S-matrix is a typical in-out object. Another classical example is the wavefunction itself. If we also insert the product of local fields, as for example in $\langle\phi| U \prod_i \mathcal{O}_i(\mathbf{x}_i) |\chi\rangle$ this is usually called an "in-out" correlator. Indeed the Lehmann-Symanzik-Zimmermann (LSZ) reduction formula projects in-out correlators onto S-matrix elements. In-out quantities should be thought of as probability amplitudes: they are complex-valued functions and are not per se observables, just like the wavefunction in quantum mechanics.
- **in-in:** A matrix element $\langle\phi| U^\dagger \mathcal{O}_i(\mathbf{x}_i) U |\phi\rangle$, which contains *two* time-evolution operators moving in opposite directions, is usually called an *in-in* object, because it has the interpretation of taking some initial "in" state in the past, evolving it to future and then evolving it back to the same initial state. An example are cross sections in particle physics. Another typical example are the expectation values we learn about in quantum mechanics, e.g. $\langle x^n\rangle$ or $\langle x^n p^m\rangle$. In-in correlators of Hermitian operators are real-valued functions and are legitimate quantum mechanical observables.

These two types of objects are related by the schematic equation

$$\text{in-in} \sim \sum_{\text{out}} |\text{in-out}|^2 \,, \tag{3.19}$$

In-out objects are somewhat more primitive and usually display a richer mathematical structure. Once they are understood, they can be used to compute a desired observable.

3.2 Cosmology and correlators

The situation in cosmology is different from that in particle physics in three major respects:

- **Broken Poincaré symmetry:** As already mentioned, the FLRW background on which cosmological perturbations propagate has four isometries less than the maximally-symmetric Minkowski spacetime. In particular, Lorentz boosts and time translations are spontaneously broken.

- **Field vs particles** The natural observables in cosmology are spatial averages of products of the distribution of stuff in the universe, as e.g. in (0.1). This means that we don't measure individual particles like in a scattering experiments (or the hadronic jets they produce), but we measure classical fields. In QFT, when a field takes a large classical value, it can be thought in principles as a coherent "condensate" of very many particles, where quantum effects are suppressed by the large occupation numbers. To describe these configurations fields are a much more convenient language than particles. As a result, in cosmology we do not want to project onto states with a definite number of particles as we do for the S-matrix, but onto eigenstates $|\phi\rangle$ of the field operators, $\hat{\phi}(\mathbf{x})\,|\phi\rangle = \phi(\mathbf{x})\,|\phi\rangle$.

- **In-in vs in-out:** At early times, cosmological perturbations were effectively in flat space and we can define an initial state, such as for example the Bunch-Davies state in the discussion of Sec. 2. The natural out states to project onto at late times are $\{|\phi\rangle\}$. The resulting in-out object is the field theoretic wavefunction

$$\Psi[\phi] = \langle\phi|\Omega\rangle \ , \tag{3.20}$$

  which is a functional of the fields. The wavefunction approach to cosmology has attracted a lot of attention recently and many new results and techniques have become available. However, due to the lack of time, we will not discuss it here. Instead, we will be interested directly in the cosmological observables, which are "in-in" expectation values.

- **Cosmic variance:** In an expanding universe with a finite age, causality imposes that there is only a finite volume that we can access observationally. If the expansion decelerates, $\ddot{a} < 0$, we can wait long enough and observe any other spacetime point. Instead, the expansion of our universe is currently accelerating $\ddot{a}/a \sim (10^{17}\text{sec})^{-2}$ (pretty slowly). If this acceleration continues in the future, the largest spatial volume we can ever observe is of order the Hubble volume today, $H_0^{-3} \sim (4\text{Gpc})^3$. Hence we cannot observe fluctuations in the whole universe and so our measurements have an intrinsic sample variance, known in this context as *cosmic variance*. This is in stark contrast for example with particle physics, where in principle we can repeat an experiment over and over again to beat down statistical fluctuations.

We are ready to define our main object of interest: an *in-in correlator* is the expectation value of some operator $\mathcal{O}$ on some state Ω to be discussed shortly,

$$\langle\mathcal{O}\rangle = \langle\Omega|\,\mathcal{O}\,|\Omega\rangle \ . \tag{3.21}$$

Here $\mathcal{O}$ will always be the *equal-time* product of local operators at different space points, $\mathcal{O} \sim \prod_i \phi_i(\mathbf{x}_i)$. The time ordering of these operators is therefore irrelevant. As familiar from quantum mechanics, correlators of Hermitian operators are observable and must be real (unlike scattering amplitudes)

$$\langle\Omega|\,\mathcal{O}\,|\Omega\rangle^* = \langle\Omega|\,\mathcal{O}^\dagger\,|\Omega\rangle = \langle\Omega|\,\mathcal{O}\,|\Omega\rangle \in \mathbb{R}\,. \tag{3.22}$$

For formal manipulations, the Heisenberg picture is very convenient. But explicit calculations are most easily performed in the *interaction picture*. We already know from (4.6) that

$$\mathcal{O}_H(t) = U_I^\dagger(\tau,-\infty)\mathcal{O}_I(\tau)U_I(\tau,-\infty)\,, \tag{3.23}$$

where all the fields on the right-hand side are in the interaction picture, i.e. they are just the free fields we studied in Sec. 2.

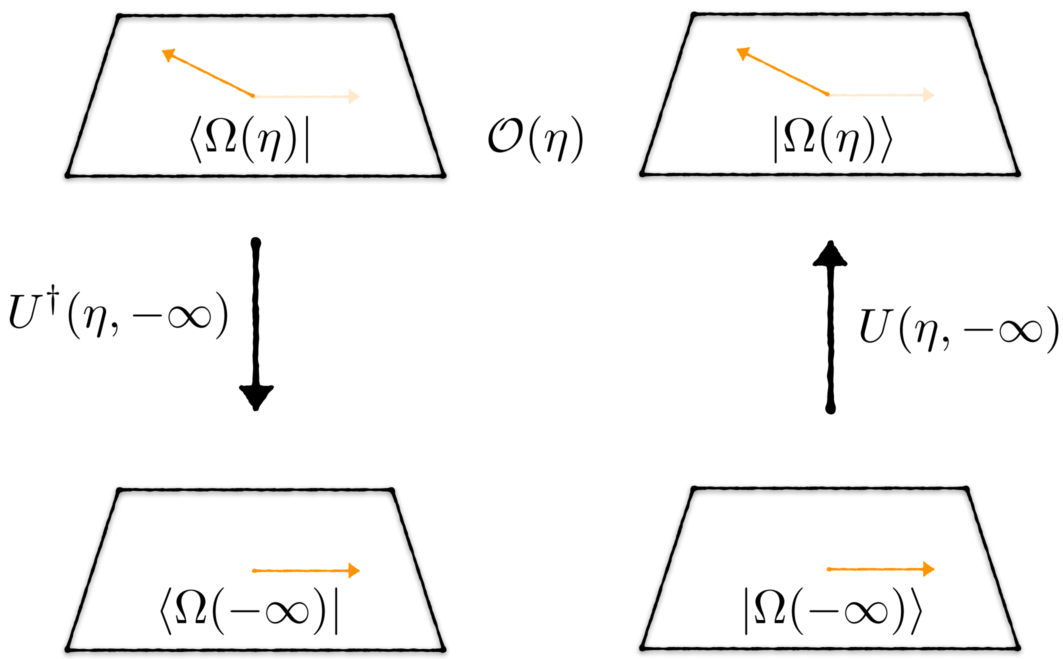


Figure 4: The figure provides a graphical representation of cosmological correlators. Moving from left to right in (3.33) we start from a state in the infinite past (usually the Bunch-Davies state) and evolve until some time η with a unitary operator U that depends on the model. We then insert an operator $\mathcal{O}(\eta)$ and then evolve back to the infinite past with the same U. The resulting object is an observable in-in correlator, which features two time-evolution operators and two separate and opposite time orderings.

The interacting "vacuum" The last thing we need is to define $|\Omega\rangle$. We will only be interested in the case in which $|\Omega\rangle$ is the "vacuum" of the interacting theory, which in the far past asymptotes the free theory vacuum $|0\rangle$, defined by $a_{\mathbf{k}}|0\rangle = 0$ in Sec. 2:

$$\lim_{\tau\to-\infty} |\Omega\rangle = |0\rangle \ . \tag{3.24}$$

For adiabatic evolution energy levels never cross, so $|\Omega\rangle$ must be the lowest energy state of the full theory, just as $|0\rangle$ is the lowest energy level of the free theory. Also, $|\Omega\rangle$ must minimize both $\hat{H}_0$ and $\hat{H}_{int}$ separately. We can then relate $|\Omega\rangle$ to $|0\rangle$ by the following heuristic argument. Let us expand $|\Omega\rangle$ in terms of energy eigenstates $|n\rangle$ of the interaction Hamiltonian $\hat{H}_{int}$

$$\begin{align} e^{-i\hat{H}_{int}(\tau-\tau_i)}|\Omega\rangle &= \sum_n e^{-i\hat{H}_{int}(\tau-\tau_i)}|n\rangle\langle n|\Omega\rangle \tag{3.25}\\ &= e^{-iE_0(\tau-\tau_i)}|0\rangle\langle 0|\Omega\rangle + \sum_{n\neq 0} e^{-iE_n(\tau-\tau_i)}|n\rangle\langle n|\Omega\rangle \ . \tag{3.26}\end{align}$$

For $\tau_i \to -\infty$, we want $|\Omega\rangle \to |0\rangle$ and so all the terms in the sum over $n \neq 0$ must drop out. To achieve this, we choose to add to τ a small and negative imaginary part

$$\tau \to \tau(1 - i\epsilon)\,, \tag{3.27}$$

where $0 < \epsilon \ll 1$ is some real number (not the homonimous slow-roll parameter). The expression $e^{-iE_n(\tau-\tau_i)}$ then acquires a factor $e^{-\epsilon E_n(\tau-\tau_i)}$. All states with energy larger than E_0 are then exponentially suppressed in the limit $\tau_i \to -\infty$ and we recover the result. A similar argument applies to the conjugate of this expression relating $\langle\Omega|$ to $\langle 0|$. In this case one finds that the opposite sign is needed for the $i\epsilon$ shift to project onto the vacuum of the free theory. From now on we will therefore always assume that the time integral in the evolution operator has been slightly rotated

$$\begin{align} U_I(\tau,-\infty) &= T\exp\left(-i\int_{-\infty(1-i\epsilon)}^{\tau} d\tau'\hat{H}_{int}(\tau')\right)\,, \tag{3.28}\\ U_I^\dagger(\tau,-\infty) &= \bar{T}\exp\left(i\int_{-\infty(1+i\epsilon)}^{\tau} d\tau'\hat{H}_{int}(\tau')\right)\,, \tag{3.29}\end{align}$$

so that we can write

$$\begin{align} \langle\mathcal{O}\rangle &= \langle\Omega|\, U_I^\dagger(\tau,-\infty)\mathcal{O}_I(\tau)U_I(\tau,-\infty)\,|\Omega\rangle \tag{3.30}\\ &= \langle 0|\, U_I^\dagger(\tau,-\infty)\mathcal{O}_I(\tau)U_I(\tau,-\infty)\,|0\rangle\ |\langle 0|\Omega\rangle|^2\,. \tag{3.31}\end{align}$$

But taking the expectation value of the unit operator $\mathcal{O} = \mathbb{I}$, we find that

$$|\langle 0|\Omega\rangle|^2 = \frac{\langle\Omega|\Omega\rangle}{\langle 0|0\rangle} = 1\,. \tag{3.32}$$

We come therefore to our final formula for correlators

$$\langle\mathcal{O}(\tau)\rangle = \langle 0|\left[\bar{T}e^{\left(i\int_{-\infty(1+i\epsilon)}^{\tau} d\tau'\hat{H}_{int}(\tau')\right)}\right]\mathcal{O}(\tau)\left[Te^{\left(-i\int_{-\infty(1-i\epsilon)}^{\tau} d\tau'\hat{H}_{int}(\tau')\right)}\right]|0\rangle\ , \tag{3.33}$$

where all fields appearing in $\mathcal{O}(\tau)$ and $\hat{H}_{int}$ are the free fields we introduced in Sec. 2. A cartoon of cosmological correlators is shown in Fig. 4.

In perturbation theory we can expand the exponentials as we did in (3.15) and find schematically

$$\langle \mathcal{O}(\tau) \rangle = \langle \bar{T} \left[1 + i \int H - \frac{1}{2} \left(\int H \right)^2 \right] \mathcal{O}\, T \left[1 - i \int H - \frac{1}{2} \left(\int H \right)^2 \right] \rangle \tag{3.34}$$

$$= \langle \mathcal{O} + i \left(\int H\mathcal{O} - \mathcal{O} \int H \right) - \frac{1}{2} \left[\left(\int H \right)^2 \mathcal{O} + \mathcal{O} \left(\int H \right)^2 \right] \tag{3.35}$$

$$+ \left(\int H \right) \mathcal{O} \left(\int H \right) + \ldots \rangle \,. \tag{3.36}$$

There is an equivalent version of this formula that is sometimes useful when performing perturbative calculations [102]:

$$\langle \mathcal{O}(\tau) \rangle = \sum_{N=0}^{\infty} i^N \int_{-\infty}^{\tau} d\tau_N \int_{-\infty}^{\tau_N} d\tau_{N-1} \cdots \int_{-\infty}^{\tau_2} d\tau_1 \tag{3.37}$$
$$\times \langle 0 | \, [\hat{H}_{int}(\tau_1), [\hat{H}_{int}(\tau_2), \ldots [\hat{H}_{int}(\tau_N), \mathcal{O}(\tau)] \ldots]] \, |0\rangle \,.$$

Sometimes people refer to (3.33) as the *factorized form* and to (3.37) as the *commutator form*. To prove that (3.33) and (3.37) are indeed equivalent, we proceed by induction. To zeroth and first order in $\hat{H}_{int}$ they obviously agree. Starting from (3.33) we expand

$$\langle \mathcal{O}(\tau) \rangle_{0^{\text{th}}} = \langle 0 | \, \mathcal{O}(\tau) \, | 0 \rangle \,, \tag{3.38}$$

$$\langle \mathcal{O}(\tau) \rangle_{1^{\text{st}}} = \langle 0 | \left[i \int \hat{H}_{int}(\tau') d\tau' \right] \mathcal{O}(\tau) + \mathcal{O}(\tau) \left[-i \int \hat{H}_{int}(\tau') d\tau' \right] |0\rangle \tag{3.39}$$

$$= i \int_{-\infty}^{\tau} \langle 0 | \, [\hat{H}_{int}(\tau'), \mathcal{O}(\tau)] \, |0\rangle \,. \tag{3.40}$$

Now assume (3.33) and (3.37) give the same result up to order $(N-1)$. Then take the time derivative of each expression at order N. They can be re-written as the expectation value of some other operator at order $(N-1)$ and so they must agree up to a constant. Since they both give the same result for $\tau \to -\infty$, at arbitrary order, the constant must be zero. You will go through the details in Example Sheet 1. Notice that terms coming from U and $U^\dagger$ combine to form the commutation in (3.37). So one has to be careful in keeping track of the correct $i\epsilon$ prescription to project Ω onto $|0\rangle$.
As an aside, there are two other formalisms to compute correlators that are useful in different applications. One is the path integral or Schwinger-Keldysh formalism, in which the correlator is expressed as a path integral from some initial time to the time at which the operators are evaluated and back to the initial time (see e.g. [26]). The second is the Schrödinger picture of quantum mechanics, where the wave function is a functional of the fields and it is often referred to as the *wave function* of the universe (see e.g. [5, 51, 67]).

3.3 Feynman rules for correlators

Just like for amplitudes, the easiest way to perform calculations is with the aid of diagrams. The diagrammatic rules to compute correlators, sometimes called Feynman-Witten rules because of their resemblance to the calculation of AdS correlators [105], can be derived either in the canonical or in the path integral formalism (see e.g. [26]).

Because of translation invariance all correlators must contain one or more delta functions of momentum conservation. Correlators that are proportional to two or more delta functions are called *disconnected* correlators. Given any disconnected correlator, one can always subtract from it all possible products of lower-point functions. This process leaves only contributions proportional to a single delta function, which are called *connected* correlators and are usually the main object of interest. For connected correlators we will introduce the notation

$$\left\langle \prod_{a=1}^{n} \phi(\mathbf{k}_a) \right\rangle_c \equiv (2\pi)^3 \delta\left(\sum_{a=1}^{n} \mathbf{k}_a \right) B_n(\{\mathbf{k}\}) \ . \tag{3.41}$$

We will usually quote results for $B_n(\{\mathbf{k}\}) = B_n(\mathbf{k}_1, \ldots, \mathbf{k}_n)$.

The rules to compute momentum-space in-in correlators B_n are as follows:

- For an n-point correlator draw a diagram consisting of V vertices, I internal lines, each connecting two vertices, and n external lines, each connecting a vertex to the future conformal boundary of dS represented by a horizontal line at the top of the diagram, as in Fig. 5 and 6. Time runs vertically from bottom at $\eta \to -\infty$ to top at $\eta \to 0$.

- Each vertex can be either (i) a "right" vertex, which is labelled with an "r" and represents an interaction from U_I in (3.33) which results in an H_{int} to the right of the operator as in $\langle \mathcal{O}(\mathbf{k}) H_{int} \rangle$, or (ii) a left vertex, which is labelled by "l" and represents an interaction from $U_I^\dagger$ resulting in a Hamiltonian to the left of the operator as in $\langle H_{int} \mathcal{O}(\mathbf{k}) \rangle$. One should sum over the 2^V ways to label the V vertices.

- To each of the n external lines associate a spatial momentum $\mathbf{k}_a$ with $a = 1, \ldots, n$ with a specific direction of flow (see e.g. Fig. 5). Because of invariance under spatial translations, there should be a delta function of the sum of all spatial momenta ending on each vertex, and all internal momenta $\mathbf{p}_m$ with $m = 1, \ldots, I$ should be integrated over. Using the topological graph identity $I - V + 1 = L$, there are L integrals over internal momenta d^3p that remain to be performed as expected for a diagram with L loops (the $+1$ accounts for the overall delta function $\delta(\sum_a^n \mathbf{k}_a)$, which does not depend on internal momenta and is dropped in B_n, see (3.41)). Notice that there is no "energy" conservation, so the norms of momenta ending on a vertex do not add up to zero in general.

- Internal lines are associated to a Bulk-to-Bulk (B2B) propagator. Depending on the type of vertices that an internal line joins, there are four possible B2B propagators:

$$\bullet - \bullet = G_{rr}(\eta_1, \eta_2, p) = \langle 0 | T\phi(\eta_1, \mathbf{p})\phi(\eta_2, \mathbf{p}') | 0 \rangle' \tag{3.42}$$

$$= f_p(\eta_1) f_p^*(\eta_2) \theta(\eta_1 - \eta_2) + f_p^*(\eta_1) f_p(\eta_2) \theta(\eta_2 - \eta_1) \tag{3.43}$$

$$\circ - \bullet = G_{lr}(\eta_1, \eta_2, p) = \langle 0 | \phi(\eta_1, \mathbf{p})\phi(\eta_2, \mathbf{p}') | 0 \rangle' = f_p(\eta_1) f_p^*(\eta_2) \tag{3.44}$$

$$\bullet - \circ = G_{rl}(\eta_1, \eta_2, p) = \langle 0 | \phi(\eta_2, \mathbf{p}')\phi(\eta_1, \mathbf{p}) | 0 \rangle' = G_{lr}^*(\eta_1, \eta_2, p) \tag{3.45}$$

$$\circ - \circ = G_{ll}(\eta_1, \eta_2, p) = \langle 0 | \bar{T}\phi(\eta_1, \mathbf{p})\phi(\eta_2, \mathbf{p}') | 0 \rangle' = G_{rr}^*(\eta_1, \eta_2, p) \tag{3.46}$$

$$= f_p^*(\eta_1) f_p(\eta_2) \theta(\eta_1 - \eta_2) + f_p(\eta_1) f_p^*(\eta_2) \theta(\eta_2 - \eta_1) \, , \tag{3.47}$$

  where η_1 and η_2 are the times associated to the vertices joined by G (see below), and T and $\bar{T}$ indicate time- and anti-time-ordering. Notice that G_{rr} and G_{ll} are symmetric in their time variables, $G_{rr,ll}(\eta_1, \eta_2) = G_{rr,ll}(\eta_2, \eta_1)$. The G_{rr} and G_{ll} propagators are just the familiar Feynman propagator we encounter in amplitudes. The theta functions enforce the time ordering appearing in U_I and the anti-time ordering in $U_I^\dagger$. No theta functions appear in G_{rl} and G_{lr} because interactions in U_I are not time-ordered with respect to interactions in $U_I^\dagger$ (this follows from the doubling of fields in the path integral formulation).

- External lines going to the boundary at $\eta = \eta_0 \to 0$ are associated to Bulk-to-boundary (B2b) propagators. There are two B2b propagators

$$\bullet\!- = G_r(\eta, p) = f_p(\eta_0) f_p^*(\eta) \,, \qquad \circ\!- = G_l(\eta, p) = f_p^*(\eta_0) f_p(\eta) \,. \tag{3.48}$$

  Notice that $G_r(\eta, p) = G_l^*(\eta, p)$ and that a B2b propagator is just a B2B propagator where one vertex has been pushed to the boundary, $G_r(\eta, p) = G_{rr}(\eta, \eta_0, p) = G_{rl}(\eta, \eta_0, p)$.

- Vertices are associated to a time η_A for $A = 1, \ldots, V$ and to an integral $d\eta$ with a factor $\sqrt{-g} = (\eta_V H)^{-4}$. The boundaries of integration are

$$\text{right vertex:} \quad -\infty(1 - i\epsilon) < \eta \leq \eta_0 \,, \tag{3.49}$$

$$\text{left vertex:} \quad -\infty(1 + i\epsilon) < \eta \leq \eta_0 \,. \tag{3.50}$$

  When the $\eta_0 \to 0$ limit is finite, such as for example for massless scalars with enough derivative interactions, I will simply write $\eta_0 = 0$. Furthermore, every vertex gets a vertex factor that depends on the theory. For simple polynomial interactions such as $H_{int} \supset \lambda\phi^n$, the vertex is simply $-i\lambda$ on every right vertex and a $+i\lambda$ on every left vertex. Every spatial derivative in a Hamiltonian interaction should act on all lines ending on the vertex giving a factor $\partial_x \to (-i\mathbf{k})$ (it would be a $+i\mathbf{p}$ in the Fourier space Hamiltonian, which then gets integrated over $\delta(\mathbf{k}+\mathbf{p})$). For example, for a right vertex we have

$$H_{\text{int}} \supset +\lambda\phi(\partial_i\phi g^{ij}\partial_j\phi) \quad \Rightarrow \quad -i\lambda \int \frac{d\eta}{(\eta H)^4} (H\eta)^2 \left[(-i)^2(\mathbf{k}_1\mathbf{k}_2) + 2 \text{ perm's}\right] . \tag{3.51}$$

  Similarly, time derivatives in the Hamiltonian interaction are accounted for by time derivatives acting on all the (B2b or B2B) propagators ending on the associated vertex. For example for a contact diagram with a right interaction we have

$$H_{\text{int}} \supset +\lambda\phi\dot{\phi}^2 \quad \Rightarrow \quad -i\lambda \int \frac{d\eta}{(\eta H)^4} (H\eta)^2 G_r(\eta, k_1)\partial_\eta G_r(\eta, k_2)\partial_\eta G_r(\eta, k_3) + 2 \text{ perm's} \,,$$

  where $\dot{} = \partial_t = (-H\eta)\partial_\eta$.

- The combinatorial factor is computed as follows [48]:

  1. Begin by considering only a single labelling of external legs, for example from $\mathbf{k}_1$ to $\mathbf{k}_n$ from left to right. Do not include two diagrams that are related to each other by a relabelling of the external legs. Do not include two diagrams that are related by swapping two identical interaction vertices (these permutations cancel the $1/m!$ from expanding the exponential in the time-evolution operator and are hence already taken into account).

2. Vertex combinatorial factor: For each each vertex with n lines corresponding to the same field multiply by $(n!)$. If there are n_1 lines of field 1 and n_2 lines of field 2 and so on, multiply by $(n_1!) \times (n_2!) \times \ldots$.
3. Diagram combinatorial factor[22]: When two vertices are connected by k bulk-bulk propagators corresponding to the same field, divide by a factor of $k!$. The case when both ends of a bulk-bulk propagator are attached to the same vertex can be excluded simply by normal ordering the interaction Hamiltonian, so that there are never contractions of two fields from the same vertex. If, for some reason, one insists in keeping these same-vertex loops, then the combinatorial factor

$$\frac{(2L^s - 1)!!}{(2L^s)!} = \frac{1}{(2L^s)!!} \tag{3.52}$$

for each vertex where L^s is the number of loops staring and ending on that vertex. The $(2L^s - 1)!!$ factor comes from all the possible ways of pairwise contracting the $2L^s$ lines emanating from the vertex, while the $1/(2L^s)!$ factor removes the wrong factors introduced in step 2 above.
4. Sum over channels: the result obtained so far corresponds to a single specific labellings of external legs. Add to this result all possible other labelling obtained by permuting the set of external momenta $\{\mathbf{k}_a\}$. Some of these permutations might be identical to the result we had for the initial labelling because that was already symmetric under the permutation, such as for the example the permutation $\mathbf{k}_1 \leftrightarrow \mathbf{k}_2$ in an s-channel diagram ($12 \to 34$). Other permutations lead to a different result, which is interpreted as a different channel, such as the permutation $\mathbf{k}_1 \leftrightarrow \mathbf{k}_3$ in an s-channel diagram which produces a t-channel contribution.

These rules apply also to in-in correlators in Minkowski evaluated at $t = 0$. To this end it suffices using the Minkowski mode functions for the f's in the propagator and the the substitutions $\eta \to t$ and $\sqrt{-g} = 1$.

We conclude noticing a useful property that relates diagrams with oppositely labelled vertices. Let D represent all functions that the above Fourier-space rules attribute to some diagram with V vertices and let σ_a with $j = 1, \ldots, 2^V$ be all possible ordered lists of choices of right/left vertices. Let $\bar{\sigma}_a$ represent the ordered list of opposite choices to σ_a, where all vertices have been flipped right $\leftrightarrow$ left. For example, for a two vertex diagram we have

$$\text{Example } V = 2: \quad \sigma_1 = \{rr\}\,, \quad \sigma_2 = \{rl\}\,, \quad \sigma_3 = \{lr\} = \bar{\sigma}_2\,, \quad \sigma_4 = \{ll\} = \bar{\sigma}_1\,. \tag{3.53}$$

Then[23]

$$D[\sigma] = D[\bar{\sigma}]^* (-)^{n_i}\,, \tag{3.54}$$

where n_i is the total number of spatial derivatives appearing in all the vertices of the

[22] In [48], the possibility of more than one same-vertex loop on the same diagram was not accounted for.

[23] The extra factor $(-)^{n_i}$ appears because a derivative interaction brings a factor $-i\mathbf{k}$ for both right and left vertices.

diagram. Hence

$$B_n = \sum_a D[\sigma_a] = \frac{1}{2} \sum_a [D[\sigma_a] + D[\bar{\sigma}_a]] \tag{3.55}$$

$$= \frac{1}{2} \sum_a [D[\sigma_a] + D[\bar{\bar{\sigma}}_a]^* (-)^{n_1}] \tag{3.56}$$

$$= \frac{1}{2} \sum_a [D[\sigma_a] + D[\sigma_a]^* (-)^{n_1}] \, . \tag{3.57}$$

This relation means that in practice we just need to do half of the work: it suffices to compute only 2^{V-1} labelings of the vertices of a diagram and the remaining 2^{V-1} are related to them by complex conjugation. In these notes we will restrict to parity-even interactions, but see e.g. [] for in depth discussions of the parity-odd case

3.4 Examples

Let's demonstrate the diagrammatic rules in a series of examples.

Polynomial interactions in Minkowski In Minkowski, the propagators of a massive canonical scalar field are

$$G_r(t,k) = \frac{e^{iEt}}{2E} \, , \qquad G_{rl}(t_1,t_2,p) = \frac{e^{iE(t_1-t_2)}}{2E} \, , \tag{3.58}$$

$$G_{rr}(t_1,t_2,p) = \frac{e^{iE(t_2-t_1)}}{2E} \theta(t_1-t_2) + (t_1 \leftrightarrow t_2) \, . \tag{3.59}$$

where $E = +\sqrt{\mathbf{p}^2 + m^2}$. For polynomial interactions $H_{\text{int}} = \lambda \phi^n/(n!)$ to leading order in λ we have the contact diagrams in Fig. 5 panel "a". The

$$B_n^{(r)} = -i \frac{\lambda}{n!} \times n! \times \int_{-\infty(1-i\epsilon)}^{0} dt \, \frac{e^{iE_T t}}{2^n \prod^n E_a} \tag{3.60}$$

$$= -i \frac{\lambda}{2^n \prod^n E_a} \left[\frac{e^{iE_T t}}{iE_T} \right]_{-\infty(1-i\epsilon)}^{0} = -\frac{\lambda}{2^n E_T \prod^n E_a} \, , \tag{3.61}$$

where we have introduce the total energy

$$E_T \equiv \sum_{a=1}^{n} E_a \, . \tag{3.62}$$

Notice that the $i\epsilon$ rotation in the lower boundary of the time integral is exactly the correct one that guarantees convergence at past infinity. The final correlator is then

$$B_n = B_n^{(r)} + B_n^{(l)} = 2 \, \mathrm{Re} \, B_n^{(r)} = -\frac{2\lambda}{E_T \prod_{a=1}^{n} (2E_a)} \, . \tag{3.63}$$

This result has precisely the right mass dimensions (working with $\hbar = c = 1$), as can be checked from

$$[\phi(\mathbf{x})] = M^1 \, , \qquad [\phi(\mathbf{k})] = M^{-2} \, , \qquad [\lambda] = M^{4-n} \, . \tag{3.64}$$

Notice also that the correlator is real, as it should be for parity-even correlators.

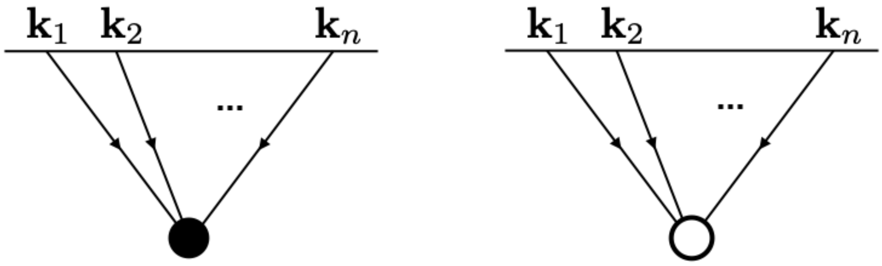


Figure 5: The Feynman-Witten diagram of a contact in-in correlator.

Derivative interactions in Minkowski As an example with derivative interactions consider

$$H_{\text{int}} = \lambda(\partial_i\phi\partial_i\phi)\dot{\phi}^2\phi\,. \tag{3.65}$$

To linear order in λ we have again a single diagram with a right vertex (plus the related corresponding diagram with a left vertex)

$$B_5^{(r)} = -i\lambda(-i\mathbf{k}_1)\cdot(-i\mathbf{k}_2)\int dt\, G_r(t,k_1)G_r(t,k_2)\dot{G}_r(t,k_3)\dot{G}_r(t,k_4)G_r(t,k_5) + \text{perm's} \tag{3.66}$$

$$= i\frac{\lambda(\mathbf{k}_1\cdot\mathbf{k}_2)}{\prod_{a=1}^5 2E_a}\int dt e^{iE_T}(iE_3)(iE_4) + \text{perm's} \tag{3.67}$$

$$= -\frac{\lambda(\mathbf{k}_1\cdot\mathbf{k}_2)E_3E_4}{\prod_{a=1}^5 2E_a}\frac{1}{E_T} + \text{perm's}\,, \tag{3.68}$$

where "perm's" refers to the other $(5!-1)$ permutations of the five momenta. In Minkowski time derivatives just amount to factors of iE_a, similarly to the Feynman rules for amplitudes.

Exchange diagram in Minkowski Next let's compute the four-point function of a scalar in Minkowski from a cubic polynomial interaction to second order

$$H_{\text{int}} = \frac{\lambda}{3!}\phi^3\,. \tag{3.69}$$

Just like for amplitudes, there are three exchange diagrams corresponding to the so-called s ($12\to 34$), t ($13\to 24$) and u ($14\to 23$) channels. It's easiest compute just the s-channel and then sum over permutations. There are two independent diagrams, which are depicted in Fig. 6. The first has two right vertices:

$$B_{4,s}^{(rr)} = (-i\lambda)^2\int dt_1dt_2 G_r(t_1,k_1)G_r(t_1,k_2)G_{rr}(t_1,t_2,E_s)G_r(t_2,k_3)G_r(t_2,k_4) \tag{3.70}$$

$$= -\lambda^2\int dt_1dt_2\frac{e^{i(E_1+E_2)t_1}e^{i(E_3+E_4)t_2}}{\prod^4(2E_a)}\frac{1}{2E_s}\left[e^{iE_s(t_2-t_1)}\theta(t_1-t_2) + (t_1\leftrightarrow t_2)\right] \tag{3.71}$$

$$= -\frac{\lambda^2}{2E_s\prod^4(2E_a)}\int dt_1\, e^{i(E_{12}-E_s)t_1}\frac{e^{i(E_{34}+E_s)t_1}}{i(E_{34}+E_s)} + (t_1\leftrightarrow t_2) \tag{3.72}$$

$$= \frac{\lambda^2}{2E_sE_T\prod^4(2E_a)}\left(\frac{1}{E_{12}+E_s}+\frac{1}{E_{34}+E_s}\right), \tag{3.73}$$

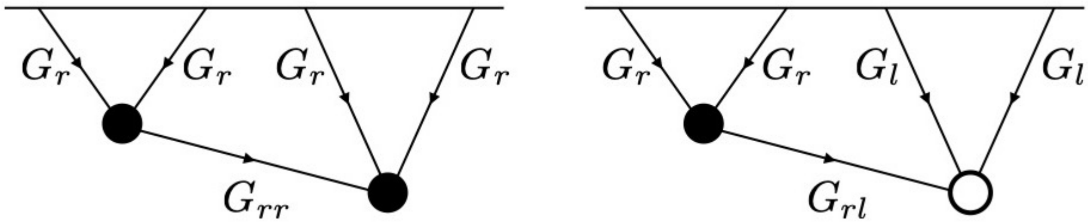
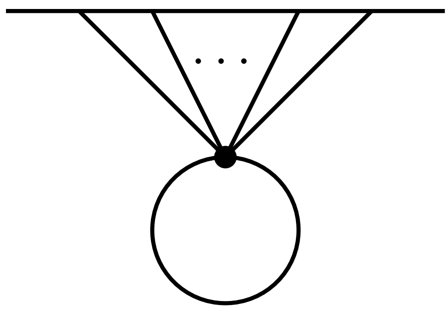


Figure 6: The Feynman-Witten diagrams for the exchange contribution to four-point in-in correlators.



where

$$E_{12} \equiv E_1 + E_2 \,, \qquad\qquad E_s = \sqrt{|\mathbf{k}_1 + \mathbf{k}_2|^2 + m^2} \,. \tag{3.74}$$

The second has one right and one left vertex:

$$B_{4,s}^{(rl)} = (-i\lambda)(i\lambda) \int dt_1 dt_2 G_r(t_1, k_1) G_r(t_1, k_2) G_{rl}(t_1, t_2, s) G_l(t_2, k_3) G_l(t_2, k_4) \tag{3.75}$$

$$= \frac{\lambda^2}{\prod^4 (2E_a)} \int dt_1 dt_2 \, e^{iE_{12}t_1} \frac{e^{iE_s(t_1-t_2)}}{2E_s} e^{-iE_{34}t_2} \tag{3.76}$$

$$= \frac{\lambda^2}{2E_s E_R E_L \prod^4 (2E_a)} \,, \tag{3.77}$$

where I introduced the left and right *partial energies* $E_{L,R}$ defined as

$$E_L \equiv E_1 + E_2 + E_s \,, \qquad\qquad E_R \equiv E_3 + E_4 + E_s \,. \tag{3.78}$$

Putting everything together

$$B_{4,s} = B_{4,s}^{(rr)} + B_{4,s}^{(rl)} + B_{4,s}^{(lr)} + B_{4,s}^{(ll)} = 2 \,\mathrm{Re} \left[B_{4,s}^{(rr)} + B_{4,s}^{(rl)} \right] \tag{3.79}$$

$$= \frac{\lambda^2}{E_s \prod^4 (2E_a)} \frac{2(E_T + E_s)}{E_L E_R E_T} \,. \tag{3.80}$$

One loop in Minkowski Let's compute now a one-loop contribution in Minkowski. The simplest case is that of a single interaction vertex, as depicted in Figure 3.4 (these "one-loop one-vertex" diagrams were discussed extensively in [63]). For simplicity we focus on the following polynomial interaction

$$\mathcal{L}_{\text{int}} = \int_{\mathbf{x}} \frac{\lambda}{6!} \phi^6 \,. \tag{3.81}$$

but it is straightforward to add derivatives. According to our Feynman rules we have

$$B_4 = 2\,\mathrm{Re}\left[\frac{i\lambda}{2}\int_{\mathbf{p}}\int_{\infty}^{0} dt\, G_{rr}(t,t,p)\prod_a^4 G_r(t,k_a)\right]. \tag{3.82}$$

An interesting simplification occurs because the two times in the bulk-bulk propagator coincide

$$G_{rr}(t,t,p) = \frac{e^{i\Omega(t-t)}}{2\Omega} = \frac{1}{2\sqrt{p^2+M^2}}. \tag{3.83}$$

The time dependence has disappeared. We can hence directly perform the loop integral over momentum. Let's start with a massless field in the loop, in which case $G_{rr} = 1/(2p)$. In dimensional regularization (dim reg) the integrals of power laws vanish because there is no scale available to write a result with the correct dimensions,

$$\int dp^d\, p^\alpha = 0 \quad \text{(dim reg)}. \tag{3.84}$$

Hence the one-loop one-vertex contribution from a massless particle in Minkowski vanish. This is similar to what happens for amplitudes in diagrams where not external momentum flows through the loop. If the field is massive the result if found to be

$$B_4 = \frac{1}{8E_1E_2E_3E_4}\frac{\lambda m^2}{16\pi^2 k E_T^{(4)}}\left(\frac{1}{\delta} + \log\frac{m}{\mu} + (\text{analytic})\right), \tag{3.85}$$

with μ the renormalization scale.

Contact interactions in dS Let's move on to de Sitter spacetime (dS). The main difference from the Minkowski example is that we have to be careful to include the correct number of scale factors. First, notice that the measure of the time integral for each vertex is

$$\int dt\sqrt{-g} = \int d\eta a^4 = \int \frac{d\eta}{H^4\eta^4}. \tag{3.86}$$

Second, notice that all derivative interaction must be covariant and so come with factors of the inverse background metric, as in $\partial_\mu\phi\partial_\nu\phi g^{\mu\nu}$. We are interested in models that can break dS boosts but preserve dS dilations, translations and rotations. So we can use time and spacial derivatives separately. Every two conformal time derivatives[24] come with a factor of $g^{00} = -1/a^2 = (\eta H)^2$, and every two spacial derivative come with $g^{ij} = \delta_{ij}/a^2 = \delta_{ij}(\eta H)^2$. These factors of η ensure dilation invariance of the final result and the factors of H give the correct mass dimensions.

More concretely, let's consider the contact interaction $H_{\text{int}} = \lambda\dot{\phi}^n/(n!)$. From the above discussion we know we should include a factor of $(\eta H)^n$, and we re-absorb the overall minus sign into the coupling constant λ,

$$\int \frac{d\eta}{H^4\eta^4}\frac{\lambda}{n!}(\phi')^n(\eta H)^n. \tag{3.87}$$

[24]Derivatives with respect to cosmological time, namely $\partial_t = a^{-1}\partial_\eta$, come with $g^{tt} = -1$ and so no factors of η.

The calculation of a contact n-point correlator proceeds straightforwardly

$$B_3 = 2\,\mathrm{Re}\left[-i\lambda \int \frac{d\eta}{(H\eta)^4}(H\eta)^n \prod_{a=1}^{n} G_r'(\eta,k_a)\right] \tag{3.88}$$

$$= 2\lambda\,\mathrm{Re}\left[-iH^{n-4}\int d\eta\,\eta^{n-4}\prod_{a}^{n}\frac{H^2}{2k_a^3}k_a^2\eta e^{ik_a\eta}\right] \tag{3.89}$$

$$= -\frac{2\lambda H^{3n-4}}{\left(\prod^n 2k_a\right)}\,\mathrm{Re}\left[\frac{i^{2n-4-1+1}(2n-4)!}{E_T^{2n-3}}\right] \tag{3.90}$$

$$= (-)^{n-1}\frac{2^{1-n}\lambda H^{3n-4}(2n-4)!}{\left(\prod^n k_a\right)E_T^{2n-3}}\,. \tag{3.91}$$

where we used the master integral

$$\int_{-\infty(1-i\epsilon)}^{0} d\eta e^{+iE_T\eta}\eta^p = \frac{i^{p-1}p!}{E_T^{p+1}} \quad \text{for } p\geq 0\,. \tag{3.92}$$

4 Correlators from $P(X,\phi)$ theories

ref

In the previous Sec. we learned how to compute cosmological correlators and went through two examples in detail. Let us now apply these results to inflation. First, we will need to expand in perturbations the class of $P(X,\phi)$ theories introduced in Sec. 1.4 and then use the in-in formalism to compute correlators of the corresponding interactions.

4.1 $P(X,\phi)$ at quadratic order and the speed of sound

Let us assume we have found some solution to the background equations of motion for some $P(X,\phi)$ theory. We will allow for perturbations around the background solution $\bar\phi(t)$ by writing

$$\phi(\mathbf{x},t) = \bar\phi(t) + \varphi(\mathbf{x},t)\,, \tag{4.1}$$

and treating $\varphi \ll \bar\phi$ perturbatively. Let us expand the Lagrangian in φ:

$$L = P(\bar X + \delta X, \bar\phi + \varphi) \tag{4.2}$$

$$= P + P_{,\phi}\varphi + P_{,X}\delta X + \frac{1}{2}\left[P_{,XX}\delta X^2 + 2P_{,X\phi}\delta X\varphi + P_{,\phi\phi}\varphi^2\right] + \dots\,, \tag{4.3}$$

where P and its derivatives are evaluated on the background $P = P(\bar X,\bar\phi)$ and we defined

$$\delta X = X - \bar X = \dot{\bar\phi}\dot\varphi - \frac{1}{2}\partial_\mu\varphi\partial^\mu\varphi\,. \tag{4.4}$$

We are going to massage this ugly looking Lagrangian into something nice that we studied before, namely (2.2). The coefficient of the terms linear in φ is nothing but the background equations of motion up to a total derivative. Since these are satisfied by assumption, we can focus directly on the quadratic terms

$$L_2 = -\frac{P_{,X}}{2}\partial_\mu\varphi\partial^\mu\varphi + \frac{1}{2}\left[P_{,XX}\left(\dot{\bar\phi}\dot\varphi\right)^2 + 2P_{,X\phi}\dot{\bar\phi}\dot\varphi\varphi + P_{,\phi\phi}\varphi^2\right]\,. \tag{4.5}$$

The term $\dot{\varphi}\varphi$ can be integrated by part in the action into a φ^2 term. Collecting all terms one finds

$$S_2 = \int d^3x dt\, a^3\, \frac{1}{2} \left[\left(P_{,X} + 2P_{,XX}\bar{X} \right) \dot{\varphi}^2 - P_{,X}\partial_i\varphi\partial^i\varphi - m^2\varphi^2 \right] , \tag{4.6}$$

where

$$m^2 = 3HP_{,X\phi}\dot{\bar{\phi}} + \partial_t \left(P_{,X\phi}\dot{\bar{\phi}} \right) - P_{,\phi\phi} , \tag{4.7}$$

and we used $\dot{\bar{\phi}} = 2\bar{X}$. Despite the ugly coefficients, the action (4.6) has the same terms as that for the massive scalar field we studied in Sec. 2.2. As discussed there, in cosmology we are interested in almost massless scalar fields, whose correlation functions survive until late times. So we will assume that the mass term is negligible as compared to the others and drop it henceforth. This step can be justified rigorously. It can be shown that all background quantities involving derivatives with respect to ϕ, e.g. $P_{,X\phi}$ or $P_{,\phi\phi}$, are suppressed by slow-roll parameters. The algebra is long and tedious in general, but it's quite simple if we look at a specific model $P = X - V$. Then, the Friedman equations become

$$3M_{\rm Pl}^2 H^2 = 2XP_{,X} - P = X + V \tag{4.8}$$

$$-M_{\rm Pl}^2\dot{H} = XP_{,X} = X . \tag{4.9}$$

Combining them we find

$$\Rightarrow \quad V = H^2 M_{\rm Pl}^2 (3 - \epsilon) . \tag{4.10}$$

Taking a time derivative on each side and using the chain rule on the left-hand side, $\partial_t = \dot{\phi}\partial_t$ we find

$$V'(\bar{\phi}) = M_{\rm Pl}H^2 \left[-\sqrt{\frac{\varepsilon}{2}}\eta - 3\sqrt{2\varepsilon} + \sqrt{2\varepsilon}\varepsilon \right] , \tag{4.11}$$

$$V''(\bar{\phi}) = H^2 \left[-\frac{3}{2}\eta + \frac{5}{2}\varepsilon\eta - \frac{1}{4}\eta^2 - \frac{1}{2}\frac{\dot{\eta}}{H} - 2\varepsilon^2 + 6\varepsilon \right] , \tag{4.12}$$

$$V'''(\bar{\phi}) = \frac{H^2}{\sqrt{2\varepsilon}M_{\rm Pl}} \left[-\frac{3}{2}\frac{\dot{\eta}}{H} - \frac{\ddot{\eta}}{2H^2} - \frac{\eta\dot{\eta}}{2H} + 9\varepsilon\eta + 3\frac{\varepsilon\dot{\eta}}{H} + 3\varepsilon\eta^2 - 9\varepsilon^2\eta + 4\varepsilon^3 - 12\varepsilon^2 \right] .$$

Here we see that all ϕ derivatives of P are suppressed by one or more slow-roll parameters and are therefore negligible to leading order. Later on, we will give an argument of why this is the case using gauge transformations.

The speed of sound After neglecting the mass term, we focus on the two remaining terms and rewrite the action as

$$S_2 \simeq \int d^3x dt\, a^3\, \frac{1}{2} P_{,X} \left[\frac{\left(P_{,X} + 2P_{,XX}\bar{X} \right)}{P_{,X}} \dot{\varphi}^2 - \partial_i\varphi\partial^i\varphi \right] . \tag{4.13}$$

We can get rid of the overall factor by rescaling φ into a canonically normalized φ_c,

$$\varphi_c = \sqrt{P_{,X}}\varphi . \tag{4.14}$$

This generates some other mass term when the time derivatives act on $P_{,X}$, but those are also slow-roll suppressed and we neglect. In terms of φ_c the action finally looks much nicer,

$$S_2 \simeq \int d^3x dt\, a^3\, \frac{1}{2} \left[c_s^{-2} \dot{\varphi}_c^2 - \partial_i \varphi \partial^i \varphi \right] , \tag{4.15}$$

where we introduced the quantity

$$c_s^2 = \frac{P_{,X}}{\left(P_{,X} + 2P_{,XX}\bar{X}\right)} . \tag{4.16}$$

Here c_s^2 is the relative coefficient between $\dot{\varphi}^2$ and $\partial_i\varphi^2$ and it cannot be removed by redefining φ_c. What is does c_s mean? By dimensional analysis it must have dimension length^2/time^2, which looks just like a speed squared. Indeed, in Lorentz invariant theories this coefficient is the speed of light squared, c^2, which we usually set to one. To convince ourselves that c_s is really a velocity, let's look at the equations of motion from the action (4.15)

$$\ddot{\varphi} + 3H\dot{\varphi} - \frac{c_s^2}{a^2} \partial_i \partial_i \varphi = 0 . \tag{4.17}$$

We know how to solve this equation in general, but let's first solve it in Minkowski, namely setting $a = 1$ and so $H = 0$. Then

$$\ddot{\varphi} - c_s^2 \partial_i \partial_i \varphi = 0 \quad \Rightarrow \quad \varphi(x) \sim e^{\pm i c_s k_p t} e^{\pm i \mathbf{k}_p \mathbf{x}} , \tag{4.18}$$

where $\mathbf{k}_p$ is the physical momentum in Minkowski. This shows that the dispersion relation, which related frequency to wavenumber, is $\omega^2 = c_s^2 k_p^2$ and so c_s indeed describes the velocity at which perturbations φ propagate on the $\bar{\phi}(t)$ background. We will call c_s the *speed of "sound"* to distinguish it from the speed of light and by analogy with sounds waves, which are also described by a scalar field.
To find the exact solutions of (4.17) we can simply follow the same steps used for the $c_s = 1$ case in Sec. 2.1. The only difference is that the mode functions are now

$$f_k = \frac{H}{\sqrt{2c_s k^3}} (1 + i c_s k \tau) e^{-i c_s k \tau} , \tag{4.19}$$

and its complex conjugate is the second independent solution. Notice that $f_k(\tau)$ stops oscillating at some time $-c_s k\tau \sim 1$, and freezes out. This is often referred to as the *crossing of the sound horizon*, where the length c_s/H is the sound horizon. For small speed of sound, $c_s \ll 1$, this happens much earlier than the crossing of the Hubble horizon H^{-1}, discussed around (2.24).

The power spectrum What is the power spectrum of a massless field with a speed of sound c_s? We can simply compute it from $|f_k|^2$ using the above mode functions and taking the $\tau \to 0$ limit. But it is also instructive to proceed in a different way. Since c_s is the only constant with dimension of velocity in the action (4.15), we can simply use dimensional analysis. We will still take $\hbar = 1$, but we should carefully account for the distinction between space and time $c_s T \sim L$, as well as momentum and energy $c_s k \sim E$. The dimensions of the field follow from the action being dimensionless,

$$[S] = 1 \quad \Rightarrow \quad [\phi(\vec{x})] = L^{-1/2} T^{-1/2} \quad \Rightarrow \quad [\phi(\mathbf{k})] = L^3 [\phi(\mathbf{x})] = L^{5/2} T^{-1/2} . \tag{4.20}$$

The power spectrum has units

$$[P(k)] = \frac{[\langle \phi(\mathbf{k})^2 \rangle]}{[\delta_D^3(\mathbf{k})]} = \frac{(L^{5/2}T^{-1/2})^2}{L^3} = \frac{L^2}{T} \,. \tag{4.21}$$

In (2.31) we found the power spectrum for $c_s = 1$. By requiring that it has the right units we find

$$\frac{L^2}{T} = [P(k)] \stackrel{!}{=} \frac{[c_s^n H^2]}{[k^3]} = \frac{(L/T)^n T^{-2}}{L^{-3}} \quad \Rightarrow \quad n = -1 \,. \tag{4.22}$$

We conclude that a field with action (4.15) has a power spectrum

$$P(k) = \frac{H^2}{2c_s k^3} \,. \tag{4.23}$$

It is straightforward to check that this agrees with $|f_k|^2$ from (4.19).

4.2 Cubic interactions and the cosmological bootstrap

To know the interactions that appear in a $P(X, \phi)$ theory we have to expand the Lagrangian to cubic order

$$\begin{aligned} L_3 &= P(\bar{X} + \delta X, \bar{\phi} + \varphi)|_3 & (4.24)\\ &= \frac{1}{2}\left[\frac{1}{3}P_{,XXX}\delta X^3 + P_{,XX}\delta X^2 + 2P_{,X\varphi}\delta X\varphi + P_{,X\varphi\varphi}\delta X\varphi^2\right. & (4.25)\\ &\quad \left. + P_{,XX\varphi}\delta X^2\varphi + \frac{1}{3}P_{,\phi\phi\phi}\varphi^3\right]_3 \\ &= \frac{1}{6}P_{,XXX}\dot{\bar{\phi}}^3\dot{\varphi}^3 - \frac{1}{2}P_{,XX}\dot{\bar{\phi}}\dot{\varphi}(\partial_\mu\varphi)^2 & (4.26)\\ &\quad - \frac{1}{2}P_{,X\varphi}\varphi(\partial_\mu\varphi)^2 + \frac{1}{2}P_{,X\varphi\varphi}\dot{\bar{\phi}}\dot{\varphi}\varphi^2 + \frac{1}{2}P_{,XX\varphi}\dot{\bar{\phi}}^2\dot{\varphi}^2\varphi + \frac{1}{6}P_{,\phi\phi\phi}\varphi^3 \,. \end{aligned}$$

It is a long expression, but it is conceptually very simple. All cubic interactions appear that are allowed by the symmetries. In this case, the only symmetry is rotation invariance, which enforces that every spatial derivative ∂_i is contracted with another spatial derivative. Notice that time and space derivatives appear separately, as opposed to in the Lorentz-invariant combination $(\partial_\mu \phi)^2$, which is familiar from particle physics. We started from the Lorentz invariant Lagrangian $P(X, \phi)$, but we broke boosts and time translations *spontaneously* by choosing a time-dependent vacuum $\phi = \bar{\phi}(t)$. In fact, we can check that by setting $\dot{\bar{\phi}}$ to zero, all non-Lorentz invariant operators disappear.

Again we can use the fact that all background terms with ∂_ϕ derivatives are slow-roll suppressed (see discussion around (4.8)). Hence we focus on the only two operators that do not have this suppression, namely $\dot{\varphi}^3$ and $\dot{\varphi}(\partial_i\varphi)^2$. The operators stand out from other because φ appears always with one derivative. In more general theories each φ can have two or more derivatives. However, interactions where φ appears without derivatives are always slow-roll suppressed[25]. As a consequence, up to slow-roll corrections the

[25]One should be careful here about integration by parts. Of course can always re-write $\dot{\varphi}\partial_i\varphi^2$ as $-\varphi\partial_i(\dot{\varphi}\partial_i\varphi)$. A more precise statement is that interactions that are not shift symmetric are slow-roll suppressed.

action is invariant under a so-called *shift symmetry*, namely (see e.g. [42, 43] for an in-depth discussion)

$$S[\varphi + \text{const}] = S[\varphi] + \mathcal{O}(\text{slow roll }) \,. \tag{4.27}$$

The calculation of the bispectrum induced by these two operators is completely analogous to that of the previous section, with the exception that we should use the mode functions in (4.19). The results are

$$B_{\varphi'^3} = \frac{H^5 \left(P_{,XXX}\dot{\bar{\phi}}^3 + 3P_{,XX}\dot{\bar{\phi}}\right)}{2k_1k_2k_3k_T^3} \,, \tag{4.28}$$

$$\begin{aligned} B_{\varphi'(\partial_i\varphi)^2} = -\frac{1}{8}\frac{H^5 P_{,XX}\dot{\bar{\phi}}}{(k_1k_2k_3)^3\, k_T^3}\Bigg[& 24\,(k_1k_2k_3)^2 - 8k_T\,(k_1k_2k_3)\left(\sum_{a<b} k_a k_b\right) \\ & - 8\,k_T^2 \left(\sum_{a<b} k_a k_b\right)^2 + 22\,k_T^3\,(k_1k_2k_3) - 6k_T^4\left(\sum_{a<b} k_a k_b\right) + 2k_T^6\Bigg] \,. \end{aligned} \tag{4.29}$$

General properties and the cosmological bootstrap When considering models of inflation beyond $P(X,\phi)$, which may contain higher derivative interactions, one finds additional possible cubic interactions[26]. The resulting bispectra have some general properties, which are already visible in the few examples we have discussed so far, e.g. (4.28) and (4.29):

- The bispectrum in principle depends on three vectors, $\{\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3\}$. However, because of statistical homogeneity, momentum is conserved and these three vectors have to form a triangle (hence be co-planar). Then we can always re-write one the $\mathbf{k}_a$'s in terms of the other two. Furthermore, because of statistical isotropy correlators must be invariant under rotations and so can depend only on dot products of vectors. Equivalently, we can say that B_3 depends on the shape of the triangle formed by $\{\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3\}$, but not on its orientation. The shape of a triangle can be specified by three numbers, for example the length of the three sides or the length of two sides and the angle in between. In the following we will use the former parameterization,

  $$B_3(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = B_3(k_1, k_2, k_3) \quad \text{(homogeneity \& isotropy)} \,. \tag{4.30}$$

  More generally, an n-point correlator B_n with $n \geq 3$ depends on $3n-3-3 = 3(n-2)$ variables. This is true non-perturbatively.

- A lemma of the above result is that a bispectrum is always invariant under parity[27]. This is obvious from (7.69) since B_3 only depends on norm of the momenta $\mathbf{k}_a$. Parity violation for scalars first occur in the trispectrum B_4. This is true non-perturbatively.

[26] The relative importance of higher derivative interactions is constrained within the framework of effective field theory by the principle of naturalness, as discussed in Sec. 5.1.

[27] In theoretical physics and especially particle physics, by "parity" we mean point inversion, i.e. the simultaneous inversion of all components of the coordinates, $\mathbf{x} \to -\mathbf{x}$.

- Scale invariance removes one additional variable. For example,

$$B_3(\lambda k_1, \lambda k_2, \lambda k_3) = \frac{B_3(k_1,k_2,k_3)}{\lambda^6} \quad \Rightarrow \quad B_3(k_1,k_2,k_3) = \frac{B_3(1, k_2/k_1, k_3/k_1)}{\lambda^6}\,. \tag{4.31}$$

  So a scale invariant bispectrum is a function of just two variables, which can be taken to be for example k_2/k_1 and k_3/k_1. This is true non-perturbatively.

- Correlators must be symmetric under any permutation σ of the wavenumbers by Bose symmetry

$$B_n(\mathbf{k}_1, \dots, \mathbf{k}_n) = B_n(\mathbf{k}_{\sigma_1}, \dots, \mathbf{k}_{\sigma_n})\,. \tag{4.32}$$

  This is true non-perturbatively.

- At tree level in de Sitter the bispectrum of a massless scalar is a rational function of the norms of the momenta. For manifestly local interactions, namely interactions involving the product of fields and positive powers of their derivatives at the same spacetime point, the denominator is fixed by locality and the choice of the Bunch-Davies initial state to be

$$B_3 = \frac{\text{Poly}_{p+3}(k_1,k_2,k_3)}{(k_1k_2k_3)^3 k_T^p}\,, \tag{4.33}$$

  where p is a non-negative integer that is equal to the number of derivative in the interaction.

- At tree level in de Sitter all bispectra of massless scalars generated by local interactions obey the condition

$$\partial_{k_a}\left[(k_1k_2k_3)^3 B_3\right]_{k_a=0} = 0\,, \tag{4.34}$$

  where the derivative is taken keeping $k_{2,3}$ fixed.

As it is now clear there are many consistency requirements for a function B_3 of the norm of the momenta to be the bispectrum of some theory. So much so that all tree level bispectra of a massless scalar field to all orders in derivatives are actually uniquely determined by the above properties, without the need to perform any explicit in-in calculation! More generally, the idea of determining an observable from a list of consistency requirements that it must satisfy is ofter referred to as "*bootstrapping*" and the above discussion is the starting point of the (boostless) cosmological bootstrap.

5 Pure gravity

ref

So far we have discussed a quantum scalar field on a classical, fixed spacetime. In this section, we tackle the problem of quantizing small spacetime perturbations. To this end, first we review the ideas of Effective Field Theory (EFT), which will give us confidence and guidance in quantizing small metric fluctuations on top of a large classical FLRW spacetime background. We'll begin by consider pure gravity, without any matter present and we will work perturbatively to leading order in the fluctuations (i.e. the linear theory). We will find that the only dynamical degree of freedom is a massless spin-2 particle known as the graviton. We will then describe its free propagation in Sec. 5.3 and its leading self interactions in Sec. 5.4.

Before we proceed we have to exorcise a demon: quantum gravity. There is no obstacle whatsoever in *perturbatively* quantizing gravity. One can also compute all kind of observables in pretty much the same way as for non-Abelian gauge theories. In fact, we know since the early 70's that pure general relativity is even finite at one loop [93]. When people talk about the difficulties of "quantum gravity" they have in mind a full non-perturbative treatment. That is a very hard problem. String theory per se or via the gauge-gravity duality might be a way forward. Instead, here we will only need to study the theory perturbatively, for small deviations from a classical background. In this regime, there is a very well-defined procedure to compute observables. Indeed gravity fits perfectly well within the paradigm that has dominated field theory research in the past half a century: Effective Field Theory (EFT).

5.1 Effective Field Theory

To understand the idea behind EFT, consider a theory with a characteristic scale E_0. For example, it could be the mass of some particle such as the W and Z bosons in the standard model; or, for gravity, this could be the Planck scale $M_{\rm Pl}$ appearing in front of the action. Then, suppose that we are interested in making an experiments at some energy E. If there is a *separation of scales* such that $E \ll E_0$, we can dramatically simplify our description of the system. For example, we might want to make predictions for a particle collider that has energy E much smaller than the mass of the W and Z boson. For the study of inflation, it turns out that the relevant energy E is the Hubble scale H during inflation. This can be understood from the fact that the mode functions of a massless scalar stop oscillating when its energy $\omega = c_s k$ is of order Hubble. From that moment onward all correlators become approximately constants. So in practice we can think of the correlators and a measurements of the interactions of a theory at energy $E = H$. Moreover, H during inflation is bounded by the non-observations of gravitational wave to $H < 10^{-5} M_{\rm Pl}$, and so we have a separation of scales.

The idea of EFT goes as follows. Choose a *cutoff* Λ well above E and close to, but below E_0

$$E \ll \Lambda \lesssim E_0 \,. \tag{5.1}$$

Since the choice of Λ is arbitrary, Λ better cancel out in the final result. Sometimes people use the word "cutoff" also to refer to E_0, which is the highest scale that Λ can be pushed to. Now, divide the fields into a low (L) and a high (H) frequency part

$$\phi = \phi_L + \phi_H \,, \tag{5.2}$$

such that ϕ_L vanishes when its frequency is high, $\omega > \Lambda$, while ϕ_H vanishes when its frequency is low $\omega < \Lambda$. Now the full theory can be formulated in terms of a path integral. Imagine being able to perform the path integral over ϕ_H

$$\int \mathcal{D}\phi_H \mathcal{D}\phi_L e^{iS(\phi_L,\phi_H)} = \int \mathcal{D}\phi_L \left(\int \mathcal{D}\phi_H e^{iS(\phi_L,\phi_H)} \right) \tag{5.3}$$

$$\equiv \int \mathcal{D}\phi_L e^{iS_\Lambda(\phi_L)} \,. \tag{5.4}$$

The new quantity $S_\Lambda(\phi_L)$ is known as *Wilsonian effective action*. Observables computed from $S_\Lambda(\phi_L)$ are UV-finite because ϕ_L vanishes at high energies. But what good does this do us if we cannot compute it?

The effective field theory expansion The key insight is that we can Taylor expand this unknown functional in the low-frequency fields

$$S_\Lambda(\phi_L) = \int d^4x \sum g_a \mathcal{O}_a \,, \tag{5.5}$$

where g_a are some coupling constants and $\mathcal{O}_a$ are *all possible local operators compatible with the symmetries of the problem*. The $\mathcal{O}$'s are build from products of fields and their derivatives at the same spacetime point[28]. The sum contains an infinite number of terms and it is useful only if there is some regime in which we can truncate it by making a negligible mistake. Remarkably, this is precisely what happens at low energies, $E \ll \Lambda \lesssim E_0$. To see this, we must learn how to compare different operators. We will do this using dimensional analysis. For example, say $\mathcal{O}_a$ has mass dimension Δ_a. Then, since the action is dimensionless (in units $\hbar = 1$), we have

$$[\mathcal{O}_a] = \Delta_a \quad \Rightarrow \quad [g_a] = 4 - \Delta_a \,. \tag{5.6}$$

It is convenient to make the dimension of g_a explicit by redefining

$$g_a \equiv \frac{\lambda_a(\Lambda)}{\Lambda^{\Delta_a - 4}} \,, \tag{5.7}$$

where $\lambda_a(\Lambda)$ are dimensionless parameters whose size depends on the arbitrary choice of Λ. We make now the following assumption: when the cutoff is taken close to the characteristic scale E_0 the λ_a's are order unity, unless $\lambda_a = 0$ enlarges the symmetry group of the theory. This assumption goes under the name of *naturalness*, or sometimes "technical" naturalness [92]. In terms of equations, natural theories obey

$$g_a \sim \frac{\mathcal{O}(1)}{E_0^{\Delta_a - 4}} \quad \Leftrightarrow \quad \lambda_a(E_0) \sim \mathcal{O}(1) \,. \tag{5.8}$$

Naturalness is grounded in the study of renormalization in QFT and has been substantiated by countless experimental confirmations. Two major exceptions are of course the small value of the cosmological constant and of the Higgs mass, which are indeed major open problems.

The last ingredient we need is that we want S_Λ to define a weakly coupled theory, which, by definition, gives predictions that are close to the free theory. In this case, we can use the free kinetic term to estimate how large the fields and their derivatives are. For example, consider a massless scalar field[29]

$$S_\Lambda = \int d^4x \, \frac{1}{2} \partial_\mu \phi \partial^\mu \phi \,. \tag{5.9}$$

[28] Some non-locality does emerge at space or time distances of order Λ, but at $E \ll \Lambda$ this can be approximated by local interactions

[29] Dimensional analysis becomes much simpler if the free kinetic term does not contain any dimensionful constant, just as in this scalar field action. This can always be achieved by an appropriate rescaling of the fields.

The free action (5.9) fixes the mass dimension of the field to be $[\phi] = (4-2)/2 = 1$. So when probing the theory at energy E, we can estimate $\phi \sim E$ and $\partial_\mu \sim E$. For example, we estimate

$$\phi^n \sim E^n \,, \quad (\partial\phi)^{2n} \sim E^{4n} \,, \quad (\partial^2\phi)^n \sim E^{3n} \,, \tag{5.10}$$

and so on. We are finally in the position to estimate the size of the terms in the infinite sum (5.5)

$$\int d^4x \, g_a \, \mathcal{O}_a = \frac{1}{E^4} \times \frac{\mathcal{O}(1)}{E_0^{\Delta_a - 4}} \times E^{\Delta_a} \sim \mathcal{O}(1) \left(\frac{E}{E_0}\right)^{\Delta_a - 4} \,. \tag{5.11}$$

This is an important result. It tells us that if $\Delta > 4$, then the operator is very small at low energies $E \ll E_0$. These are called *irrelevant operators*. Conversely, for $\Delta < 4$ the operator is called *relevant* and indeed becomes large at low energies. *Marginal operators* with $\Delta = 4$ are in between. Their faith depends on whether loop corrections push their dimension above or below four.

Most field theories you can think of are EFT's. For example the Fermi theory of weak interaction is an EFT below $E_0 \sim m_{W,Z} \sim 80$ GeV. The chiral Lagrangian that describes the interaction of pions is an EFT below the confinement scale of QCD, $E_0 \sim$ GeV. The standard model of particle physics, when extended to include neutrino masses is an EFT.

Renormalizability In the old days, theories that include irrelevant operators used to be called "non-renormalizable" theories. This is a misnomer. It comes from the observations that, if an irrelevant operator is present, it can be shown to generate infinitely many other irrelevant operators via loop corrections. Naively one would then need to know/measure the infinitely many coupling constant with ever increasing dimension and the theory seems to be doomed. But now we understand this is not the case! Operators with large dimension give very small correction to low-energy processes. To be more precise, imagine you want to make a prediction for an experiment at energy $E \ll E_0$ that has precision δ. For example $\delta = 10^{-2}$ for percent level predictions and so on. By the estimate (5.11), you only need to include operators up to dimension Δ_{max} such that

$$\left(\frac{E}{E_0}\right)^{\Delta_{\text{max}} - 4} < \delta \,. \tag{5.12}$$

There is a finite number of such operators and so you only need a finite number of couplings. In this precise sense there is no problem in renormalizing and EFT (see [77] for a rigorous proof).

5.2 Gravity as an Effective Field Theory

So what about gravity? We might recall hearing our nursery friends saying that "General Relativity is non-renormalizable", because there are irrelevant operators. Let us see how this works. The starting point is the Einstein-Hilbert action

$$S = \frac{M_{\text{Pl}}^2}{2} \int d^4x \, \sqrt{-g} R \,. \tag{5.13}$$

Let us expand it in small perturbations around some background $g_{\mu\nu} = \bar{g}_{\mu\nu} + h_{\mu\nu}$. Dropping all indices, the expanded action looks like

$$S = \frac{M_{\rm Pl}^2}{2}\int d^4x\, \left[\partial h \partial h + h\partial h \partial h + \dots\right] . \tag{5.14}$$

where the dots contain terms with more powers of h. To connect to our discussion in the previous section, we need to normalize the free action so that no dimensionful constants appear. This is achieved by

$$h_{\mu\nu} \to \tilde{h}_{\mu\nu} = h_{\mu\nu} M_{\rm Pl} . \tag{5.15}$$

In terms of the canonically normalized field $\tilde{h}_{\mu\nu}$, the action looks like

$$S = \frac{1}{2}\int d^4x\, \left[\partial\tilde{h}\partial\tilde{h} + \frac{1}{M_{\rm Pl}}\tilde{h}\partial\tilde{h}\partial\tilde{h} + \dots\right] , \tag{5.16}$$

which has now the same schematic form as (5.9). We recognize that the second term has dimension five and so it's an irrelevant operator. The infinitely many other terms hidden in the dots have even larger dimension and are even more irrelevant. We also recognize that $M_{\rm Pl}$ plays the role of the characteristic scale, $E_0 \sim M_{\rm Pl}$. Therefore, we know that we can quantize GR and make predictions at energies well-below the Planck scale. This is just what we will do in the remainder of this section.

Let me briefly mention some simplifying assumptions we have made in this discussion. First, a theory might have more than one scale. Estimating the size of operators then requires more care, as for example when considering higher derivative corrections to GR, which we did not discuss. Second, we were very vague as to how high- and low-frequency fields should be separated. In fact, the frequency itself is not a Lorentz invariant concept. Even worse, in a time dependent background, both energy and momentum are red- or blu-shifted. Finally, there exists important non-perturbative effects, such as for example tunneling in quantum mechanics. These can often be computed within the EFT, but important subtleties arise.

5.3 Quantizing general relativity: gravitons

The EFT discussion has emboldened us to quantize gravity just like we would quantize any other EFT (e.g. the scalar field in Sec. 2). Here we will start with the pure general relativity without any matter. First, we will quantized the linearized theory and then discuss the leading interactions in Sec. 5.4. At this level, we are allowed to neglect the technical complication coming from the fact that we are quantizing a gauge theory. In the next section, where will discuss these completions and also include matter.

Familiar challenges Let's divide the metric into a classical background $\bar{g}_{\mu\nu}$ and small fluctuations $h_{\mu\nu}$

$$g_{\mu\nu}(x,t) \quad = \quad \bar{g}_{\mu\nu}(t) + h_{\mu\nu}(x,t) . \tag{5.17}$$

We wish to quantize $h_{\mu\nu}$. To do this we face two interconnected challenges that are familiar from quantizing the gauge field A_μ of electromagnetism or the worldsheet of string theory. The first challenge is that some components of $h_{\mu\nu}$ have a vanishing conjugate momentum and hence are non-dynamical. The second and related challenge is that changing coordinates induces a gauge transformation of $h_{\mu\nu}$. Just like for E&M or string theory we have two choices on how to proceed (see e.g. the nice discussion in [96]). The first option is to fix a convenient gauge and explicitly solve for all non-dynamical fields so that they are removed from the action in favour of the dynamical ones. This is what we do when we quantize E&M in Coulomb gauge and is the path we take in these notes. As we will see, the price to pay is that neither Lorentz invariance nor locality are manifest. The second option is keep things manifestly Lorentz invariant and gauge invariant by artificially giving dynamics to the non-dynamical degrees of freedom. Then, after canonical quantization we will have to put constraints of the now too large Hilbert space to reduced it down to the set of physical states. For non-Abelian gauge theories and gravity this leads to the Fadeev-Popov ghosts and BRST quantization. We will not discuss this second possibility here.

A priori, there are ten independent components of $h_{\mu\nu}$. However, four components in $\delta g_{\mu\nu}$ can be set to zero by a change of coordinates $x^\mu \to x^\mu + \epsilon^\mu(x)$, where $\epsilon^\mu(x)$ is a set of four arbitrary functions. This changes the metric by (the symmetrization of indices is defined in (A.5))

$$g_{\mu\nu} \to g_{\mu\nu} + 2\nabla_{(\mu}\epsilon_{\nu)} \,. \tag{5.18}$$

Moreover, four additional component, namely $h_{0\mu}$ obey four constraint equations (see Box 1), which are at most first order in time derivatives, and therefore are not dynamical. It turns out that to derive the action to order $(2n+1)$ in perturbations it suffices to solve these constraints to order n. Hence, for the quadratic and cubic action we need only to solve the linearized constraints. The four constraints are organized into one scalar and one one vector equation and hence cannot contain the graviton at linear order, since γ_{ij} is a two-index tensor. We will hence neglect the completely here and rectify this in the next section.

The graviton The remaining $10-4-4=2$ components of $h_{\mu\nu}$ are dynamical and describe the two helicities of a massless spin-2 particle known as the *graviton*. A convenient choice of coordinates to study the linear dynamics of gravitons on an FLRW spacetime is

$$ds^2 = -dt^2 + a^2\left(\delta_{ij} + \gamma_{ij}\right) dx^i dx^j \quad \text{(linear order)}\,, \tag{5.22}$$

where γ_{ij} is transverse, $\partial_i \gamma_{ij} = 0$, and traceless, $\gamma_{ii} = 0$, and so has indeed only $6-3-1=2$ independent components[30]. For the moment $a(t)$ is arbitrary. We should now expand the Einstein-Hilbert action,

$$S = \frac{M_{\rm Pl}^2}{2}\int d^4x\sqrt{-g}R\,, \tag{5.23}$$

[30] Consider a plane wave $\gamma_{ij}(k\hat{x})$ and a rotation $R(\theta\hat{x})$ by θ around $\hat{x}$. From $h'_{ij} = (R^T(1+\gamma)R)_{ij}$ it is straightforward to see that γ_{ij} is a linear combination of helicity ± 2 components that transform as $\gamma_{ij}^{\pm} \to e^{\pm 2i\theta}\gamma_{ij}^{\pm}$.

Box 5.1 The Bianchi identities The Riemann tensor must obey the following partial differential equation

$$\nabla_\lambda R_{\alpha\beta\mu\nu} + \nabla_\nu R_{\alpha\beta\lambda\mu} + \nabla_\mu R_{\alpha\beta\nu\lambda} = 0\,, \tag{5.19}$$

which follows from the fact that covariant derivatives must obey the Jacobi identities. A contracted version of this relation is known as (contracted) Bianchi identity

$$\nabla^\mu G_{\mu\nu} \equiv \nabla^\mu \left(R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R \right) = 0\,. \tag{5.20}$$

Writing out the covariant derivative we find

$$\partial_t G^{t\nu} = -\partial_k G^{k\nu} - \Gamma^\alpha_{\alpha\gamma} G^{\nu\gamma} - \Gamma^\nu_{\alpha\gamma} G^{\alpha\gamma}\,. \tag{5.21}$$

Since the right-hand side has at most second derivatives of the metric, we conclude that $G^{t\nu}$ has at most one time derivative. But then the metric must appear with just one time derivative in four of the Einstein equations, namely $G^{t\nu} = T^{t\nu}$. These must then be constraint equations that limit the set of consistent initial data $\{h_{\mu\nu}, \dot{h}_{\mu\nu}\}$ that one can specify. As we will see shortly in Sec. 6.1, we can freely specify g_{ij} while $g_{0\mu}$ are instead fixed by a set of four constraint equations, which are first order in time derivatives.

to quadratic order in γ_{ij}. At the end of a long but straightforward calculation[31], one finds

$$S_2 = \frac{M_{\rm Pl}^2}{8} \int d^3x d\tau\, a^2 \left[\gamma'_{ij}\gamma'_{ij} - \partial_k \gamma_{ij} \partial_k \gamma_{ij} \right]\,. \tag{5.26}$$

This action could have been easily guessed as it contains the only two terms allowed by the symmetries of the problem[32]. As we do for the photon, we can expand the graviton in plane waves by writing[33]

$$\gamma_{ij}(x) = \int_{\mathbf{k}} \sum_s \epsilon^s_{ij}(\mathbf{k}) \gamma_s(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}}\,, \tag{5.27}$$

where $\epsilon^s_{ij}(\mathbf{k})$ are *polarization tensors* and s takes two values corresponding to the two polarizations. There are many possible choices of basis for the two polarization tensors. A common choice in the study of gravitational waves at interferometers are the plus and

[31] Some useful steps along the way include (see Sec. 5.1 of [104])

$$\delta\Gamma^i_{jk} = \frac{1}{a^2}\left[\partial_{(k} h_{ij)} - \frac{1}{2}\partial_i h_{jk}\right], \qquad \delta\Gamma^i_{j0} = \frac{1}{2a^2}\left[\dot{h}_{ij} - H h_{ij}\right], \qquad \delta\Gamma^0_{ij} = \frac{1}{2}\dot{h}_{ij}, \tag{5.24}$$

$$\delta R_{ij} = \frac{1}{2a^2}\partial_i^2 h_{ij} - \frac{1}{2}\ddot{h}_{ij} + \frac{H}{2}\dot{h}_{ij} - 2H^2 h_{ij}, \quad \delta R_{00} = \delta R_{0j} = \delta\Gamma^0_{00} = 0, \quad \delta\Gamma^i_{00} = \delta\Gamma^0_{i0} = 0. \tag{5.25}$$

[32] By rotational invariance one has to contract the two indices of γ_{ij}. Any contraction with δ_{ij} or ∂_i gives zero, so the only possibility is contracting with another γ_{ij}. The Ricci scalar contains two derivatives, which can act on the background (e.g. on $a(t)$) or on γ_{ij}. The only terms with two derivatives on perturbations are those in S_2. The relative factor is what we call *speed of light* and has been set to unity here. Terms with one time derivative can be integrated by part into terms without any derivatives. Finally, terms without any derivatives cannot be invariant under diffs, (5.18), so they must all cancel out.

[33] Notice that $\gamma_{ij}(x)$ and ϵ^s_{ij} are dimensionless, while $[\gamma^s(\mathbf{k})] \sim M^{-3}$.

cross polarizations, $s = +, \times$. Another common choice inherited from particle physics are circular polarization, $s = \pm 2$, where each polarization is an eigenvector of rotations around the momentum. We will not need to make any specific choice here, but explicit expressions are derived in Appendix C. In general, polarization tensors are complex and satisfy

$$\begin{aligned}
\epsilon^s_{ii}(\mathbf{k}) = k^i \epsilon^s_{ij}(\mathbf{k}) &= 0 && \text{(transverse and traceless)}\,, && (5.28)\\
\epsilon^s_{ij}(\mathbf{k}) &= \epsilon^s_{ji}(\mathbf{k}) && \text{(symmetric)}\,, && (5.29)\\
\epsilon^s_{ij}(\mathbf{k})\epsilon^s_{jk}(\mathbf{k}) &= 0 && \text{(lightlike)}\,, && (5.30)\\
\epsilon^s_{ij}(\mathbf{k})\epsilon^{s'}_{ij}(\mathbf{k})^* &= 2\delta_{ss'} && \text{(normalization)}\,, && (5.31)\\
\epsilon_{ij}(\mathbf{k})^* &= \epsilon_{ij}(-\mathbf{k}) && (\gamma_{ij}(x)\text{ is real})\,. && (5.32)
\end{aligned}$$

Let's re-write the action using this decomposition:

$$S_2 = \frac{M_{\rm Pl}^2}{4} \int_{\mathbf{k}} d\tau\, a^2 \sum_s \left[\gamma_s'(\mathbf{k})\gamma_s'(-\mathbf{k}) - \frac{k^2}{a^2}\gamma_s(\mathbf{k})\gamma_s(-\mathbf{k}) \right] . \tag{5.33}$$

Now this action consists of two independent copies of the action for a massless scalar field (2.3), up to an overall factor of $M_{\rm Pl}^2/2$. The two polarizations γ_s are now canonically normalized. To quantize the theory we can then proceed exactly as we did in Sec. 2.1. We promote $\gamma_s(\mathbf{k})$ to an operator and write it in terms of creation and annihilation operators

$$\gamma_s(\mathbf{k}) = \frac{\sqrt{2}}{M_{\rm Pl}} \left(f_k a^s_{\mathbf{k}} + f_k^* a^{s\dagger}_{-\mathbf{k}} \right) . \tag{5.34}$$

where the commutation relations are the usual ones,

$$[a^s_{\mathbf{k}}, a^{s'}_{\mathbf{k}'}] = 0 \qquad\qquad [a^s_{\mathbf{k}}, a^{s'\dagger}_{\mathbf{k}'}] = (2\pi)^3 \delta^3_D \left(\mathbf{k} - \mathbf{k}'\right) \delta_{ss'}\,. \tag{5.35}$$

If we assume a dS background, i.e. $a = e^{Ht}$, the mode functions f_k are the same as for the massless scalar field, (2.23). The graviton power spectrum, often called the *tensor power spectrum* $P_T(k)$ can be easily computed

$$\begin{aligned}
\langle \gamma_{ij}(\mathbf{k})\gamma_{ij}(\mathbf{k}')\rangle &= \sum_{s,s'} \epsilon^s_{ij}(\mathbf{k})\epsilon^{s'}_{ij}(\mathbf{k}')\langle\gamma_s(\mathbf{k})\gamma_{s'}(\mathbf{k}')\rangle && (5.36)\\
&= \frac{2}{M_{\rm Pl}^2}\sum_{s,s'} \epsilon^s_{ij}(\mathbf{k})\epsilon^{s'}_{ij}(\mathbf{k}')(2\pi)^3\delta^3_D(\mathbf{k}+\mathbf{k}')|f_k|^2 && (5.37)\\
&= \frac{2}{M_{\rm Pl}^2}\frac{H^2}{2k^3}(2\pi)^3\delta^3_D(\mathbf{k}+\mathbf{k}')\sum_{s,s'} 2\delta_{ss'} && (5.38)\\
&= (2\pi)^3\delta^3_D\left(\mathbf{k}+\mathbf{k}'\right)P_T \quad \text{with} \quad P_T = \frac{4}{k^3}\frac{H^2}{M_{\rm Pl}^2}\,. && (5.39)
\end{aligned}$$

A few comments are in order. First, just like in the scalar case, the tensor power spectrum in de Sitter is scale invariant, i.e. $P_T(k) \propto k^{-3}$. Second, the amplitude $(2H/M_{\rm Pl})^2$ is completely fixed by the scale of inflation. When $H \sim M_{\rm Pl}$ the amplitude becomes of order one and the semiclassical, perturbative approach we have followed so far breaks down. To study that regime we would need a full theory of quantum gravity.

Luckily for us, even though H during is the second most uncertain scale in science[34], we already know that $H \ll M_{\rm Pl}$ because we have not yet see any primordial gravitons. The current leading observational bound[35] comes from the non-detection of primordial B-mode polarization of the cosmic microwave background. Any positive detection would non only confirm the general picture proposed by inflation but would also measure the energy scale at which it took place. Finally, notice that the two polarization of the graviton have the same power spectrum. This is a consequence of parity invariance of the GR action, since parity maps a helicity $s = +2$ graviton into one with $s = -2$.

5.4 Graviton self interactions

Once we have defined the free theory, the EFT approach allows us to consistently discuss small interactions in perturbation theory. The leading interactions arise from expanding the Einstein Hilbert action to the next order, namely $\mathcal{O}(\gamma^3)$. Before expanding, we need to amend our definition of γ_{ij}. In principle we could keep the definition in (5.22) beyond linear order, but it turns out that it is more convenient to define

$$g_{ij} = a^2 \left(e^{\gamma}\right)_{ij} = a^2 \left[\delta_{ij} + \gamma_{ij} + \frac{1}{2}\gamma_{il}\gamma_{lj} + \frac{1}{3!}\gamma_{il}\gamma_{lm}\gamma_{mj} + \dots\right] , \tag{5.40}$$

where γ_{ij} is again symmetric, traceless and transverse and e^{γ} denotes the *matrix exponential* of the matrix γ_{lm}. Notice that that this agrees with the previous definition in (5.22) to linear order. One reason why the new definition is convenient is that the metric determinant is independent of γ_{ij} to all orders

$$a^{-6} \det(g_{ij}) = \exp\left[\log\det(e^{\gamma})\right] = \exp\left[\mathrm{Tr}\log(e^{\gamma})\right] = \exp\left[\mathrm{Tr}\,\gamma\right] = 1\,. \tag{5.41}$$

Using this definition a lengthy but straightforward calculation leads to the cubic action for gravitons

$$\mathcal{L}_3 = -\frac{M_{\rm Pl}^2}{8} a\gamma_{ij}\partial_j\gamma_{km}\left(\partial_i\gamma_{km} - 2\partial_m\gamma_{ik}\right) . \tag{5.42}$$

This result is partially as expected: we know that a cubic interaction should contain three γ_{ij}'s and two derivatives, and that all the spatial indices should be contracted. The only surprising thing is the absence of time derivatives, but that is an accident of the cubic order and time derivatives do appear from $\mathcal{O}(\gamma^4)$ onward. The above interaction generates a graviton bispectrum, which can be computed using the in-in formalism just as we did for the scalar bispectrum. The result is [67]

$$\begin{aligned}\langle\gamma^{s_1}(\mathbf{k}_1)\gamma^{s_2}(\mathbf{k}_2)\gamma^{s_3}(\mathbf{k}_3)\rangle' = -&\frac{H^2}{2M_{\rm Pl}^4}\frac{\epsilon^{s_1}_{ii'}\epsilon^{s_2}_{jj'}\epsilon^{s_3}_{ll'}t_{ijl}t_{i'j'l'}}{(k_1k_2k_3)^3} \\ &\times\left[\frac{k_1k_2k_3}{k_T^2} + \frac{k_1k_2 + k_1k_3 + k_2k_3}{k_T} - k_T\right] ,\end{aligned} \tag{5.43}$$

[34] Values of Hubble during inflation in the range 10^{13}GeV $< H \ll 10^{-26}$GeV are still compatible will observations. This is a 39 orders-of-magnitude interval! The upper bound comes from the non-detection of primordial gravitons, also called "tensor" modes. The lower bound comes from demanding that the reheating temperature at the beginning of the hot big bang is well above that of big bang nucleosynthesis, $T_{\rm BBN} \sim 0.1$MeV.

[35] These are usually expressed in terms of the *tensor-to-scalar ratio* $r \equiv \frac{P_T}{P_{\mathcal{R}}}$, where $P_{\mathcal{R}}$ is the power spectrum of the gauge invariant scalar $\mathcal{R}$ to be introduced in Sec. 6.6. The current bound is $r < 0.036$ at 95% confidence [2].

where we introduce the tensor

$$t_{ijl} = k_2^i \delta_{jl} + k_3^j \delta_{il} + k_1^l \delta_{ij} \,. \tag{5.44}$$

While this result is still rather compact, things get out of hand pretty quickly as one move on to higher orders, where one has to solve the constraint equations and the number of terms proliferates. For reference, the four-point correlator in de Sitter of graviton was first written down in complete form only in [20] (see [45, 64] for earlier partial results, and [79] for expressions in anti-de Sitter).

6 Gravity and matter

ref

An empty spacetime is pretty boring. In this section, we learn how to describe a system where both matter, in the form of a scalar field, and spacetime are dynamical. This requires discussing three more ingredients: the Arnowit-Deser-Misner (ADM) formalism [9] and the associated scalar-vector-tensor decomposition, the constraint equations and gauge transformations. The final result will be the definition of an appropriate (perturbatively) gauge-invariant observable, which we will call $\mathcal{R}$, and a prescription to compute its correlators.

6.1 Constraint equations from the ADM formalism

Let's introduce the ADM formalism due to Arnowitt, Deser and Misner. Given a 3+1 decomposition of spacetime specified by a family of spatial hypersurfaces $\Sigma(t)$ with t a time function[36] and h_{ij} the spatial metric on $\Sigma(t)$, we can write down the most generic spacetime metric $g_{\mu\nu}$ as

$$ds^2 = -N^2 dt^2 + h_{ij}(dx^i + N^i dt)(dx^j + N^j dt) \,, \tag{6.1}$$

where $N(x)$ is called the *lapse* and $N^i(x)$ is called the *shift*. The lapse has a natural interpretation if we write down a one-form field n_μ that is everywhere perpendicular to $\Sigma(t)$ and normalized so that $n^\mu g_{\mu\nu} n^\nu = -1$,

$$n_\mu = \begin{pmatrix} -N \\ \mathbf{0} \end{pmatrix}, \qquad\qquad n^\mu = \frac{1}{N}\begin{pmatrix} 1 \\ -N^i \end{pmatrix}. \tag{6.2}$$

The spacetime metric is hence decomoposed into time-time, time-space and space-space parts

$$g_{\mu\nu} = \begin{pmatrix} -N^2 + N_k N^k & N_j \\ N_i & h_{ij} \end{pmatrix}, \quad g^{\mu\nu} = \frac{1}{N^2}\begin{pmatrix} -1 & N^j \\ N^i & N^2 h^{ij} - N^i N^j \end{pmatrix}, \tag{6.3}$$

where spatial indices are lowered and raised with the spatial metric h_{ij}. The determinant of the $g_{\mu\nu}$ takes the simple form

$$\sqrt{-g} = \sqrt{h} N \,. \tag{6.4}$$

[36] A *time function* is a function from the spacetime to the real numbers whose gradient is always future-directed and timelike.

Box 6.1 The ADM formalism General covariance, namely the freedom of choosing coordinates, creates the following problem in quantizing general relativity. We know that the correct initial value problem in GR consists in specifying only the spatial part of the metric and its time derivative at an instant of time, subject to four constraint equations. This is essential to ensure that one can described the same solution using a different set of coordinates that happen to agree with the original ones at that instant of time. To make progress we therefore need to decompose spacetime into space and time. This is called a "3+1 decomposition" and is achieved by the ADM formalism. In the ADM formalism we derive and solve the constraint equations for the non-dynamical components of the metric and then we can proceed as usual to canonically quantization.

The 3+1 decomposition can be written in two way: by keeping covariance manifest or by choosing convenient coordinates. The former approach proceeds as follows. First, introduce an artificial and arbitrary infinite family of spacial hypersurfaces $\Sigma(t)$ parameterized by some time function t that foliate spacetime. The hypersurfaces are defined everywhere by a perpendicular vector field n^μ that is normalized to $n^\mu n_\mu = -1$ and always future-pointing. Whenever you want to take the "time component" of some tensor, just contract the desired index with n^μ. Then introduce the projector $h_{\mu\nu} = g_{\mu\nu} + n_\mu n_\nu$ onto $\Sigma(t)$, which is immediately seen to be perpendicular to n^μ. If you now want to take the spatial component of some tensor, just contract the desired index with $h_{\mu\nu}$.

The second approach of choosing coordinates is somewhat more convenient for explicit calculations. We start again with a time function t. Hypersurfaces $\Sigma(t)$ of constant t are an infinite family of spacelike hypersurfaces that foliate spacetime. Let's parameterize the hypersurface $\Sigma(0)$ at $t = 0$ with the three spatial coordinates x^i. Points on $\Sigma(t)$ for $t \neq 0$ can be brought onto $\Sigma(0)$ by moving them along lines that are everywhere tangent to the gradient of t. This gives us a set of global coordinates $\{t, x^i\}$, which are the one we used in the main text.

The normal vector n^μ need not be the same everywhere in space and time. Its variation tells us how $\Sigma(t)$ is embedded into spacetime. In general $\nabla_\mu n_\nu$ is non-zero and given by[37],

$$\nabla^\mu n^\nu = \begin{pmatrix} 0 & a^i \\ 0 & K^{ij} \end{pmatrix} . \tag{6.5}$$

The spatial matrix K_{ij} is called the *extrinsic curvature*[38] and represents the change of the spatial components of n^μ as one moves along Σ. In terms of the metric components the extrinsic curvature is

$$\begin{aligned} K_{ij} \equiv n_{i;j} = n_{i,j} - \Gamma^\lambda_{ij} n_\lambda = N\Gamma^0_{ij} &= \frac{1}{2} N g^{0\mu} \left(g_{\mu j,i} + g_{i\mu,j} - g_{ij,\mu}\right) & (6.6)\\ = -\frac{1}{2N}\left(g_{0j,i} + g_{i0,j} - g_{ij,0}\right) &+ \frac{1}{2N} N_l h^{lm} \left(h_{mj,i} + h_{im,j} - h_{ij,m}\right) & (6.7)\\ = \frac{1}{2N}\left(\dot h_{ij} - N_{i,j} - N_{j,i}\right) + \frac{1}{N}{}^{(3)}\Gamma^l_{ij} N_l &= \frac{1}{2N}\left(\dot h_{ij} - 2^{(3)}\nabla_{(i}N_{j)}\right) , & (6.8) \end{aligned}$$

where we symmetrise with weight one, i.e. $X_{(ij)} = (X_i + X_j)/2$, the connection ${}^{(3)}\Gamma^i_{jk}$ refers to the 3-dimensional metric h_{ij} and ${}^{(3)}\nabla_i$ the related 3-dimensional covariant

[37] Since $n^\mu n_\mu = -1$, we have that $n^\nu \nabla_\mu n_\nu = 0$ which is why some components of (6.5) vanish.

[38] In the literature one sometimes sees a rescaled extrinsic curvature, $E_{ij} \equiv NK_{ij}$, so that E_{ij} is independent of the lapse N.

derivative. The spatial vector a^i in (6.5) is called the *acceleration*, and is given by,

$$a_i = n^\mu \nabla_\mu n_i = \frac{^{(3)}\nabla_i N}{N} \, . \tag{6.9}$$

Decomposing $\nabla_\mu \phi$ into its spatial components and an orthogonal component along n^μ gives

$$\begin{aligned} \phi_n &:= n^\mu \nabla_\mu \phi = \frac{\left(\partial_t - N^i \partial_i\right) \phi}{N} \\ \phi_i &:= {}^{(3)}\nabla_\mu \phi = \partial_i \phi \, . \end{aligned} \tag{6.10}$$

Then, the scalar $X = -\frac{1}{2} g^{\mu\nu} \nabla_\mu \phi \nabla_\nu \phi$ is given simply by[39]

$$X = -\frac{1}{2} \left[-(\phi_n)^2 + \partial_i \phi \partial_j \phi h^{ij} \right] \, . \tag{6.11}$$

So like with the line element, using n^μ as a time-like direction effectively removes the cross-terms like $g^{0i} \dot{\phi} \partial_i \phi$ which would be present had we used the naive co-ordinate time direction (i.e. $\dot{\phi} = \partial_t \phi = \hat{t}^\mu \nabla_\mu \phi$ and $\hat{t}^\mu$ is not orthogonal to the spatial directions).
The *Gauss-Codazzi equation* relates the 4-dimensional Ricci scalar R to the three dimensional one ${}^{(3)}R$ as

$$R = {}^{(3)}R + \left(K_{ij} K^{ij} - K^2\right) - 2 \nabla_\alpha \left(n^\beta \nabla_\beta n^\alpha - n^\alpha \nabla_\beta n^\beta \right) \, . \tag{6.12}$$

Notice that ${}^{(3)}R$ depends on h_{ij} but not on N or N_i. The last term leads to a total derivative in the action and so drops out[40]. This formula allows us to re-write the action in terms of the ADM variables

$$S = \frac{M_{\rm Pl}^2}{2} \int d^4 x \sqrt{h} N \left[{}^{(3)}R + K_{ij} K^{ij} - K^2 \right] \, . \tag{6.13}$$

We see explicitly that N appears in the action without any derivatives, while N_i appears with spatial but no time derivatives. This remains true as we couple (minimally) a matter sector to gravity. To see this, let's also write the $P(X, \phi)$ action in the ADM formalism:

$$S = \int d^4 x \, \sqrt{-g} \, P(X, \phi) = \int d^3 x dt \, N \sqrt{h} \, P(X, \phi) \, , \tag{6.14}$$

where of course there are some N and N_i hiding in X. For example

$$\frac{\partial g^{\mu\nu}}{\partial N} = -\frac{2}{N} \left(g^{\mu\nu} - \delta^\mu_i \delta^\nu_j h^{ij} \right) \, , \tag{6.15}$$

As we vary the action with respect to $\{N, N^i\}$ we obtain so called *constraint equations*, in which N and N^i appear without time derivatives and h_{ij} with at most first time derivatives[41]:

$$\frac{\delta S}{\delta N} = 0 \qquad \Rightarrow \qquad {}^{(3)}R - K_{ij} K^{ij} + K^2 + \frac{2}{M_{\rm Pl}^2} \left[P - P_{,X} \left(2X + \partial_i \phi \partial^i \phi \right) \right] = 0 \, , \tag{6.16}$$

$$\frac{\delta S}{\delta N^j} = 0 \qquad \Rightarrow \qquad \nabla_j [K^j_i - \delta^j_i K] + \frac{P_{,X}}{M_{\rm Pl}^2 N} \partial_i \phi \left(N^j \partial_j \phi - \dot{\phi} \right) = 0 \, . \tag{6.17}$$

[39] Similarly, a second derivative like $\Box \phi = g^{\mu\nu} \nabla_\mu \nabla_\nu \phi$ can be decomposed as $\Box \phi = -\phi_{nn} + {}^{(3)}g^{ij} \phi_{ij}$ where $\phi_{nn} = \frac{(\partial_t - N^i \partial_i) \phi_n}{N} - a^i \phi_i$ and $\phi_{ij} = {}^{(3)}\nabla_i \phi_j - K_{ij} \phi_n$ (see e.g. [61, Sec. IV.A]).

[40] The boundary term cancels the Gibbons-Hawking-York boundary term.

[41] These expressions agree with (21) of [86], which was written in flat gauge.

6.2 The scalar-vector-tensor decomposition

Before solving these complicated equations, let's discuss two important tools that we can employ to simplify them. The first one is the Scalar-Vector-Tensor (SVT) decomposition. The main idea is the usual one: choose your variables according to the symmetries of the problems. Since all FLRW backgrounds are homogeneous and isotropic, it is a good idea to work with objects that transform nicely under spatial rotations and translations. Mathematically, these are the irreducible representations of the ISO(3) isometry group, which can be obtained using the same method of induced representations, which is also used to define particles in particle physics (see Sec. 10.9 of [71]). Here we follow a more pedestrian approach.

Spatial translations are easily diagonalized by working in Fourier space. For rotations, we separate objects with zero, one and two spatial indices and call them rotation-scalars, rotation-vectors and rotation-tensor, respectively. For example, the field perturbation φ has no spatial indices and already transforms as a scalar under rotations. In particular, for $x^i \to x'^{i'} = R_i^{i'} x^i$, it transforms as

$$\varphi'(x') = \varphi(x) \,. \tag{6.18}$$

The metric perturbation $h_{\mu\nu}$ instead is more complicated. It is a symmetric 4×4 matrix with 10 independent entries. These can be separated into rotation-scalars, rotation-vectors and rotation-tensors with the following definitions

$$h_{i0} = N_i \equiv a^2 \partial_i \psi + N_i^V \tag{6.19}$$
$$h_{ij} \equiv a^2 \left[\delta_{ij} A + \partial_{ij} B + \partial_{(i} C_{j)} + \gamma_{ij} \right] \,, \tag{6.20}$$

where all the rotation-vectors are also transverse, in the sense that $\partial_i N_i^V = \partial_i C_i = 0$ and the rotation-tensor is both transverse and traceless $\gamma_{ii} = \partial_i \gamma_{ij} = 0$. Let's check that the number of variables matches the 10 independent entries of $h_{\mu\nu}$. We have four rotation-scalars h_{00}, A, B and ψ, accounting for 4 variables; two transverse rotation-vectors C_i and N_i^V, which with their two "polarizations" each account for $2 + 2 = 4$ variables; finally one transverse traceless rotation-tensor γ_{ij} with its two polarizations accounts for the remaining 2 variables. A similar decomposition can be performed for all other variables in the problem, e.g for the energy-momentum tensor, but we will not need this here.

Now the crucial point: rotation-scalars, transverse rotation-vectors and transverse traceless rotation-tensors *decouple from each other at linear order*, meaning that in solving the equations of motion for one we can set the others to zero. After finding solutions, we can simply add them up. In the rest of these notes, we will drop the word "rotation-" and simply call the various components scalars, vectors or tensors. You should be aware though that the word "scalar" sometimes refers to a Lorentz scalar, such as ϕ, while sometimes is refers to a rotation-scalar, such as φ or h_{00}.

6.3 Diffs as gauge transformations

Let's move on to gauge transformations. In GR, we can always perform a coordinate transformation to simplify the equations. Consider the coordinate transformation

$$x^\mu \to x'^\mu = x^\mu + \epsilon^\mu(x) \,, \tag{6.21}$$

for arbitrary $\epsilon^\mu(x)$. We know that $g_{\mu\nu}$ and ϕ transform as a two-tensor and a scalar respectively, namely as

$$\phi'(x') = \phi(x)\,, \quad g'_{\mu'\nu'} = \frac{\partial x^\mu}{\partial x'^{\mu'}} \frac{\partial x^\nu}{\partial x'^{\nu'}} g_{\mu\nu}\,. \tag{6.22}$$

But how do the perturbations $h_{\mu\nu}$ and φ transform? We have the freedom to specify how the background part and the perturbations transform separately, while keeping their sum covariant. A convenient and very common way to resolve this ambiguity is to work with so called *gauge transformations*, in which case the background is kept fixed and all the transformation of the full tensor is attributed to the perturbations. More in detail, the rules are the following

1. Transform the full tensor covariantly, as in (6.22), and keep the background unchanged.
2. Drop the prime from the new coordinates.
3. Attribute all the transformation to the perturbations.

For example for a scalar field $\phi(x) = \bar{\phi} + \varphi$, one finds the transformation $\Delta\varphi$ to be

$$\Delta\varphi \equiv \phi'(x) - \phi(x) = \phi(x-\epsilon) - \phi(x) = -\epsilon^\mu \partial_\mu \phi(x) + \mathcal{O}(\epsilon^2)\,, \tag{6.23}$$

where we used

$$\phi'(x') = \phi(x) \quad \Rightarrow \quad \phi'(x) = \phi(x-\epsilon)\,. \tag{6.24}$$

For a homogeneous background, $\bar{\phi}(x) = \bar{\phi}(t)$, to linear order, this simplifies to

$$\Delta\varphi = -\epsilon^0 \dot{\bar{\phi}} \quad \text{(linear order)}\,. \tag{6.25}$$

The same rules apply to tensors

$$\begin{aligned}
\Delta h_{\mu\nu}(x) &\equiv g'_{\mu\nu}(x) - g_{\mu\nu}(x) \\
&= g'_{\mu\nu}(x') - \epsilon^\lambda \partial_\lambda g_{\mu\nu}(x) - g_{\mu\nu}(x) \\
&= g_{\lambda\kappa}(x) \frac{\partial x^\lambda}{\partial x'^\mu} \frac{\partial x^\kappa}{\partial x'^\nu} - \epsilon^\lambda \partial_\lambda g_{\mu\nu}(x) - g_{\mu\nu}(x) \\
&= -g_{\lambda\mu} \partial_\nu \epsilon^\lambda - g_{\lambda\nu} \partial_\mu \epsilon^\lambda - \epsilon^\lambda \partial_\lambda g_{\mu\nu}(x) \\
&= -\nabla_\mu \epsilon_\nu - \nabla_\nu \epsilon_\mu = -2\nabla_{(\mu}\epsilon_{\nu)}\,.
\end{aligned} \tag{6.26}$$

In differential geometry, the above transformation are known as Lie derivatives (up to a sign). How do the SVT components transforms? Using Eq. (6.26) and the SVT decomposition Eq. (6.20), we find the following linear gauge transformations of the SVT components for the metric[42]

$$\begin{aligned}
\Delta A &= 2H\epsilon_0\,, \quad \Delta B = -\frac{2}{a^2}\epsilon^S\,, \\
\Delta C_i &= -\frac{1}{a^2}\epsilon_i^V\,, \quad \Delta\gamma_{ij} = 0\,, \quad \Delta h_{00} = -2\delta N = 2\dot{\epsilon}^0\,, \\
\Delta\psi &= \frac{1}{a^2}\left(-\epsilon_0 - \dot{\epsilon}^S + 2H\epsilon^S\right)\,, \quad \Delta N_i^V = -\dot{\epsilon}_i^V + 2H\epsilon_i^V\,,
\end{aligned} \tag{6.27}$$

[42]Notice that $\epsilon_0 = -\epsilon^0$.

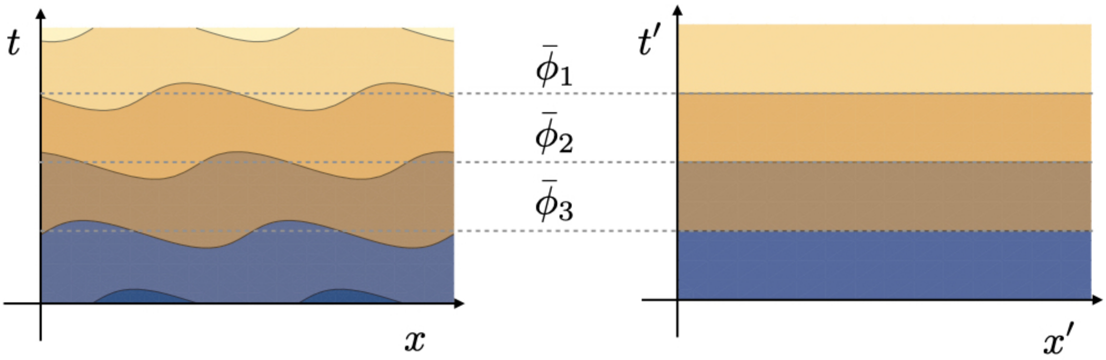


Figure 7: The figure depicts the idea that perturbations can be set to zero by a gauge transformation, namely a change of coordinates where the background is kept fixed. In the left panel we see that in coordinates x and t the lines of constant $\phi(t,x) = \bar{\phi}(t) + \delta\phi(t,x)$ (continuous lines separating areas of different color) are different from the lines of constant $\bar{\phi}$ (straight horizontal dashed gray lines). Hence there is a spatially varying, non-vanishing perturbation $\varphi(t,x)$. In the right panel, we have changed to coordinates x' and $t' \sim \phi(t,x)$. Now the constant ϕ lines coincide with the constant $\bar{\phi}$ lines, which are independent of x', and hence the perturbation $\delta\phi$ vanishes.

where we have SVT-decomposed gauge parameter

$$\epsilon^\mu = \{\epsilon^0, \partial^i \epsilon^S + \epsilon^i_V\}\,, \tag{6.28}$$

with $\partial_i \epsilon^i_V = 0$. It is important to notice that to derive these transformations from the general transformation of $h_{\mu\nu}$ in (6.26), one needs to use inverse Laplacians and this is a valid step only if we assume that $\epsilon^\mu(x)$ vanishes for $|\mathbf{x}| \to \infty$.
Two comments are in order. First, perturbations do *not* transform covariantly, rather they all shift by something linear in ϵ^μ. This means that by carefully choosing ϵ^μ we can set to zero some of the perturbations. In other words, we can choose a gauge (i.e. coordinates) such that A or B or h_{00} or some other component vanishes. This idea is depicted in Fig. 7. Second, unlike scalars and vectors, tensor perturbations γ_{ij} are gauge invariant to linear order. The intuitive reason is that the gauge parameter ϵ^μ has only a scalar and a vector component.

We now can proceed in two ways:

- We can *fix the gauge* and work in a particular set of coordinates, or
- we can work with *gauge-invariant variables*, namely specific combinations of the perturbations for which the gauge transformations cancel (e.g. (6.27) at linear order).

The gauge fixing approach is useful for explicit calculations, while the gauge-invariant variables are convenient to express the final result of a calculation. So in the following we will discuss both possibilities in Sec. 6.4 and Sec. 6.6.

6.4 Different gauges

Since vectors decay in cosmology, we will neglect them henceforth. The idea of fixing the gauge is to choose coordinates that correspond to the constant hypersurfaces of some of the perturbations, so that those perturbations appear constant. In other words, we can choose ϵ^0 and ϵ^S in Eq. (6.28) in such a way to cancel whatever profile of some of the scalar perturbations, using the transformation properties in (6.27), as depicted in Fig. 7. Notice that the gauge parameters ϵ^μ need to vanish at spatial infinity in the same way as the physical perturbations they need to cancel. In this sense these are *small gauge transformations*. Large gauge transformations will be discussed in Sec. 7.1. There are infinitely many choices of gauge, but we will discuss only four that are commonly used.

Newtonian gauge Using the gauge transformations in (6.27), we see that

$$\begin{cases} \epsilon^S = a^2 B/2 \\ \epsilon_0 = a^2\psi - \frac{a^2}{2}\dot{B} \end{cases} \quad \Rightarrow \quad \begin{cases} B' = B + \Delta B = B - B = 0 \\ \psi' = \psi + \Delta\psi = \psi - \psi = 0\,. \end{cases} \tag{6.29}$$

In a more compact form, we will simply write the gauge condition as

$$B = 0 \qquad \psi = 0\,. \tag{6.30}$$

Notice that these two conditions determine ϵ^0 and ϵ^S completely, so small scalar gauge transformations are fully fixed by these requirements. The scalar part of the metric has then only diagonal perturbations, namely in h_{00} and h_{ii}. Traditionally these perturbations are called Φ and Ψ and collectively referred to as *Newtonian potentials*. So, with the identification $h_{00} = -2\Phi$ and $A = -2\Psi$, we find [43]

$$ds^2 = -\left(1 + 2\Phi\right) dt^2 + a^2 dx^i dx^j \left[\left(1 - 2\Psi\right)\delta_{ij} + \gamma_{ij}\right] \quad \text{(Newtonian gauge)}\,. \tag{6.31}$$

This is the perturbed metric in Newtonian gauge. This is particularly useful in the study of the formation of Large Scale Structures.

Spatially-flat gauge In the study inflation, it is sometimes useful to choose coordinates such that the spatial part of the metric is free from any scalar perturbation,

$$A = B = 0\,. \tag{6.32}$$

Then

$$g_{00} = -1 + h_{00}\,, \qquad g_{0i} = N_i = a^2\partial_i\psi\,, \qquad g_{ij} = a^2\left(\delta_{ij} + \gamma_{ij}\right) \quad \text{(flat gauge)}\,, \tag{6.33}$$

which has only tensor perturbations. When tensors are neglected, this is just the metric of *flat* FLRW background, hence the name. Of course $h_{0\mu}$ does not vanish in this gauge and can be written in terms of the lapse and shift. Since these are fixed by constraints, in this gauge there are no dynamical degrees of freedom in the metric. The only dynamical scalar degrees of freedom is φ. We will shortly use this gauge to solve the constraints (6.16).

[43] Be aware that this is possibly the least universal convention in physics. You might find references where the definitions of Φ and Ψ as well as their signs are exchanged. Here we follow Weinberg's notation, which differ from Dodelson's notation by $\Phi_W = \Psi_D$ and $\Psi_W = -\Phi_D$.

Comoving gauge Another option, often employed to study inflation, is comoving gauge[44], sometime also called "ζ-gauge":

$$\varphi = 0 \quad \text{and} \quad B = 0\,. \tag{6.34}$$

In this gauge the metric takes the form

$$ds^2 = (-1 - 2\delta N)\, dt^2 + 2N_i dx^i dt + a^2 dx^i dx^j \left[\delta_{ij}(1 + A) + \gamma_{ij}\right]\,. \tag{6.35}$$

It is straightforward to check that $\varphi = 0$ fixes ϵ^0, while ϵ^S is completely fixed by the condition $B = 0$. This gauge was employed by Maldacena in his seminal paper on primordial non-Gaussianity [67].

Synchronous gauge* An alternative choice of gauge makes the temporal scalar part of the metric $h_{0\mu}$ vanish identically, namely one chooses ϵ^0 and ϵ^S such that

$$g_{00} = -1 \qquad\qquad g_{0i} = 0\,. \tag{6.36}$$

The perturbed metric takes the form

$$ds^2 = -dt^2 + a^2 dx^i dx^j \left[\delta_{ij}(1 + A) + \partial_i\partial_j B + \gamma_{ij}\right]\,. \tag{6.37}$$

This gauge is sometimes used in the study of the Cosmic Microwave Background.

6.5 Small gravitational interactions and the decoupling limit

We are finally in the position to perturbatively quantize general relativity in the presence of a scalar field and compute the contribution of dynamical gravity to cosmological correlators. Using the scalar vector tensor decomposition and a convenient choice of gauge we will solve perturbatively the constraints. In principle we should plug these solution into the action which would finally depend only on unconstrained dynamical field. However, we will show that the gravitational interactions implied by the solution of the constraints are always slow-roll suppressed and so in the regime of phenomenological interest they are smaller than the inflaton self interaction.

The constraints in (6.16) become manageable in perturbation theory. We claim[45] that to compute the action to cubic order it suffices to solve the constraints to linear order. Let's consider a toy model with only N. Adding N_i is a straightforward extension. We want to compute the Lagrangian $\mathcal{L}(h, \bar{N})$ where $\bar{N}$ is a solution of a constraint, namely

$$\frac{\delta\mathcal{L}}{\delta N}(h, \bar{N}, \bar{N}_i) = \frac{\partial\mathcal{L}}{\partial N} - \partial_i \frac{\partial\mathcal{L}}{\partial\partial_i N} = 0\,. \tag{6.38}$$

Imagine we found an n-th order solution $\bar{N} = N^{(n)}$ that solves the constraints up to corrections of order $(n+1)$. Now, write the *full action* and expand it around this solution in powers of $\bar{N} - N^{(n)}$,

$$\mathcal{L}(h, \bar{N}) = \mathcal{L}(h, N^{(n)}) + \left(\bar{N} - N^{(n)}\right)\frac{\delta\mathcal{L}}{\delta N} + \mathcal{O}\left(\left(\bar{N} - N^{(n)}\right)^2\right)\,. \tag{6.39}$$

[44] This is not the same as comoving orthogonal gauge, where one imposes $\psi = 0$ instead of $B = 0$.

[45] To leading order this was noticed in [67]. More generally, the n^{th} order solution of the constraints is sufficient to obtain the $(2n+1)^{\text{th}}$ order action (see App A.3 of [73]).

We recognise that the coefficient of $\left(\bar{N} - N^{(n)}\right)$ is just the constraint equation evaluated on the n-th order solution, which is at least order $n+1$. We conclude that $\mathcal{L}(h, N^{(n)})$ differs from the full action by terms of order $2(n+1)$. In our case, we solve the constraints at linear order, so $n = 1$ and the action we obtained is incorrect only at order $2(1+1) = 4$. This proves the claim.

Flat gauge and the decoupling limit Scalars, vectors and tensors decouple at linear order. In solving the constraints to linear order we can therefore focus on scalars and set all other perturbations to zero. It is convenient to work in flat gauge (6.33), where we find the linear order relations

$$ {}^{(3)}R \simeq 0\,, \qquad N \simeq 1 + \delta N\,, \qquad N_i \simeq a^2 \partial_i \psi\,, \tag{6.40} $$

$$ N K_{ij} \simeq a^2 \left(H\delta_{ij} - \partial_i \partial_j \psi\right)\,, \qquad N K \simeq 3H - \partial_i \partial_i \psi\,. \tag{6.41} $$

Using these to expand the Hamiltonian constraint in (6.16) to linear order we find

$$ \left.\frac{\delta S}{\delta N}\right|_0 \propto 3H^2 M_{\rm Pl}^2 + P - 2XP_{,X} = 0\,, \tag{6.42} $$

$$ \left.\frac{\delta S}{\delta N}\right|_1 \propto -2HM_{\rm Pl}^2 \partial_{ii}\psi + \varphi\left(P_{,\phi} - 2XP_{,X\phi}\right) + (P_{,X} + 2XP_{,XX})\delta X = 0\,, \tag{6.43} $$

where to first order

$$ \delta X = \frac{1}{2}\dot{\phi}^2 (1 - 2\delta N) + \frac{1}{2}\dot{\phi}\dot{\varphi} + \dots\,. \tag{6.44} $$

The momentum constraint in (6.16) to linear order becomes (it's trivial to zeroth order)

$$ \left.\frac{\delta S}{\delta N^i}\right|_1 \propto \partial_i \left[2HM_{\rm Pl}^2 \delta N - P_{,X}\dot{\phi}\dot{\varphi}\right] = 0\,. \tag{6.45} $$

This can be straightforwardly solved for δN

$$ \delta N = \frac{\dot{\phi}}{2HM_{\rm Pl}^2}\varphi = \frac{\epsilon H}{\dot{\phi}}\varphi\,. \tag{6.46} $$

where we used repeatedly the background equations of motion, (1.61) and (1.62). Using this in the Hamiltonian constraint we can find

$$ \partial_i \partial_i \psi = -\frac{\epsilon}{c_s^2}\partial_t \left(\frac{H\varphi}{\dot{\phi}}\right)\,, . \tag{6.47} $$

These equations tell us how spacetime is deformed by the presence of the scalar field perturbations φ, a phenomenon that is called *backreaction*. The explicit solution (6.47) shows that backreaction is suppressed by the first slow-roll parameter $\epsilon \ll 1$. One can keep these terms and substitute them into the action, (6.13) and (6.14), and see what cubic interactions are generated. This was first done by Maldacena in the seminal paper [67]. As one might expect, the size of the resulting interactions is suppressed by ϵ. This leads to small effects that are beyond observational reach, for at least the next century. Conversely, recall from Sec. 4.2 that in $P(X, \phi)$ theories we can make the interactions large by choose a function P with appropriately large derivatives with respect to X (see (4.28)). Hence in the following, we neglect these slow-roll suppressed terms.

We will see in Sec. 7.4 that they can be partially recovered from soft theorems.

Finally, notice that if we are only interested in computing the bispectrum of φ, we can also neglect the tensor perturbations in the metric (6.33). The reason is that tensors don't mix with scalars at linear order. Interactions between φ and γ_{ij} only appear in the cubic action (or higher), e.g. in the form $\gamma_{ij}\partial_i\varphi\partial_j\varphi$. But these terms cannot contribute at tree level to the bispectrum of φ.

6.6 Gauge invariant variables: curvature perturbations

Instead of choosing a specific set of coordinates one can work with gauge invariant variables[46]. This is not particularly convenient during the calculation, but it is a useful way to present the final result.

The idea is to find combinations of variables whose gauge transformations cancel each other. There are infinitely many options. For example, it is easy to see from (6.27) that the following combinations are gauge invariant at linear order

$$\begin{aligned} &\gamma_{ij} \ , && \Phi_i = N_i^{(V)} - \frac{a^2}{2}\dot{C}_i \ , \\ &-2\Psi = A + Ha^2\left(2\psi - \dot{B}\right) \ , && -2\Phi = h_{00} - \partial_t\left(a^2\left(2\psi - \dot{B}\right)\right) \ . \end{aligned} \tag{6.48}$$

These are called the *Bardeen variables* and were introduced in [15]. In the inflationary literature it is more common to use alternative gauge invariant variables.

Curvature perturbations Let's introduce the new combination[47]

$$\mathcal{R} \equiv \frac{A}{2} - \frac{H}{\dot{\phi}}\varphi \,, \tag{6.49}$$

From the gauge transformations in (6.27) it is straightforward to check that $\mathcal{R}$ is gauge invariant at linear order. We will refer to $\mathcal{R}$ as *curvature perturbations on comoving hypersurfaces*. To understand this name, notice that in comoving gauge we have $\varphi = 0$ from (6.34) and so $\mathcal{R} = A/2$. Then $\mathcal{R}$ appears in the metric as

$$g_{ij} = a^2 dx^i \delta_{ij} dx^j (1 + 2\mathcal{R}) \quad \text{(comoving gauge)} \,. \tag{6.50}$$

Then the Ricci scalar of a spatial hypersurface is ${}^{(3)}R = -4\partial_i^2\mathcal{R}$ and so $\mathcal{R}$ generates a position dependent curvature. Nomen est omen.

[46] By "gauge invariant" here we have mean something distinct and weaker than what is meant in discussions of quantum gravity. Here a variable is "gauge invariant" if it does not transform at linear or higher order in perturbation theory under a change of coordinates *after the new coordinate is identified with the old one* (see Sec. 6.3). When discussing quantum gravity, "gauge invariant" has a stronger meaning and requires an observable to not transform at all as one changes coordinates. In this stronger sense a local scalar field operator is not gauge invariant, since $\phi(x) \to \phi(x - \epsilon)$. Hence there is no contradiction between our discussion here and the often stated fact that there are no local and gauge-invariant observables in quantum gravity.

[47] Unfortunately, different conventions for the names of these variables exists. For example, in [67], $\mathcal{R}$ is called ζ. This has produces a schism in the subsequent literature. A useful summary of the many possible choices in the literature is given in App A of [98].

Since $\mathcal{R}$ is gauge invariant, we can and will compute it in any gauge we want. In particular, in flat gauge $A = 0$, and so

$$\mathcal{R} = -\frac{H}{\dot{\bar{\phi}}}\varphi \quad \text{(flat gauge)}\,. \tag{6.51}$$

This relation teaches us something interesting. Notice that $H/\dot{\bar{\phi}}$ is a time dependent function. Its time derivative is slow-roll suppressed but it is not zero. This implies that $\mathcal{R}$ and φ cannot be both constant in time. In the next section we will prove a very general theorem stating that $\mathcal{R}$ is constant on superHubble scales and so φ must evolve. This is another important reason why the predictions of the early universe should always be computed in terms of superHubble correlators of $\mathcal{R}$, which is gauge invariant and conserved, rather than those of ϕ, which evolves with time and changes from gauge to gauge.

Beyond linear order Since much of the goal of these notes is to study cosmological perturbation beyond linear order, it would be nice to define a variable that is gauge invariant not just to linear order. Indeed, a tree-level n-point function of $\mathcal{R}$ is gauge invariant iff $\mathcal{R}$ is gauge invariant to at least order $n-1$. With a bit of work we could find a second-order version of (6.51), but this involves a lot of algebra as we have to recompute all the gauge transformations to second order. Instead, we'll try to be clever. We'll define the gauge-invariant variable $\mathcal{R}$ to be the quantity that in comoving gauge appears in the metric as[48]

$$g_{ij} = a^2 e^{2\mathcal{R}}\delta_{ij} \quad \text{(comoving gauge)}\,. \tag{6.52}$$

In other gauges, $\mathcal{R}$ is given by its value in comoving gauge plus all the terms induced by the gauge transformation to the required order. This agrees with the previous definition (6.49) at linear, because $\varphi = 0$ and $e^{2\mathcal{R}} = 1 + A$.

The non-linear definition of $\mathcal{R}$ lends itself to a nice interpretation, which is depicted in Fig 8. In particular, we can think of the spatial variation of $\mathcal{R}(\mathbf{x})$ as modulating the local amount of expansion compared to the expansion $a(t)$ prescribed by the background. For example, in de Sitter a point $\mathbf{x}$ experience a local amount of expansion corresponding to a number of efoldings

$$N_e = \log a(t)e^{\mathcal{R}(\mathbf{x})} = Ht + \mathcal{R}(\mathbf{x})\,. \tag{6.53}$$

This mean that the number of efoldings of expansion in a perturbed universe depends on the point, $N_e = N_e(\mathbf{x}, t)$. Then one can define perturbations to N_e as usual by subtracting the average $\delta N = N_e - \bar{N}_e$, where $\bar{N}_e$ is the spatial average of N_e. In an appropriate gauge, $\mathcal{R}$ coincides with δN to all orders. This is the starting point of the so-called δN formalism that describes the evolution of superHubble perturbations of $\mathcal{R}$ in an expansion in spatial derivatives [82, 83].
Let's see what $\mathcal{R}$ looks like at the next order: quadratic. The manifestly gauge invariant expression is messy. For our purpose it will be sufficient to see how $\mathcal{R}$ is related to φ at order $\mathcal{O}(\varphi^2)$. To find this relation we have to change the gauge from comoving to flat.

[48] This assumes we set tensors to zero, $\gamma_{ij} = 0$. Later one we will see that γ_{ij} can be included by the substitution $\delta_{ij} \to \exp(\gamma_{ij})$.

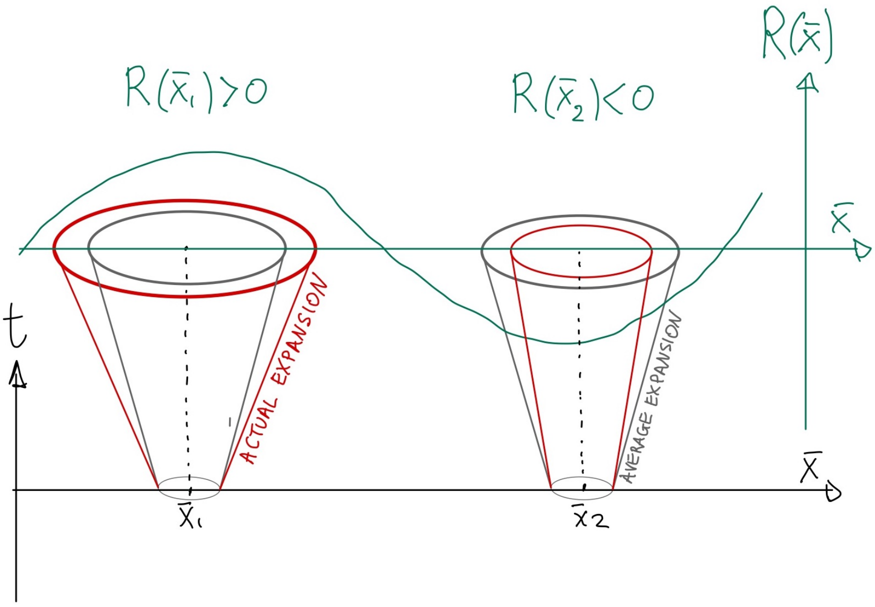


Figure 8: The figure shows the intuitive meaning of $\mathcal{R}(\vec{x})$ perturbations in comoving gauge. The unperturbed homogeneous universe expands in time (along the vertical axis) by the same amount at every point (the "average expansion" is represented by gray lines). But in the perturbed universe, points with different values of $\mathcal{R}(\vec{x})$ experience a larger (smaller) amount of expansion if $\mathcal{R}(x) > 0$ ($\mathcal{R}(x) < 0$) (the "actual expansion" is represented by red lines), as indicated in (6.53).

The calculation is conceptually the same as at linear order but it becomes algebraically more involved. The final result is (see app A of [67] for details)

$$\mathcal{R} = -H\frac{\varphi}{\dot{\bar{\phi}}} + \frac{H\dot{\varphi}\varphi}{\dot{\bar{\phi}}^2} + \frac{1}{2}\frac{\varphi^2}{\dot{\bar{\phi}}^2}\left(\dot{H} - \frac{H\ddot{\bar{\phi}}}{M_{\rm Pl}^2}\right) - \left(1 - \frac{\partial_i\partial_j}{\partial^2}\right)\left[\frac{\partial_i\varphi\partial_i\varphi}{4a^2\dot{\bar{\phi}}^2} - \frac{1}{2}\frac{\partial_i\psi\partial_i\varphi}{\dot{\bar{\phi}}}\right].$$

Remarkably, all quadratic terms turn out to be small. The term $\dot{\varphi}\varphi$ decays on superHubble scales because φ becomes approximately constant, up to slow-roll corrections. The term $(\partial_i\varphi)^2$ decays as $a^{-2} \propto \tau^2$. The term $\partial_i\varphi\partial_i\psi$ is slow-roll suppressed because $\partial_i\psi$ is given by (6.47). Finally, the term φ^2 is also slow-roll suppressed because $\dot{H} = -H^2\epsilon$ and $\ddot{\bar{\phi}} \propto \eta$.

In summary, up to small slow-roll corrections, gauge invariant perturbations $\mathcal{R}$ are linearly related to scalar field perturbations φ in flat gauge by the simple rescaling in (6.51). We will use this to express our results for the correlators of φ into correlators of $\mathcal{R}$.

6.7 Primordial non-Gaussianity from self interactions

Let's summarize where stand: all interactions induced by gravity are slow-roll suppressed while there is a regime in which the interactions implied by the scalar action $P(X, \phi)$ (see Sec. 4.2) are much larger. We realized in Sec. 6.3 that φ is not gauge invariant, and so neither are its power spectrum $P(k)$ in (2.31), nor its bispectrum B_3 in (4.28). To conclude our discussion we would like to write down the correlators computed from

$P(X, \phi)$ in terms of the gauge invariant variable $\mathcal{R}$.

Recall that the scalar perturbations φ are related to a canonically normalized scalar φ_c by the rescaling $\varphi = \varphi_c/\sqrt{P_{,X}}$. Using this, the power spectrum for a canonical scalar in dS, (2.31), and (6.51), we can compute the power spectrum of $\mathcal{R}$:

$$P_{\mathcal{R}}(k) = \frac{H^2}{\dot{\phi}^2 P_{,X}} P_{\varphi_c}(k) = \frac{1}{4\epsilon c_s}\left(\frac{H}{M_{\rm Pl}}\right)^2 \frac{1}{k^3} \,. \tag{6.54}$$

This an important result. It tells us that the measured amplitude of primordial perturbations

$$\Delta^2_{\mathcal{R}} \equiv \frac{k^3 P_{\mathcal{R}}(k)}{2\pi^2} = \frac{1}{8\pi^2 \epsilon c_s}\left(\frac{H}{M_{\rm Pl}}\right)^2 = 3.047 \pm 0.014 \quad (68\% \text{ CL}) \tag{6.55}$$

is a measurement of the scale of inflation H in Planck units, divided by ϵ and c_s.
To translate the bispectrum of φ into the bispectrum of $\mathcal{R}$ we need to find the relation between these two variables to second order. We implicit defined the gauge invariant $\mathcal{R}$ by specifying it in comoving gauge to all orders in
This means that, up to decaying terms and terms suppressed by slow-roll corrections, we can simply use the linear order relation (6.51) and write

$$\langle \mathcal{R}(\mathbf{k}_1)\mathcal{R}(\mathbf{k}_2)\mathcal{R}(\mathbf{k}_3)\rangle \simeq -\left(\frac{H}{\dot{\phi}}\right)^3 \langle \varphi(\mathbf{k}_1)\varphi(\mathbf{k}_2)\varphi(\mathbf{k}_3)\rangle + \mathcal{O}(\epsilon, \eta, \dots) \tag{6.56}$$

for $\tau \to 0$, where the bispectra of φ were given in (4.28). It is interesting to notice that for a canonical inflaton, with $P = X - V$, all scalar self-interactions in (4.26) vanish except for $P_{,\phi\phi\phi} = -V'''$. This interaction is slow-roll suppressed since $V''' \sim \xi_3$. Since also gravitational interactions are slow-roll suppressed, we conclude that *primordial non-Gaussianity from canonical-field inflation are expected to be very small*, of order $\mathcal{O}(\epsilon, \eta)$. Since we haven't seen any evidence for primordial non-Gaussianity, canonical single-field inflation is still a very successful model for the early universe.

7 Symmetries and soft theorems

ref

So far we focus on computing the correlators directly for a given class of theories. But there is also much we can learn about correlators from the symmetries of the problem alone. This model independent approach leads to some very powerful results. First we will discuss the concept of adiabatic modes and then use it to derive soft theorems that dictate the behavior of cosmological correlators when one of the momenta is soft, namely much smaller than the others.

7.1 Adiabatic modes

We found some gauge invariant predictions for primordial correlations from inflation to leading order in slow roll. To see how these manifest themselves in observables we need to evolve them in time until today. The problem though is that we don't know the constituents of the universe at energies much larger than those probed at colliders, say above 10 TeV. So we don't even know what the right equations to solve are. Luckily for us, under very general conditions $\mathcal{R}$, γ_{ij} and their correlators are conserved in time (see e.g. [66, 82]). This result, which we will prove in this section, is one of the most important in cosmology: It tells us that we can use the sub-eV photons of the CMB to learn something about the laws of physics tens of orders of magnitude higher. This remarkable connection of low-energy observables to high-energy physics has been a tremendous drive for the field of cosmology and has open new possibility to probe fundamental physics.

We are now ready to state an important theorem [100]:

Whatever the constituents of the universe and outside the sound horizon, $c_s k \ll aH$, there is always at least[49] *one conserved scalar "adiabatic" mode, i.e. $\dot{\mathcal{R}} = 0$ and one conserved tensor mode, i.e. $\dot{\gamma}_{ij} = 0$.*

This theorem is valid to all orders in perturbation theory around a flat FLRW spacetime, but we will prove it only at linear order. Also, the theorem applies to gravity coupled to a $P(X, \phi)$ theory, but also holds much more generally. So, in this section, we consider a general matter sector, which is described by a generic energy-momentum tensor $T_{\mu\nu}$, with SVT decomposition

$$\begin{aligned} \delta T_{00} &= -\bar{\rho} h_{00} + \delta\rho \,, \\ \delta T_{i0} &= \bar{p} h_{0i} - (\bar{\rho} + \bar{p}) \left[\partial_i \delta u + \delta u_i^V\right] \,, \\ \delta T_{ij} &= \bar{p} h_{ij} + a^2 \left[\delta_{ij}\delta p + \partial_{ij}\pi_{ij}^S + \partial_{(i}\pi_{j)}^V + \pi_{ij}^T\right] \,, \end{aligned} \tag{7.1}$$

where π^S, π_i^V and π_{ij}^T are known as *anisotropic inertia* and depend on the substance under consideration. For example, all anisotropic inertia vanishes for a scalar field or for a perfect fluid (see (1.12)). In the above decomposition, we recognize four scalars ($\delta\rho$, δp, δu and π^S), two transverse vectors ($\partial_i \pi_i^V = 0 = \partial_i \delta u_i^V$) and one transverse traceless tensor ($\pi_{ii}^T = \partial_i \pi_{ij}^T = 0$), adding up again to 10 components. Notice that we SVT-decomposed the fluid velocity with a *lower* index:

$$u_\mu = \{-1 + \delta u_0, \partial_i \delta u + \delta u_i^V\} \,, \tag{7.2}$$

Our scalar field can be written in this language using the identifications (1.60).

Proof The general strategy to prove the result consists of the following two steps (summarized in Fig. 9):

- Start from an unperturbed universe (horizontal black line in Fig. 9). Perform a large gauge transformation that diverges at spatial infinity and that generates non-vanishing perturbations (diagonal blue line in Fig. 9). These are still a solution to all equations of motion by general covariance. The new solution is just an unperturbed universe written in awkward coordinates.
- Under certain conditions, the above solution can be deformed by an appropriately small amount to vanish at spatial infinity. In Fourier space, the deformation is small in the sense that it is $\mathcal{O}(k^2)$ with $c_s k \ll aH$. The deformed solution describes

[49] Actually one can prove the existence of many other decaying modes, including vector modes, see [31, 52, 72] for details.

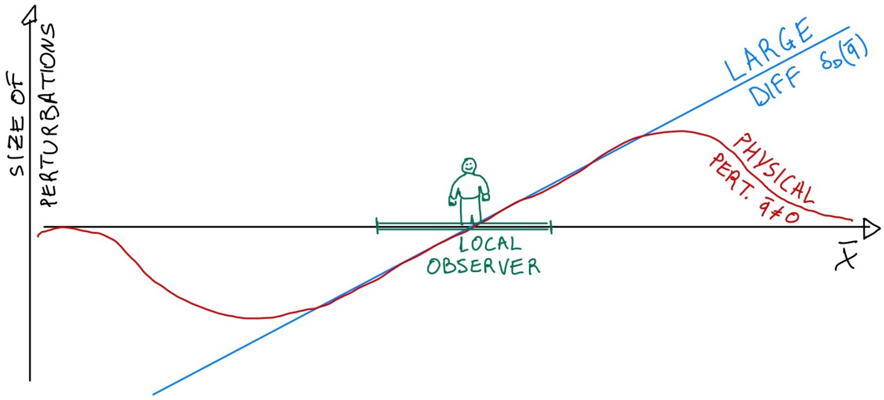


Figure 9: The figure summarized the construction of adiabatic modes. We start from an unperturbed FLRW spacetime a perform a large diff. This generates perturbations to the metric and matter fields (blue line) that solve Einstein's equations but do not vanish at spatial infinity. We can find some physical modes (red line) that locally, i.e. nearby some observer's location, look the same as the large diff but do vanish at $|\vec{x}| \to \infty$. These are adiabatic modes, namely physical solutions that, to the best of my drawing abilities, are locally indistinguishable from a change of coordinates.

a new configuration of the system, which corresponds to a perturbed universe with perturbations that vanishes at spatial infinity (wavy red line in Fig. 9). In position space, the deformation is small in the sense that the large diff and the deformed physical solution coincide locally, nearby some observer, but differ further away

We will prove the theorem working in comoving gauge, as in [31, 52, 72]. A derivation in Newtonian and synchronous gauge was presented in the original paper [100]. In this context, comoving gauge is defined by the gauge condition

$$\delta u = 0\,, \qquad\qquad B = 0\,, \tag{7.3}$$

with the metric given in (6.35). With the identification $\delta u = \varphi/\dot{\bar{\phi}}$, this is coincides with our definition of comoving gauge in Sec. 6.4, namely $\varphi = 0 = B$.
Consider the change of coordinates

$$\epsilon^\mu = \left\{0, \omega_{ij}(t)x^j\right\}\,, \tag{7.4}$$

with ω_{ij} some time-dependent 3×3 symmetric matrix, $\omega_{ij} = \omega_{ji}$. We have chosen $\epsilon^0 = 0$ so that we don't spoil the gauge condition $\delta u = 0$. Notice that ϵ^μ doesn't vanish at spatial infinity. Therefore, the transformation of $h_{\mu\nu}$ is still given by (6.26) but the transformations of the SVT components (6.27) cannot be used, since they were derived under the assumption that ϵ vanishes at spatial infinity. If we start from an unperturbed, flat FLRW universe, after this gauge transformation we find some non-trivial perturbations given by

$$h_{00} = -2\delta N = 0\,, \qquad\qquad \mathcal{R} = \frac{A}{2} = -\frac{1}{3}\omega_{ii}\,,$$

$$N^i = \partial_i\psi + N^i_V = -\dot{\omega}_{ij}x^j \qquad\qquad \psi = f(t) - \frac{1}{6}\dot{\omega}_{kk}x^i x^j \delta_{ij} \tag{7.5}$$

$$N^i_V = -\dot{\omega}_{<ij>}x^j \qquad\qquad \gamma_{ij} = -2\omega_{<ij>}\,, \tag{7.6}$$

where $< .. >$ indicated the symmetric traceless part

$$A_{<ij>} \equiv \frac{1}{2}\left(A_{ij} + A_{ji}\right) - \frac{1}{3}\delta_{ij} A_{kk}\,, \tag{7.7}$$

and $f(t)$ is an arbitrary time-dependent integration constant. All other perturbations $\{\delta\rho, \delta p, \delta u, B, \pi^{S,V,T}\}$ vanish and so does $\delta T_{\mu\nu}$. Notice that since we still have $B = 0 = \delta u$ after the gauge transformation, so we are still in comoving gauge. The above transformations are completely different from those valid for small gauge transformations, Eq. (6.27). For example, the tensor perturbations γ_{ij} now do change. What do the perturbations in (7.6) represent? Since GR is a covariant theory and we started from and unperturbed FLRW, which is a solution of GR, the perturbations in Eq. (7.6) must also be solution. But because ϵ^{μ} did not vanish at spatial infinity, this solution is an unusual one: perturbations are constant in space and don't vanish at spatial infinity either. This is depicted in the blue line in Fig. 9. In fact, this solution is just an unperturbed FLRW written in silly coordinates!

The clever insight of Weinberg is to ask when the above solution can be *extended to a physical solution*, with perturbations that do vanish at spatial infinity and can hence arise dynamically (see red line in Fig 9). To answer this, it's easiest to work in Fourier space, where the perturbations in Eq. (7.6), being all constant or power-law in $\mathbf{x}$, are proportional to $\delta_D(\vec{k})$ or its derivative. In particular they have support only at $\mathbf{k} = 0$. Any physical perturbation must vanish at spatial infinity and so its Fourier transform must be continuous at $\mathbf{k} = 0$. So, any non-vanishing perturbation in Fourier space must have support on $\mathbf{k} \neq 0$ as well. When $\mathbf{k} \neq 0$, we are not guaranteed that Eq. (7.6) is still a solution. We have to check. For those equations of motion that do not have an overall factor of $\mathbf{k}$, Eq. (7.6) is still a solution up to an arbitrary small correction. For example, for the tensor perturbations,

$$\ddot{\gamma}_{ij} + 3H\dot{\gamma}_{ij} + \frac{k^2}{a^2}\gamma_{ij} = 0\,. \tag{7.8}$$

The solution with $k^2 = 0$ and with k^2 small but non-vanishing are very similar: they differ only at order $k^2/(Ha)^2$, which can be made arbitrarily small for superHubble perturbations, $k \ll Ha$. Then, by continuity we know that physical solutions $\gamma_{ij}(k,t)$ exist, which in the limit $k \to 0$ look like (7.6), namely are constant in time, up to $\mathcal{O}(k^2)$ corrections. On the other hand, the extension to a physical, non-constant solution can be obstructed when some equations of motion vanish identically for $\mathbf{k} = 0$. For example, the ii- and $0i$-components of Einstein's equation in comoving gauge and Fourier space are

$$k_i k_j \left(\delta N + \mathcal{R} + \dot{\psi} + H\psi\right) = 0\,, \tag{7.9}$$

$$k_i \left(H\delta N - \dot{\mathcal{R}}\right) = 0\,, \tag{7.10}$$

$$k_i \left(\dot{N}^V_j + HN^V_j\right) = 0\,. \tag{7.11}$$

For the solution in (7.6), these equations were trivially solved because $k_i = 0$. But if we want physical solutions with $\mathbf{k} \neq 0$, we need to impose that these equations are non-trivially solved. Let us first focus on the trace part ω_{kk} of ω_{ij}, which appear only in the first two of the above equations. The "physicality condition" that (7.10) is satisfied for $k_i \neq 0$ implies

$$\dot{\mathcal{R}} = H\delta N = 0 \quad \Rightarrow \quad \mathcal{R} = -\frac{1}{3}\omega_{kk} = \text{const}\,, \tag{7.12}$$

which in turn gives the solution of (7.9) as

$$\psi = \frac{\omega_{kk}}{3a}\int^t dt' a(t')\,. \tag{7.13}$$

We conclude that a physical solution with $\mathcal{R}$ non-vanishing and constant must always exist in the superHubble limit as consequence of diffeomorphism invariance. As mentioned previously, this is the reason why we want to give the predictions of inflation in terms of $\mathcal{R}$, rather than other fields, such as φ, which continue to evolve in time.
We can also look at the constraint imposed by (7.11) on the traceless part $\omega_{<ij>}$ of ω_{ij}:

$$\ddot{\omega}_{<ij>} + 3H\dot{\omega}_{<ij>} = 0 \quad \Rightarrow \quad \omega_{<ij>} = \bar{\omega}^{(1)}_{<ij>} + \bar{\omega}^{2}_{<ij>}\int^t \frac{dt'}{a(t)^3}\,. \tag{7.14}$$

where $\bar{\omega}^{(1,2)}_{<ij>}$ are integration constant traceless matrices. Notice that the above equation is precisely the same as that of tensor modes in (7.8) for $k \to 0$. We conclude that, whatever the constituents of the universe, there is always a solution to the equations of motion with a constant, non-vanishing $\gamma_{ij} = \bar{\omega}^{(1)}_{<ij>}$, up to corrections that vanish in the superHubble limit. This solution represent the conservation of superHubble *primordial gravitational waves*. This conservation gives us a unique opportunity to use measurements of the late universe, such the CMB, to probe GR in the early universe and its perturbative quantization. The second solution $\bar{\omega}^{(2)}_{<ij>}$ is also general, but it decays with time, so it is not very relevant observationally.

To summarize, we have demonstrated the existence of adiabatic modes, namely physical solutions (which vanish at spatial infinity) that are locally indistinguishable from a change of coordinates (see Fig 9). In Sec. 7.4 we will see that the existence of adiabatic modes implies the existence of a symmetry for perturbations, which in turn will lead to soft theorems. Before we get there, let us remind ourselves of the role of symmetry in field theory.

7.2 Symmetry symmetry symmetry

In this section we recall some basic facts about the role of symmetries in field theory.

Recall that symmetries in field theory are transformations $\Delta\phi$ of the fields ϕ (used in this section to denote collectively fields of any spin) that leave the action invariant, or equivalently that change the Lagrangian by a total derivative

$$\Delta\mathcal{L} = \partial_\mu F^\mu\,. \tag{7.15}$$

What symmetries do for a living is to take some solution ϕ_{sol} of the dynamics and generate another, different one $\phi'_{sol} = \phi_{sol} + \Delta\phi_{sol}$. If one imposes that two states that differ by a symmetry transformation are the same physical state, i.e. all observables give precisely the same values in both states, then the symmetry is called a *gauge symmetry*. A familiar example is electrodynamics, where A^μ and $A^\mu + \partial_\mu\alpha$ represent the same

$$
\begin{tikzcd}
A(t1) \arrow[rr, leftrightarrow, "H"] \arrow[d, leftrightarrow, "Q"'] & & A(t2) \arrow[d, leftrightarrow, "Q"] \\
B(t1) \arrow[rr, leftrightarrow, "H"] & & B(t2)
\end{tikzcd}
$$

Figure 10: The equivalence of two definitions of symmetry: a transformation that generates new solutions and a transformation that commutes with the Hamiltonian H. Some solution A at time t_1 can be evolved to time t_2 and then transformed by Q into $B(t_2) \neq A(t_2)$. This gives the same result as first transforming to $B(t_1)$ and then evolving because $[Q, H] = 0$. By doing this process at every time t from a solution $A(t)$ one can the generate a new solution $B(t)$.

physical state[50] (with appropriate boundary conditions on α). If ϕ_{sol} and ϕ'_{sol} are physically distinguishable, the transformation is called a *global symmetry*. In the following I'll focus on global symmetries unless otherwise stated.

The fact that Q generates new solutions is equivalent to saying that symmetries commute with the Hamiltonian $[Q, H] = 0$ and so the diagram in Fig. 10 commutes. By Nöther theorem there exists a conserved current

$$J^\mu = \frac{\delta \mathcal{L}}{\delta \partial_\mu \phi} \Delta\phi - F^\mu \quad \text{with} \quad \partial_\mu J^\mu = 0\,. \tag{7.16}$$

If you wish, you can make things look more covariant by defining $\tilde{J}^\mu \equiv J^\mu (-g)^{-1/2}$ so then $\nabla_\mu \tilde{J}^\mu = (\partial_\mu J^\mu)\,(-g)^{-1/2} = 0$. If the current vanishes sufficiently fast at spatial infinity, then one can define a conserved current Q by

$$Q \equiv \int \sqrt{h} J^\mu n_\mu d^3x\,, \tag{7.17}$$

where n^μ is a time-like vector field that defines some "constant-time" hypersurface over which we integrate. The conservation of J^μ implies $\dot{Q} \equiv n^\mu \partial_\mu Q = 0$. What Q does for a living is to generate the transformations of the fields from which it originally was derived through Nöther theorem:

$$i[Q, \phi] = \Delta\phi\,. \tag{7.18}$$

Since Q is Hermitian, $Q = Q^\dagger$, we can exponentiate this generator to define a unitary symmetry operator

$$\text{Finite unitary transformation:} \quad U \equiv e^{i\alpha Q}\,, \tag{7.19}$$

[50] Often gauge symmetries have parameters that are functions of spacetime as e.g. $\alpha(x)$ in electrodynamics. But this does not have to be the case in general. For example, consider a quantum mechanical particle on a circle of length L. We can describe the system using $x \in \{0, L\}$ but it is sometimes convenient to use the variable $x \in \{-\infty, +\infty\}$ with the identification $x \approx x + nL$ with $n \in \mathbb{N}$. The transformation $x \to x + nL$ is a gauge symmetry even if n is not time dependent.

for some parameter α of the transformation. We say that the symmetry generated by Q is *unbroken* in the state $|\Omega\rangle$ iff

$$\text{Unbroken symmetry:} \quad \langle\Omega|\,[Q,\phi]\,|\Omega\rangle = 0\,, \tag{7.20}$$

otherwise it is spontaneously-broken

$$\text{Spontaneously-broken symmetry:} \quad \langle\Omega|\,[Q,\phi]\,|\Omega\rangle \neq 0\,. \tag{7.21}$$

In words, the laws of nature are invariant under a given symmetry (i.e. $[Q,H]=0$), but the solution of those laws is not ($U\,|\Omega\rangle \neq |\Omega\rangle$). If Q annihilates $|\Omega\rangle$, namely $Q\,|\Omega\rangle = 0$, so that $U\,|\Omega\rangle = |\Omega\rangle$ then Q is unbroken. Conversely, if Q is *spontaneously broken*[51], $|\Omega\rangle$ is not invariant under Q namely $Q\,|\Omega\rangle \neq 0$. For spontaneously-broken symmetries $\Delta\phi$ must contain a constant term, when expanded in power of ϕ (assuming we are working with fields such that $\langle\phi\rangle = 0$, which is always true up to a field redefinition $\phi \to \phi - \langle\phi\rangle$). So *a spontaneously broken symmetry must always be non-linearly realized*[52]:

$$\text{Non-linearly realized symmetry:} \quad i[Q,\phi] = \Delta\phi = \text{const} + \mathcal{O}(\phi)\,. \tag{7.22}$$

7.3 Correlators and linearly-realized symmetries

In this section, we discuss the observational consequences of linearly realized symmetries in cosmology. We will focus on spacetime symmetries and discuss translations and rotations for FLRW spacetime and then dilations and special conformal transformations for de Sitter spacetime.

FLRW: translations and rotations If we assume a Lorentz-invariant theory and expand around a flat FLRW background, all primordial correlators must be invariant under translations and rotations. To see this more formally, consider the generators of spatial translations P^i and spatial rotations L^i, acting on scalar[53] operators

$$i[P^i, \phi(\mathbf{x})] = -\partial_i \phi(\mathbf{x})\,, \tag{7.24}$$

$$i[L^i, \phi(\mathbf{x})] = -\epsilon^{ijk} x_j \partial_k \phi(\mathbf{x})\,. \tag{7.25}$$

If these generators commute with the Hamiltonian then the same expressions hold for the Heisenberg operators at any time. These generators exponentiate to finite translations and rotations as in

$$U^{-1}(\vec{\alpha},\vec{\omega})\phi(\mathbf{x})U(\vec{\alpha},\vec{\omega}) = \phi(R^{ij}x^j + \alpha_i)\,, \tag{7.26}$$

with

$$R_{ij} = \exp\left(\epsilon_{ijk}\omega^k\right)\,, \qquad U(\vec{\alpha},\vec{\omega}) = \exp\left(iP^i\alpha_i\right)\exp\left(iL^i\omega_i\right)\,. \tag{7.27}$$

[51]This should not be confused with *explicit symmetry breaking*, which describes a situation in which the transformation is just not a symmetry anymore.

[52]To avoid confusion, let us stress that the commutator is a linear operation on ϕ and so $[Q,\lambda\phi] = \lambda\Delta\phi$ for any constant λ. By "non-linearly realized" we mean that the transformation acts non-linearly on the solutions of the theory, namely given two solutions $\phi_{sol,1} = \lambda\phi_{sol,2}$ one finds $\Delta\phi_{sol,1} \neq \Delta\phi_{sol,2}$.

[53]For generic operators with spin, the action or rotations is simply replaced by

$$i[L^i, \mathcal{O}_S^A(\mathbf{x})] = -D(L)^A_B \epsilon^{ijk} x_j \partial_k \mathcal{O}_S^B(\mathbf{x})\,, \tag{7.23}$$

where $D(L)^A_B$ is the representation of the algebra $\mathfrak{so}(3)$ relevant for $\mathcal{O}$.

and $U^\dagger U = 1$. Then we see that

$$\langle\Omega|\prod_a \phi(x_a)|\Omega\rangle = \langle\Omega| UU^{-1}\phi(x_1)UU^{-1}\phi(x_2)\dots\phi(x_n)UU^{-1}|\Omega\rangle \tag{7.28}$$

$$= \langle\Omega| U^{-1}\phi(x_1)UU^{-1}\phi(x_2)\dots\phi(x_n)U|\Omega\rangle \tag{7.29}$$

$$= \langle\Omega|\prod_a \phi(t_a, R\mathbf{x}_a + \vec{\alpha})|\Omega\rangle\,, \tag{7.30}$$

where in the second step we used the invariance of the vacuum and in the last that U commutes with the Hamiltonian. It is useful to re-write this expression as an operator annihilating the correlation function. To this end, we expand (7.30) to linear order in $\vec{\alpha}$ and $\vec{\omega}$ and cancel the zeroth order piece with the left hand side. The remaining term is

$$\sum_{a=1}^{n} \frac{\partial}{\partial \mathbf{x}_a}\langle\phi(x_1)\phi(x_2)\dots\phi(x_n)\rangle \stackrel{!}{=} 0\,, \tag{7.31}$$

$$\sum_{a=1}^{n}\left(x_a^i\frac{\partial}{\partial x_a^j} - x_a^j\frac{\partial}{\partial x_a^i}\right)\langle\phi(x_1)\phi(x_2)\dots\phi(x_n)\rangle \stackrel{!}{=} 0\,. \tag{7.32}$$

These relations must be obeyed by all cosmological correlators. The general solution of the first condition is that the correlator only depends on the distance among points, i.e. only on $n-1$ of the n point appearing. For example, this can be chosen to be $\mathbf{x}_a - \mathbf{x}_1$ for $a = 2, \dots n$. The second condition implies that the correlator must be a function of scalar products $\mathbf{x}_a \cdot \mathbf{x}_b$. The full n-correlator then depends on $3n-3-3$ variables.
It is easier to deal with translation invariance in Fourier space

$$\phi(t,\mathbf{k}) \equiv \int d^3x e^{-i\mathbf{k}\cdot\mathbf{x}}\phi(t,\mathbf{x})\,, \qquad \phi(t,\mathbf{x}) \equiv \int \frac{d^3k}{(2\pi)^3}e^{i\mathbf{k}\cdot\mathbf{x}}\phi(t,\mathbf{k})\,. \tag{7.33}$$

The generators acting on Fourier space correlators are then

$$P_i : -ik_i \quad \text{and} \quad R_{ij} : -i\left(k_i\partial_j - k_j\partial_i\right)\,, \tag{7.34}$$

and therefore

$$\sum_{a=1}^{n} \mathbf{k}_a\langle\phi(\mathbf{k}_1)\phi(\mathbf{k}_2)\dots\phi(\mathbf{k}_n)\rangle \stackrel{!}{=} 0\,, \tag{7.35}$$

$$\sum_{a=1}^{n}\left(k_a^i\frac{\partial}{\partial k_a^j} - k_a^j\frac{\partial}{\partial k_a^i}\right)\langle\phi(\mathbf{k}_1)\phi(\mathbf{k}_2)\dots\phi(\mathbf{k}_n)\rangle \stackrel{!}{=} 0\,. \tag{7.36}$$

The first condition is satisfied if the correlator is proportional to a Dirac delta function of the sum of all momenta

$$\langle\phi(\mathbf{k}_1)\phi(\mathbf{k}_2)\dots\phi(\mathbf{k}_n)\rangle \equiv (2\pi)^3\delta_D^3\left(\sum_a \mathbf{k}_a\right)B_n(\mathbf{k}_1,\dots,\mathbf{k}_n)\,, \tag{7.37}$$

where the prime denotes the *stripped* correlator, i.e. with the delta function and $(2\pi)^3$ removed. The second condition implies again that the correlator only depends on the rotational invariant contractions $\mathbf{k}_a \cdot \mathbf{k}_b$.

De Sitter spacetime: dilations Cosmological observations of the power spectrum tell us that primordial perturbations are not only translation and rotation invariant, but also approximately scale invariant. This can be seen for example in the large scale behavior of the CMB temperature anisotropy angular power spectrum C_l^{TT}, where the transfer function is just approximately constant for $l \ll 50$ (the so-called Sachs-Wolfe approximation). On these large scales one finds $C_l^{TT} \propto l^{-2}$, which in angular space implies that the correlation of anisotropies is approximately independent of angle. The leading paradigm to explain such scale invariance is to postulate a phase of quasi-de Sitter expansion in the very early universe. De Sitter spacetime in flat slicing is given by

$$ds^2 = -dt^2 + e^{2Ht} d\mathbf{x}^2 = \frac{-d\tau^2 + d\mathbf{x}^2}{\tau^2 H^2}\,, \tag{7.38}$$

for some constant Hubble parameter H and with $\tau = -e^{-Ht}/H$. This is a maximally symmetric spacetime with ten isometries, arranged according to the group SO(4,1) (the Lorentz group in (4+1)-dimensions or equivalently the conformal group in 3 euclidean dimension). Besides spatial rotations and translations, de Sitter is also invariant under dilations and dS boosts. Let's focus on dilations for the moment, namely

$$\text{dilation:} \quad \tau \to \tau(1+\lambda)\,, \qquad \mathbf{x} \to \mathbf{x}(1+\lambda)\,, \tag{7.39}$$

which trivially leave (7.38) invariant. In real space, the dilation generator is[54]

$$D: \, -\tau\partial_\tau - x^i \partial_i\,, \tag{7.41}$$

> **Box 7.1 Diagonal symmetries and cosmological condensed matter** To conclude that D generates a symmetry of the problem, we have to check that dilations leave invariant not only spacetime but also whatever fills spacetime. In our case that would be some time-dependent scalar with a background $\bar{\phi}(t)$. In general this is *not* invariant under dilations. However, it can happen that the change of $\bar{\phi}(t)$ induced by a dilation is invisible to φ perturbations. This happens very generally in a class of models in which the action for ϕ has a shift symmetry *and* the background is approximately linear in time, i.e. $\dot{\bar{\phi}} \sim$ const. (see e.g. [6, 28, 73, 84])

If all other non-gravitation background quantities also respect this symmetry (see Box 1), then this additional isometry further constrains cosmological correlators. Let's work this out. As before, the sum of D acting on *each* operator in a correlator must vanish by symmetry:

$$\sum_{a=1}^{n} D_a \langle \phi(\tau_1, \mathbf{x}_1)\phi(\tau_2, \mathbf{x}_2)\dots\phi(\tau_n, \mathbf{x}_n)\rangle \stackrel{!}{=} 0\,, \tag{7.42}$$

[54] Check that indeed $\epsilon^\mu = \{-\tau, -x^i\}$ is a Killing vector for the dS metric in (7.38), namely it solves

$$\mathcal{L}_\epsilon g_{\mu\nu} = -(\nabla_\mu \epsilon_\mu - \nabla_\mu \epsilon_\mu) = 0\,. \tag{7.40}$$

where $\mathcal{L}$ is the Lie derivative.

where D_a is the operator in (7.41) acting on $\{\tau_a, \mathbf{x}_a\}$. It's useful to translate this discussion to Fourier space. When acting on a single Fourier-space field $\phi(\tau, \mathbf{k})$, the dilation generator becomes

$$D: \quad (3 - \tau\partial_\tau) + k^i \partial_{k^i} \,, \tag{7.43}$$

where the 3 comes from the d^3k in the Fourier transform (derive it!). When acting on correlators, we find the annoyance that the derivative hits the delta function. This can be bypasses by noticing that we can phrase dilation invariance directly at the level of primed correlators, where the delta function has been removed: simply impose that the combination $-3 + \sum_a D_a$ annihilates the primed correlator

$$\left[-3 + \sum_{a=1}^{n} D_a\right] \langle \phi(\tau_1, \mathbf{k}_1)\phi(\tau_2, \mathbf{k}_2) \ldots \phi(\tau_n, \mathbf{k}_n)\rangle' \overset{!}{=} 0\,. \tag{7.44}$$

There is one further simplification that is very important for cosmology. We observe correlators at late times, namely at the so-called future conformal boundary of dS, $\tau \to 0$ and so we would like to discuss scale invariance directly on this boundary rather than at arbitrary τ_a's. To achieve this we assume that as $k\tau \to 0$ the fields have some time dependence that to leading order is fixed by their free equations of motion

$$\phi(\tau, \mathbf{x}) = \tau^{\Delta_+}\phi_+(\mathbf{k}) + \tau^{\Delta_-}\phi_-(\mathbf{k})\,, \tag{7.45}$$

where the exponents of τ depend on the mass of the field[55] according to (2.38), hence

$$\Delta_\pm = \frac{3}{2} \pm \nu = \frac{3}{2} \pm \sqrt{\left(\frac{3}{2}\right)^2 - \frac{m^2}{H^2}}\,. \tag{7.46}$$

Notice that $\Delta_+ + \Delta_- = 3$, the number of space dimensions, and that in our conventions $\mathrm{Re}\,\Delta_+ \geq \mathrm{Re}\,\Delta_-$. For "light" fields with $m < (3/2)H$, we find that $\Delta_\pm$ are real and we can drop the subleading τ^{Δ_+} dependence in (7.45). At late time we can hence substitute $3 - \tau\partial_\tau$ with Δ_+ and in Fourier space we find[56]

$$D: \quad \Delta_+ + k^i \partial_{k^i}\,. \tag{7.47}$$

Using these results in (7.44) we conclude that as $\tau \to 0$ for light fields the correlators should be homogenous functions of the $\mathbf{k}_a$'s,

$$\langle \phi(\lambda\mathbf{k}_1)\phi_-(\lambda\mathbf{k}_2) \ldots \phi_-(\lambda\mathbf{k}_n)\rangle' = \lambda^{3-n\Delta_+}\langle \phi(\mathbf{k}_1)\phi_-(\mathbf{k}_2) \ldots \phi_-(\mathbf{k}_n)\rangle'\,. \tag{7.48}$$

This result is partially familiar. Since $\mathcal{R}$ becomes constant as $k\tau \to 0$, we should take $\Delta_- = 0$ and so $\Delta_+ = 3$. Then (7.48) says that the n-point correlator of $\mathcal{R}$ at $\tau \to 0$ scales as k^{3-3n}, which gives us positions scales correlators that independent of the rescaling of distances, as in (1.38), and agrees with are findings in (4.32).

[55] In general they can also depend on the spin of the field, but here we focus on scalars. A specific spin dependence of the exponents $\Delta_\pm$ only arises when assuming invariance under the full set of dS isometries, as for example discussed in [6, 46, 62]. When dS boosts are broken the time dependence is not fixed by mass and spin but is model dependent [21].

[56] This is precisely the form of the dilation general in a conformal field theory acting on an operator of conformal dimension Δ_+, a fact that in AdS is one of the pillars of the gauge-gravity correspondence.

de Sitter spacetime: boosts The final 3 isometries of dS spacetime are dS boost. Their finite action is given by

$$\tau \to \gamma\,\tau\,, \qquad\qquad \mathbf{x}^i \to \gamma\left[\mathbf{x}^i + \mathbf{b}^i\left(\tau^2 - |\mathbf{x}|^2\right)\right]\,, \tag{7.49}$$

where

$$\gamma = \left(1 - 2\mathbf{b}\cdot\mathbf{x} - |\mathbf{b}|^2(\tau^2 - |\mathbf{x}|^2)\right)^{-1}\,. \tag{7.50}$$

To linear order in $\mathbf{b}$ this reduces to

$$\text{SCT:}\quad \tau \;\to\; \tau(1 + 2\mathbf{b}.\mathbf{x})\,, \qquad \mathbf{x} \;\to\; \mathbf{x} + 2(\mathbf{b}.\mathbf{x})\,\mathbf{x} + (\tau^2 - |\mathbf{x}|^2)\,\mathbf{b}\,. \tag{7.51}$$

The name refers to the fact that, at distances much shorter than the Hubble radius, $xH \ll 1$, these reduce to the Lorentz boosts of Minkowski's spacetime.

$$\mathbf{b}\cdot\mathbf{K}:\; 2\mathbf{b}\cdot\mathbf{x}\left(\tau\partial_\tau - x^i\partial_i\right) + \left(\tau^2 - |\mathbf{x}|^2\right)b^i\partial_i \qquad\qquad (\text{SCT})\,, \tag{7.52}$$

for an arbitrary constant three-vector $\mathbf{b}$. For the correlator, this implies the constraint

$$\sum_{a=1}^{n}\mathbf{b}\cdot\mathbf{K}_a\langle\phi(\tau_1,\mathbf{x}_1)\phi(\tau_2,\mathbf{x}_2)\ldots\phi(\tau_n,\mathbf{x}_n)\rangle \stackrel{!}{=} 0\,. \tag{7.53}$$

The solutions of these equations have been studied for half a century in an attempt to better understand Conformal Field Theories (see e.g. online reviews [70, 81, 89]). For example, the 2 and 3 point functions are completely fixed up to an overall multiplicative constant, while higher n-point functions can only depends on specific invariants called cross ratios. In Fourier space, the corresponding equations take the form

$$\mathbf{b}\cdot\mathbf{K}:\quad -\left(3 - \tau\partial_\tau\right)2b^i\partial_{k^i} + \mathbf{b}\cdot\mathbf{k}\partial_{k^i}\partial_{k^i} - 2k^i\partial_{k^i}b^j\partial_{k^j}\,, \tag{7.54}$$

$$\left[\sum_{a=1}^{n}\mathbf{b}\cdot\mathbf{K}_a\right]\langle\phi(\tau_1,\mathbf{k}_1)\phi(\tau_2,\mathbf{k}_2)\ldots\phi(\tau_n,\mathbf{k}_n)\rangle' = 0\,. \tag{7.55}$$

7.4 Soft theorems*

Soft theorems[57] are constraints on correlators in the limit in which one of the momenta goes to zero, $\mathbf{k}_a \to 0$. In this Sec. we will derive the following soft theorem

$$\lim_{\mathbf{q}\to 0}\langle\mathcal{R}(\mathbf{q})\mathcal{R}(\mathbf{k})\mathcal{R}(\mathbf{k}')\rangle' = (1 - n_s)P_{\mathcal{R}}(k)P_{\mathcal{R}}(q) + \mathcal{O}(q^2)\,. \tag{7.56}$$

for any single-field model of inflation. This results is a consequence of a non-linearly realized symmetry, which is related to adiabatic modes.

In Sec. 7.1, working in comoving gauge, we found that the change of coordinates (7.57) generates a physical solution with a new $\mathcal{R}$ and γ_{ij}, as long as ω_{ij} is constant. Let us focus on the trace part of ω_{ij}, which is the only one relevant for $\mathcal{R}$. A similar discussion for γ_{ij}can be carried over using the traceless part $\omega_{<ij>}$. So let's drop the out-of-diagonal terms and consider

$$\epsilon^\mu = \left\{0, \lambda x^i\right\}\,, \tag{7.57}$$

[57] This section is not examinable for the Part III course in Lent 2024.

which looks like a constant isotropic rescaling. Since this diff maps a physical solution into another physical solution, it implies the existence of a symmetry of the action of cosmological perturbations. We would like to find this symmetry to linear order in ϵ^μ but to all orders in perturbations. To this end, recall

$$\begin{aligned} g_{ij}(x) \to g'_{ij}(x') &= g_{\mu\nu}(x)\frac{\partial x^\mu}{\partial x'^i}\frac{\partial x^\nu}{\partial x'^j} \\ &= g_{\mu\nu}(x)\delta^\mu_i(1-\lambda)\delta^\nu_j(1-\lambda) \\ &= g_{ij}(x)(1-2\lambda)+\mathcal{O}(\lambda^2)\,. \end{aligned} \tag{7.58}$$

Using the form of the spatial metric in comoving gauge we find

$$e^{2\mathcal{R}'(x')} = e^{2\mathcal{R}(x)}(1-2\lambda) \quad \Rightarrow \quad \mathcal{R}'(x) = \mathcal{R}(x) - \lambda x^i\partial_i\mathcal{R}(x) - \lambda\,. \tag{7.59}$$

So in real and Fourier space the symmetry transformation is

$$\Delta\mathcal{R}(\mathbf{x}) = -\lambda - \lambda x^i\partial_i\mathcal{R}(\mathbf{x})\,, \tag{7.60}$$

$$\Delta\mathcal{R}(\mathbf{k}) = -\lambda(2\pi)^3\delta^3(\mathbf{k}) - \lambda\left(3+\mathbf{k}\cdot\partial_{\mathbf{k}}\right)\mathcal{R}(\mathbf{k})\,. \tag{7.61}$$

We already found the shift $-\lambda$ when discussing adiabatic modes, while the linear transformation

$$\Delta_{lin}\mathcal{R}(\mathbf{k}) = -\lambda\left(3+\mathbf{k}\cdot\partial_{\mathbf{k}}\right)\mathcal{R}(\mathbf{k}) \tag{7.62}$$

was neglected there because it has one more perturbation. In the following we will keep both terms. The charge that generates this transformation needs to satisfy (7.18) and so can be written as

$$Q = Q_S + Q_{lin} \tag{7.63}$$

$$Q_S \equiv -\lambda\int d^3x\Pi(t,\mathbf{x})\,, \tag{7.64}$$

$$Q_{lin} \equiv \frac{1}{2}\int d^3x\{\Pi(t,\mathbf{x}),\Delta_{lin}\mathcal{R}(t,\mathbf{x})\}\,, \tag{7.65}$$

where the parenthesis indicate the anti-commutator and are used to make Q hermitian, while Π is the conjugate momentum of $\mathcal{R}$, namely

$$[\mathcal{R}(t,\mathbf{x}),\Pi(t,\mathbf{y})] = i\delta^3_D(\mathbf{x}-\mathbf{y})\,. \tag{7.66}$$

By the definition, the charge Q must satisfy

$$i\langle[Q,\mathcal{O}]\rangle = \langle\Delta\mathcal{O}\rangle\,, \tag{7.67}$$

which is known as Ward-Takahashi (WT) identity. Here, $\mathcal{O}$ denotes collectively the product of n curvature perturbations $\mathcal{R}$ and the variation is

$$\mathcal{O} = \prod_{a=1}^n\mathcal{R}(\mathbf{k}_a) \quad \Rightarrow \quad \Delta\mathcal{O} = \sum_{a=1}^n\mathcal{R}(\mathbf{k}_1)\ldots\Delta\mathcal{R}(\mathbf{k}_a)\ldots\mathcal{R}(\mathbf{k}_n)\,. \tag{7.68}$$

The idea is to compute the left- and right-hand sides of (7.67) in different ways.

The left-hand side On the left-hand side, Q_{lin} has one more perturbation than Q_S and it is higher order. If parity is a symmetry, $\mathcal{O}$ is Hermitian and we can use

$$i\langle[Q,\mathcal{O}]\rangle = 2\mathrm{Im}\langle\mathcal{O}Q\rangle\,. \tag{7.69}$$

We can compute $Q\,|\Omega\rangle$ in perturbation theory, where the free $\mathcal{R}$ and Π fields are given by

$$\mathcal{R}(\mathbf{x}) = \int \frac{d^3k}{(2\pi)^3}\left[a_{\mathbf{k}}f_k(t)e^{i\mathbf{k}\cdot\mathbf{x}} + a^\dagger_{\mathbf{k}}f^*_k(t)e^{-i\mathbf{k}\cdot\mathbf{x}}\right]\,, \tag{7.70}$$

$$\Pi(\mathbf{x}) = \int \frac{d^3k}{(2\pi)^3}\left[a_{\mathbf{k}}g_k(t)e^{i\mathbf{k}\cdot\mathbf{x}} + a^\dagger_{\mathbf{k}}g^*_k(t)e^{-i\mathbf{k}\cdot\mathbf{x}}\right]\,, \tag{7.71}$$

with $f_k(t)$ and $g_k(t)$ the solutions of the classical linearized equations of motion. In fact, $g_k(t) = a^3\epsilon(t)\dot{f}_k(t)$ with ϵ the Hubble slow-roll parameter, but we will not need this relation here. The canonical quantization (7.66) fixes the so-called Wronskian

$$f_k g^*_k - f^*_k g_k = i\,. \tag{7.72}$$

We can then write

$$Q_S\,|\Omega\rangle \simeq Q_S\,|0\rangle = \int d^3x\Pi(x)\,|0\rangle = g^*_0(t)a^\dagger_{\mathbf{0}}\,|0\rangle \tag{7.73}$$

$$= \frac{g^*_0(t)}{f^*_0(t)}f^*_0(t)a^\dagger_{\mathbf{0}}\,|0\rangle = \frac{g^*_0(t)}{f^*_0(t)}\mathcal{R}(\mathbf{0})\,|0\rangle\,, \tag{7.74}$$

where $\mathcal{R}(\mathbf{0}) = \mathcal{R}(\mathbf{k}=\mathbf{0})$ is the Fourier space field. So we need to compute

$$i\langle[Q,\mathcal{O}]\rangle = 2\mathrm{Im}\left[\langle\mathcal{O}\mathcal{R}(\mathbf{0})\rangle\frac{g^*_0(t)}{f^*_0(t)}\right]\,. \tag{7.75}$$

Since we care about the bispectrum, let us take

$$\mathcal{O} = \mathcal{R}(\mathbf{k})\mathcal{R}(\mathbf{k}')\,, \tag{7.76}$$

where the implicit time argument is $\tau \to 0$. By Hermiticity $\langle\mathcal{O}\mathcal{R}\rangle$ is real:

$$\begin{aligned}\langle\mathcal{O}\mathcal{R}(\mathbf{0})\rangle^* &= \langle\mathcal{R}^\dagger(\mathbf{0})\mathcal{O}^\dagger\rangle = \langle\mathcal{R}(\mathbf{0})\mathcal{R}(-\mathbf{k}')\mathcal{R}(-\mathbf{k})\rangle\\ &= \langle\mathcal{R}(\mathbf{0})\mathcal{R}(\mathbf{k}')\mathcal{R}(\mathbf{k})\rangle = \langle\mathcal{R}(\mathbf{k})\mathcal{R}(\mathbf{k}')\mathcal{R}(\mathbf{0})\rangle = \langle\mathcal{O}\mathcal{R}(\mathbf{0})\rangle\,,\end{aligned} \tag{7.77}$$

where we used $\mathcal{R}^\dagger(\mathbf{k}) = \mathcal{R}^\dagger(-\mathbf{k})$ and that all equal time $\mathcal{R}$ commute with each other. For the other factor in (7.75) we can use the Wronskian condition

$$\mathrm{Im}\,\frac{g^*_0(t)}{f^*_0(t)} = \frac{\mathrm{Im}\,[g^*_0(t)f_0(t)]}{|f_0(t)|^2} = -\frac{i}{2}\frac{[g^*_0(t)f_0(t) - g_0(t)f^*_0(t)]}{|f_0(t)|^2} \tag{7.78}$$

$$= \frac{1}{2|f_0(t)|^2} = \frac{1}{2P_{\mathcal{R}}(0)}\,. \tag{7.79}$$

Our calculation of the left-hand side is complete

$$i\langle[Q,\mathcal{R}(\mathbf{k})\mathcal{R}(\mathbf{k}')]\rangle = \frac{1}{P_{\mathcal{R}}(0)}\langle\mathcal{R}(\mathbf{k})\mathcal{R}(\mathbf{k}')\mathcal{R}(\mathbf{0})\rangle\,. \tag{7.80}$$

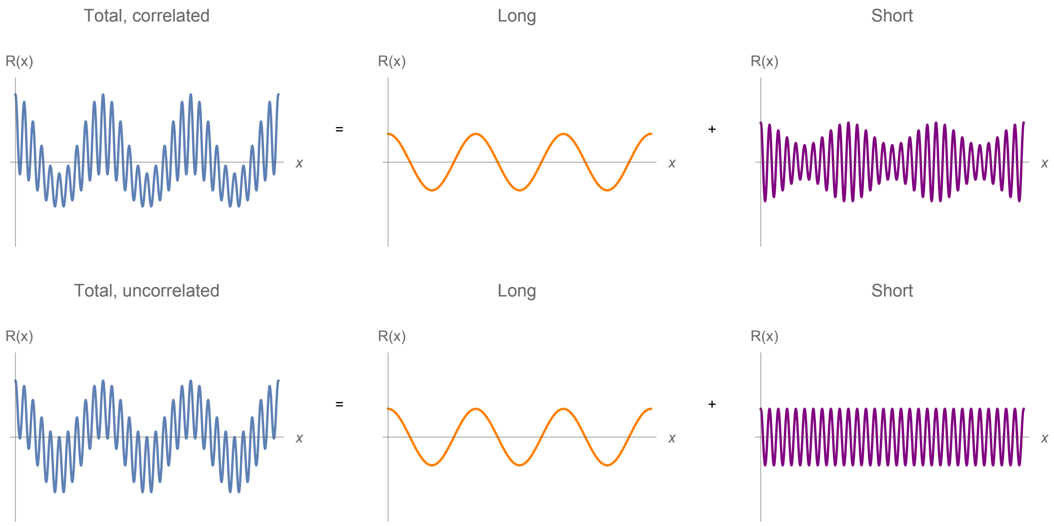


Figure 11: The figure show a profile of $\mathcal{R}(x)$ composed of a long mode (orange) and a short mode (purple). In the top panel they are uncorrelated, while in the bottom they amplitude of the short mode correlates with the long mode. This is the type of non-Gaussianity described by the squeezed limit of the bispectrum, which is fixed by the soft theorem.

The right-hand side The right-hand side of (7.67) depends only on the linear transformation Δ_{lin}. The reason is that we are interested in computing *connected* diagrams, namely diagrams that are proportional to one overall delta function. Instead the shift only contributes to disconnected diagrams, that are proportional to the product of two or more delta functions:

$$\langle \Delta \mathcal{O} \rangle \supset C \sum_{a=1}^{n} \delta_D^3(\mathbf{k}_a) \langle \mathcal{O}(\mathbf{k}_1) \dots \mathcal{O}(\mathbf{k}_{a-1}) \mathcal{O}(\mathbf{k}_{a+1}) \dots \mathcal{O}(\mathbf{k}_n) \rangle \propto \delta_D^3(\mathbf{k}_a) \delta_D^3 \left(\sum_{b \neq a} \mathbf{k}_b \right) ,$$

So, the right-hand side of the (7.67) with the choice (7.76) becomes

$$\langle \Delta \mathcal{O} \rangle = - \left(3 + \mathbf{k} \cdot \partial_{\mathbf{k}} + 3 + \mathbf{k}' \cdot \partial_{\mathbf{k}'} \right) \langle \mathcal{R}(\mathbf{k}) \mathcal{R}(\mathbf{k}') \rangle . \tag{7.81}$$

One can eliminate the Dirac delta function picking up a -3 and express this in terms of the tilt of the power spectrum

$$\langle \Delta \mathcal{O} \rangle' = - \left(3 + k \partial_k \right) P_{\mathcal{R}}(k) = (1 - n_s) P_{\mathcal{R}}(k) . \tag{7.82}$$

We conclude with the WT identity in its final form

$$\lim_{\mathbf{q} \to 0} \langle \mathcal{R}(\mathbf{q}) \mathcal{R}(\mathbf{k}) \mathcal{R}(\mathbf{k}') \rangle' = (1 - n_s) P_{\mathcal{R}}(k) P_{\mathcal{R}}(q) + \mathcal{O}(q^2) . \tag{7.83}$$

A few comments are in order:

- It is the soft limit, a.k.a. squeezed limit of the correlator that is fixed by the theorem. This represents the correlation (a.k.a. mode coupling) between one long wavelength mode $\lambda_{\text{long}} \sim 1/q$ and two short wavelength modes $\lambda_{\text{short}} \sim 1/k \ll \lambda_{\text{long}}$, as depicted in Fig. 11

- This gives the slow-roll suppressed bispectrum. Indeed one can check that the two bispectra in (4.28), which are not slow-roll suppressed, are subleading in this soft limit, $q \to 0$. If one keeps all slow-roll suppressed terms that we have neglected, one can indeed check the validity of this result via direct calculation [67].

- This relation is valid for all single field models in which $\mathcal{R}$ becomes constant (i.e. adiabatic) on superHubble scales, but it is in general violated in multifield models. Observing any deviation from this relation, e.g. in the CMB temperature anisotropy bispectrum would rule out the leading class of inflationary models.

- We derived the relation using comoving momentum $\mathbf{k}$. After relating $\mathbf{k}$ to the physical momentum $\mathbf{k}_p$ using the perturbed metric this result reduces to [74]

$$\lim_{\mathbf{q}_p \to 0} \langle \mathcal{R}(\mathbf{q}_p)\mathcal{R}(\mathbf{k}_p)\mathcal{R}(\mathbf{k}'_p)\rangle' = \mathcal{O}(q^2)\,. \tag{7.84}$$

  This is to be expected since by definition adiabatic modes are locally equivalent to a change of coordinates and so cannot affect the physics. A more formal and precise derivation of this fact uses (conformal) Fermi Coordinates [14,33,34]. The $\mathcal{O}(q^2)$ term is model dependent but has a lower bound of order η [22,32].

- Many other soft theorems exist, also with soft tensor and vectors [53,56,67,72].

8 Phenomenology

ref

It is useful to recap all of our results in a way that can be effectively communicated to a late universe observer, who tries to measure signals from the primordial universe. Our results are predictions for the correlation function of the gauge-invariant variables $\mathcal{R}$ and γ_{ij}.

8.1 Primordial non-Gaussianity

We can write the most generic scalar bispectrum as

$$\langle \mathcal{R}(\mathbf{k}_1)\mathcal{R}(\mathbf{k}_2)\mathcal{R}(\mathbf{k}_3)\rangle = (2\pi)^3 \delta_D(\sum_a \mathbf{k}_a) f_{NL} B(k_1, k_2, k_3)\,, \tag{8.1}$$

where f_{NL} gives us the overall *size* of the bispectrum and B, which is normalized to[58] we impose the conventional normalization

$$k^6 B(k,k,k) = -\frac{18}{5}(2\pi)^4 \Delta^4_{\mathcal{R}}\,, \tag{8.2}$$

gives us the *shape*, i.e. the dependence on the momenta. Above we used the fact that B is scale invariant and so $B(k,k,k) \propto k^{-6}$. Notice that in principle the bispectrum depends on 9 variables, namely 3 vectors, each with 3 components. But translation and rotation invariance each remove 3 of them, in such a way that B only depends on $9-3-3=3$ variables, which we have here chosen to be $k_{1,2,3}$. Intuitively the delta function forces $\mathbf{k}_{1,2,3}$ to form a triangle, which is fully determines by the length of it 3 sides. More generally, an n-point function depends on $3(n-2)$ variables, for $n \geq 3$.

[58] The strange factor 3/5 comes about because a seminal paper on non-Gaussianity [60] considered the scalar potential Φ instead of $\mathcal{R}$ and during matter domination $\mathcal{R} = -5/3\Phi$.

We quoted the shape of the bispectrum induced by the interactions $\dot{\varphi}^3$ and $\dot{\varphi}\partial_i\varphi^2$ in (4.28) to be

$$B_{\varphi'^3} = -\frac{18}{5}(2\pi)^4\Delta_{\mathcal{R}}^4 \frac{1}{2k_1k_2k_3k_T^3}\,, \tag{8.3}$$

$$\begin{aligned} B_{\varphi'(\partial_i\varphi)^2} = -\frac{18}{5}(2\pi)^4\Delta_{\mathcal{R}}^4 \frac{1}{102}\frac{1}{(k_1k_2k_3)^3\,k_T^3}\Bigg[&24\,(k_1k_2k_3)^2 - 8k_T\,(k_1k_2k_3)\left(\sum_{a<b}k_ak_b\right) \\ &- 8\,k_T^2\left(\sum_{a<b}k_ak_b\right)^2 + 22\,k_T^3\,(k_1k_2k_3) - 6k_T^4\left(\sum_{a<b}k_ak_b\right) + 2k_T^6\Bigg]\,. \end{aligned} \tag{8.4}$$

The slow-roll suppressed bispectrum induced by gravity for a canonical scalar field, $P = X - V$, which we did not compute in these lectures, is found to be the sum of two terms [67]

$$B_\epsilon(k_1,k_2,k_3) = -\frac{18}{5}(2\pi)^4\Delta_{\mathcal{R}}^4\cdot\frac{1}{5}\frac{1}{\prod k_i^3}\left[-3\sum_i k_i^3 + \sum_{i\neq j}k_ik_j^2 + 8\frac{\sum_{i>j}k_i^2k_j^2}{k_t}\right] \tag{8.5}$$

$$B_{\mathrm{loc}}(k_1,k_2,k_3) = -\frac{18}{5}(2\pi)^4\Delta_{\mathcal{R}}^4\cdot\frac{1}{3}\frac{\sum_i k_i^3}{\prod k_i^3}\,, \tag{8.6}$$

where the name of "local" in the second shape will become clear in the next lecture. The size of these bispectra are slow-roll suppressed as expected

$$f_{NL}^{\epsilon} = -\left(\frac{5}{12}\right)^2\epsilon\,, \tag{8.7}$$

$$f_{NL}^{\mathrm{loc}} = -\frac{5}{48}(\eta+2\epsilon) = \frac{5}{24}(n_s-1)\,. \tag{8.8}$$

Because of scale invariance, we can write

$$B(k_1,k_2,k_3) = \frac{1}{k_1^6}B\left(1,\frac{k_2}{k_1},\frac{k_3}{k_1}\right)\,. \tag{8.9}$$

While all bispectra share the k^{-6} factor, the differ in how they depend on the dimensionless ratios $x_2 \equiv k_2/k_1$ and $x_3 \equiv k_3/k_1$. To see how similar or different two bispectra are, we plot the following function of two variable

$$B(1,x_2,x_3)x_2^2x_3^2\,. \tag{8.10}$$

where the additional factor of $x_2^2x_3^2$ is added to account for the momentum space volume, see [11] for more details. For example, Figure 12 (from [11]) shows the shape of of the bispectrum in canonical slow-roll inflation, which is induced by gravity. Clearly the correlator peaks in squeezed configurations, where $x_2 \sim 1$ and $x_3 \sim 0$, which translates to $k_3 \ll k_1 \sim k_2$.

8.2 Quantum-to-classical transition

As $\mathcal{R}$ perturbations leave the Hubble radius the become effectively classical. A precise derivation of this fact is still a matter of debate in literature, but we will content ourselves

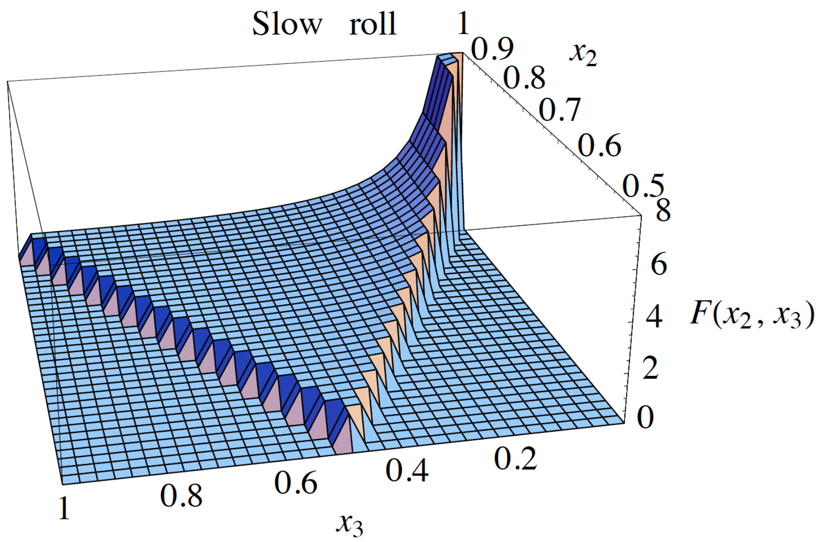


Figure 12: The shape of the bispectrum produced by gravitational non-linearities in canonical, single-field slow-roll inflation.

with a heuristic argument. Using the dS mode functions and the conversion from a canonical field φ_c to $\mathcal{R}$, we can write the free field $\mathcal{R}$ to leading order in $\tau \to 0$ as

$$\mathcal{R}(\mathbf{k}, \tau) \simeq \frac{1}{2\sqrt{\epsilon c_s k^3}} \left(a_{\mathbf{k}} + a^\dagger_{-\mathbf{k}}\right) + \mathcal{O}(\tau^2)\,, \tag{8.11}$$

$$\dot{\mathcal{R}}(\mathbf{k}, \tau) \simeq -H\tau^2 k^2 \frac{1}{2\sqrt{\epsilon c_s k^3}} \left(a_{\mathbf{k}} + a^\dagger_{-\mathbf{k}}\right) + \mathcal{O}(\tau^2)\,. \tag{8.12}$$

We notice that $\mathcal{R}$ and $\dot{\mathcal{R}}$ are proportional to each other and so must commute up to $\mathcal{O}(\tau^2)$ corrections. The defining feature of quantum mechanics, namely the non-commutation of operators, becomes harder and harder to measure as time goes on. More quantitatively, we can try to define a classicality parameter C that quantifies how precise our observations need to be to detect the non-commutation of $\mathcal{R}$ and $\dot{\mathcal{R}}$ (see e.g. [10])

$$C \equiv \frac{|\langle [\mathcal{R}, \dot{\mathcal{R}}] \rangle|}{\sqrt{\langle \mathcal{R}^2 \rangle \langle \dot{\mathcal{R}}^2 \rangle}}\,. \tag{8.13}$$

This can be readily computed and expanded for $\tau \to 0$:

$$C = \frac{|f_k \dot{f}_k^* - f_k^* \dot{f}_k|}{|f_k \dot{f}_k|} \tag{8.14}$$

$$\simeq \frac{2H\tau^3 k^3}{1 \times H\tau^2 k^2} \simeq 2\tau k \to 0\,. \tag{8.15}$$

In particular, at the end of inflation, when we match to the radiation dominated hot big bang,

$$C \sim \frac{k}{(aH)} = e^{-N} \sim e^{-50} \sim 10^{-21}\,, \tag{8.16}$$

where $N \sim 50$ is the number of efoldings between then end of inflation and when the mode k leaves the Hubble radius during inflation. So, unless we can measure the time evolution of $\mathcal{R}$ with a precision of 10^{-21}, we can safely describe correlators as classical averages, as opposed to quantum ones.

Part II

Large Scale Structures

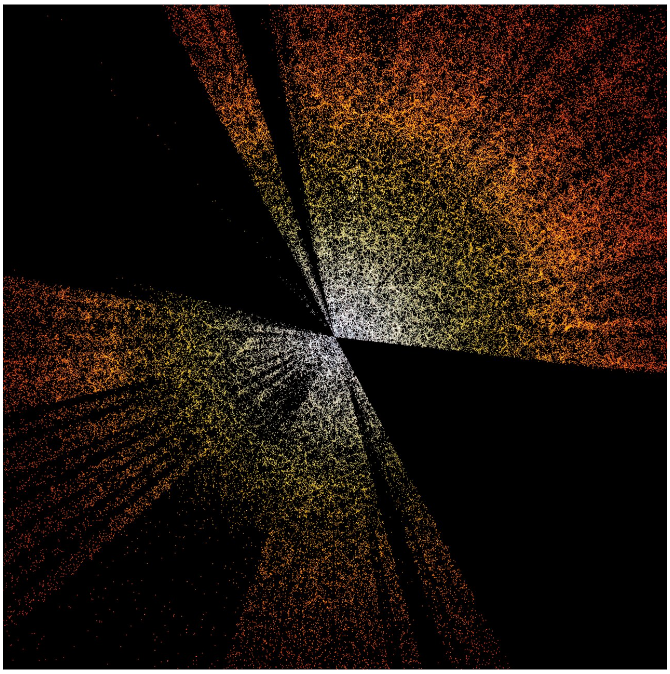


Figure 13: Galaxies measured by the Dark Energy Spectroscopic Instrument (DESI) are luminous objects that trace the underlying large scale structure of matter in the universe. The earth is at the center of the image and the black cones are where our own galaxy prevents us from taking data. Credit D. Schlegel.

In this part of the notes, we will study how matter collapses under the force of gravity and forms the large scale structures that we observe in the universe today (see Fig. 13). After electrons and protons combine into hydrogen, the universe becomes transparent to light and there is no more radiation pressure to contrast the tendency of matter to clump together. Perturbations grow and the linear approximation becomes an increasingly poor description. Here we learn how to study the non-linear dynamics using perturbation theory. We will learn how to use the ideas of effective field theory to develop a self consistent formalism. We will also briefly describe how the distribution of dark matter is traced by luminous objects such as galaxies, which we can more easily see, and we will develop a *bias model* to find quantitative predictions.

Why would we want to go through this trouble? The analysis of CMB fluctuations has tightly constrained the six parameters of the minimal ΛCDM model and further improvements are to be expected from small scale CMB observations, targeting lensing and the CMB polarization. However, many of the constraints are limited by the number of available modes ($N_{\rm CMB} \approx l_{\rm max}^2 \approx 10^6$) in the CMB and any improvement requires measuring additional independent modes. In contrast to the 2-dimensional CMB, Large Scale Structures (LSS) give us a 3-dimensional picture of the universe and hence potentially contain many more independent modes. An LSS cosmological survey is characterized by a volume $V \sim L^3 \sim k_{\rm min}^{-3}$ as well as the maximum wavenumber $k_{\rm max}$ to which one can reliably analyze the data. The number of independent modes in LSS then roughly scales as

$$N_{\rm LSS} \approx (k_{\rm max}/k_{\rm min})^3 \approx 10^8 \tag{8.17}$$

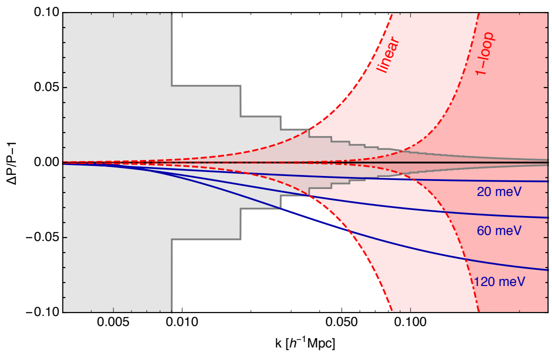


Figure 14: Relative error on the power spectrum from cosmic variance (gray shaded) and the mistake implied by perturbation theory to leading order ("linear", lighter red shade) and next-to-leading order ("1-loop", darker red shade). We also plot the effect of massive neutrinos as a physically interesting signal one would like to detect.

up to $z = 2.5$ and $k_{\rm max} = 0.3\ h{\rm Mpc}^{-1}$. This discussion is exemplified in Fig. 14, where we show the fractional error $\Delta P/P$ (gray bands) in measuring the matter power spectrum as function of the shortest scale $k = k_{\rm max}$ that we can access. As expected the error gets increasingly smaller as we add more data and push to larger k's. However actually computing the power spectrum becomes harder and harder as we go to larger k because shorter scale perturbations are larger and our perturbative approximation become less and less accurate. For example, in red we show the error that a purely linear treatment of the problem would produce and the improvement by including the next-to-learding order terms, here denoted as "1-loop". Clearly, to be sensitive to new physics, such as for example the neutrino masses, we have to account for some non-linearities. Similar conclusions apply to using LSS to constrain the properties of dark energy, dark matter and primordial non-Gaussianity.

In this section we will find predictions for the statistics of the overdensity of dark matter and of galaxies. As we saw in Sec. 8.2, primordial perturbations quickly behave classically as they leave the Hubble radius during inflation, and so we can tackle this problem by neglecting any quantum interference and approximate the evolution using classical equations of motion. However, we still have to keep track of the fact we don't know the precise initial conditions $\delta^{(1)}(\mathbf{x})$, but only their *statistical* properties, which are fixed by the quantum cosmological correlators computed during inflation. So the plan is the following. To compute correlators of matter and galaxy fluctuations $\delta_{m,g}$ we first express them deterministically in terms of the initial conditions $\delta^{(1)}$ right at the end of inflation. In practice this will lead to an expansions of the form

$$\begin{aligned}\delta_{\rm m} &= \delta^{(1)} + \delta^{(2)} + \delta^{(3)} + \dots \\ \delta_{\rm g} &= b_1(\delta^{(1)} + \delta^{(2)} + \delta^{(3)}) + b_2[\delta^{(1)}]^2 + \dots\,,\end{aligned}$$

where $\delta^{(n)} \sim \mathcal{O}([\delta^{(1)}]^n)$ and b_n are time-dependent functions known as "bias parameters" (see Sec. **??**). Then, we use predictions from inflation to compute the correlators of products of $\delta^{(1)}$'s.

9 Newtonian dynamics

ref

In this section, we derive the equations governing the evolution in time of the spatial density of dark matter. In principle, we should solve the Einstein equations and additional equations governing matter. This is not only impossible to do analytically, but it is even prohibitively difficult numerically. Fortunately, two important observations lead to a drastic simplification of the problem. First, recall that all perturbations, including those of dark matter, are fixed by the gauge invariant curvature perturbations $\mathcal{R}$, which in turn are constant on superHubble scales and have a small amplitude, of the order of $\Delta_{\mathcal{R}} \sim 10^{-4.5}$. Hence on large scales the linearized equations give an excellent approximation. As time proceed perturbations re-enter the Hubble radius and start evolving. After photons have decoupled, the evolution consist of a continuous growth of matter inhomogeneities, which hence become large on comoving scales $1/k$ that are much smaller that the comoving Hubble radius, $k \gg aH$. It is only at these subHubble scales that non-linearities become important. Second, notice that dark matter inhomogeneities move much more slowly than the speed of light. A quick and dirty estimate for the velocity is obtained by looking at fully collapsed objects such as dark matter haloes and galaxies. In these systems, the average gravitational potential is $\phi \sim 10^{-5}$. The virial theorem then says that

$$mv^2 \sim \text{kinetic energy} \sim \text{potential energy} \sim m\phi\,, \tag{9.1}$$

and so $(v/c) \sim 10^{-2.5} \ll 1$. In summary, we are interested in studying non-linear structure formation on scales that are (i) much shorter than the Hubble radius and (ii) for non-relativistic velocities. In this regime general relativity reduces to a version of Newtonian dynamics that we now discuss.

9.1 Newtoninan cosmology

Let's first derive the equations governing a single, non-relativistic particle ($v \ll 1$) in an expanding FLRW universe in the Newtonian limit. We will be interested in a universe containing non-relativistic matter, $\rho_m \propto a^{-3}$ and a cosmological constant $\rho_\Lambda = \Lambda M_{\rm Pl}^2$ and we want to keep track of the important fact that the universe is expanding. You might think that both a cosmological constant and cosmic expansion are general relativistic effects and cannot possibly be captured in the Newtonian limit, but that's not the case. As we will now see, a cosmological constant implies a small modification of the Poisson equation and cosmic expansion amounts to a rescaling of distances by the scale factor.

The equation of motion for a particle at physical (as opposed to comoving) position $\vec{r}$ is

$$\ddot{\vec{r}} = -\nabla_r \Phi\,, \tag{9.2}$$

where the dot stands for derivatives with respect to coordinate time t and Φ is the Newtonian gravitational potential. We might be tempted to write the Poisson equation as

$$\nabla_r^2 \Phi \overset{?}{=} 4\pi G_N \rho = \frac{\rho}{2M_{\rm Pl}}\,, \tag{9.3}$$

where ρ is the matter density. But this would account only for matter. In the presence of a cosmological constant, the non-relativistic Poisson equation is modified to (this is justified in Box 1)

$$\nabla_r^2 \Phi = \frac{\rho}{2M_{\rm Pl}^2} - \Lambda \,, \tag{9.4}$$

where $\Lambda = \rho_\Lambda / M_{\rm Pl}^2$ appears in the relativistic theory as in (1.42). Part of the dynamics

Box 9.1 The cosmological constant á la Newton On scales much shorter than the Hubble radius the difference between flat space an de Sitter is a small effect and a cosmological constant can be described by choosing an appropriate Newtonian potential. We start by writing de Sitter space in so-called static coordinates,

$$ds^2 = -f(r)dt^2 + f^{-1}(r)\, dr^2 + r^2 d\Omega_2^2 \,, \qquad\qquad f(r) = 1 - \frac{\Lambda}{3} r^2 \,. \tag{9.5}$$

The name of these coordinates refer to the fact that the metric is time independent, or equivalently there is a non-vanishing time-like Killing vector ∂_t. Static coordinates cover only one wedge (out of four) of the conformal diagram of dS, which is known as the static patch. The Hubble horizon corresponds to the coordinate singularity in the above metric at $r = \sqrt{3/\Lambda} = H^{-1}$. For $r \ll H^{-1}$ the metric is very close to Minkowski and we can identify the Newtonian potential from

$$-g_{00} = 1 + 2\Phi = 1 - \frac{\Lambda}{3} r^2 \quad \Rightarrow \quad \Phi = -\frac{\Lambda}{6} r^2 \,. \tag{9.6}$$

This non-trivial Newtonian potential implies a force pushing all particles away from the origin

$$\ddot{\vec{r}} = -\nabla_r \Phi = \frac{\Lambda}{3} \vec{r} . \tag{9.7}$$

This is nothing but cosmological expansion in disguise! Let's check it. Using the usual Poincaré coordinates in (1.39), consider a particle at constant comoving position $\vec{x}$ and physical distance from the origin $\vec{r} = a\vec{x}$, where $a = e^{Ht}$ is the scale factor. The particle's motion obeys

$$\ddot{\vec{r}} = \ddot{a}\vec{x} = H^2 \vec{r} = \frac{\Lambda}{3} \vec{r} , \tag{9.8}$$

in precise agreement with the Newtonian result in (9.7). Notice that Φ now obeys the modified Poisson equation $\nabla_r \Phi = -\Lambda$. There is more structure in this story than meets the eyes. The parabolic Newtonian potential $\Phi \propto r^2$ has some unique properties. It is one of only two potential for which the force generated by a spherically symmetric distribution coincides with the force that would be generated if that distribution was collapsed to the center. Of course $\Phi = 1/r$ is the other option. Moreover, the Newtonian dynamics now enjoys an enlarged set of symmetries that appear in the classification of kinematical groups as a Wigner Inönü contraction [57] of the Poincaré group [12].

dictated by equations (9.2) and (9.4) is associated to the homogeneous expansion of the universe (the so-called Hubble flow) while part of it stems from the presence of inhomogeneities. To disentangle the two we move from the set of variables $\{\vec{r}, \Phi, \rho\}$ to the new set $\{\vec{x}, \phi, \delta\}$ defined by

$$\vec{r} \equiv a(t)\vec{x} \,, \qquad \Phi = \frac{1}{6}\left(\frac{\bar{\rho}}{2M_{\rm Pl}^2} - \Lambda\right) |\vec{r}|^2 + \phi \,, \qquad \rho(t,\vec{x}) = \bar{\rho}(t)(1 + \delta(t,\vec{x})) \tag{9.9}$$

where $\vec{x}$ is the comoving coordinate of the particle, δ are matter perturbations and we will soon see that ϕ is the part of the Newtonian potential generated by inhomogeneities. The bar on $\bar{\rho}$ indicates a spatial average, i.e.

$$\bar{\rho}(t) \equiv \lim_{V\to\infty} \frac{\int_V d^3x\rho(\vec{x},t)}{\int_V d^3x} \,. \tag{9.10}$$

It will also be convenient to use conformal time[59], $ad\tau = dt$ and denote the corresponding derivatives with a prime, $\partial_\tau f = f'$. In terms of our new variables, Newton's equation in (9.2) becomes

$$\ddot{\vec{r}} = \frac{1}{a}\left(\mathcal{H}'\vec{x} + \mathcal{H}\vec{x}' + \vec{x}''\right) = -\frac{1}{a}\vec{\nabla}_x\Phi = \frac{a}{3}\left(\Lambda - \frac{\bar{\rho}}{2M_{\rm Pl}^2}\right)\vec{x} - \frac{1}{a}\nabla_x\phi \,, \tag{9.11}$$

where we introduced the conformal Hubble parameter $\mathcal{H} \equiv a'/a = aH$. The terms proportional to $\vec{x}$ cancel each other by virtue of the acceleration equation (1.24)

$$\mathcal{H}' = \frac{a^2}{3}\left(\Lambda - \frac{\bar{\rho}}{2M_{\rm Pl}^2}\right) \,, \tag{9.12}$$

and we are left with the new equations

$$\boxed{\begin{aligned} \vec{x}'' + \mathcal{H}\vec{x}' &= -\nabla_x\phi \,, \\ \nabla^2_{\vec{x}}\phi &= \frac{3}{2}\mathcal{H}^2\Omega_{\rm m}(a)\delta \,, \end{aligned}} \tag{9.13}$$

where we introduced the time-dependent fractional matter density

$$\Omega_m(a) \equiv \frac{\bar{\rho}}{3H^2M_{\rm Pl}^2} = \frac{\bar{\rho}a^2}{3\mathcal{H}^2M_{\rm Pl}^2} \,. \tag{9.14}$$

From now on all spatial derivatives will be with respect to comoving coordinates $\vec{x}$, so we drop the label from ∇_x. Defining the *canonical momentum*

$$\vec{p} = am\vec{x}' , \tag{9.15}$$

the equation of motion becomes

$$\vec{p}' = -am\vec{\nabla}_{\vec{x}}\phi \,. \tag{9.16}$$

Let us finally stress again that these equations are only true in the Newtonian regime.

9.2 The Vlasov equation

Now we would like to understand what happens when (i) we have many particles and (ii) the inhomogeneities δ that source ϕ are determined by the way these particles are distributed. We will use the formalism of kinetic theory and hydrodynamics, which we briefly review.

[59] For an arbitrary function $f(t)$ we have $a\dot{f} = f'$ and $a^2\ddot{f} = f'' - \mathcal{H}f'$. Unless otherwise quoted we will refer to dots as derivatives with respect to coordinate time and dashes as derivatives with respect to conformal time.

The number of particles in an infinitesimal phase space volume $d^3x\, d^3p$ around position $\vec{x}$ and momentum $\vec{p}$ is given by

$$dN = f(\vec{x}, \vec{p}, \tau) d^3x d^3p \,, \tag{9.17}$$

where $f(\vec{x}, \vec{p}, \tau)$ is a non-negative distribution, normalized such that the total number of particles is N. Liouville's theorem asserts that f is constant along the trajectories of the systems, i.e. when $\vec{x}$ and $\vec{p}$ solve the equations of motion. This means that if we solved for the trajectories $\vec{x}_i(\tau)$ for i running over each and every particle in the system, then we would find a solution for the phase-space distribution

$$f(\vec{x}, \vec{p}, \tau) = \sum_i \delta^{(\mathrm{D})}(\vec{x} - \vec{x}_i)\delta^{(\mathrm{D})}(\vec{p} - am\vec{x}_i') \,. \tag{9.18}$$

For a very large number of interacting particles it is no feasible to solve for all these trajectories, so we will have to develop an alternative approach. To this end, we will assume here that particles interact exclusively through gravity, so that the equations we derived in the previous section fully describe their evolution. This in particular assumes that dark matter is *collisionless*, i.e. that if we turned off gravity two dark matter particles would fly by each other without even noticing it. In practice it is plausible that dark matter has some interaction either with itself or with standard model particles, so that it can be created in the first place during the hot big bang. However, as long as all interactions are short range and the density of dark matter particles is not too big, we can safely neglect them. In summary, we can use the *Vlasov equation*[60]

$$\frac{df}{d\tau} = \frac{\partial f}{\partial \tau} + \frac{d\vec{x}}{d\tau} \cdot \frac{\partial f}{\partial \vec{x}} + \frac{d\vec{p}}{d\tau}\frac{\partial f}{\partial \vec{p}} \tag{9.19}$$

$$= \frac{\partial f}{\partial \tau} + \frac{\vec{p}}{ma} \cdot \frac{\partial f}{\partial \vec{x}} - am\vec{\nabla}\phi \cdot \frac{\partial f}{\partial \vec{p}} = 0 \,, \tag{9.20}$$

where in the second line we used the equation of motion (9.16). To close the system we have to supplement this by the Poisson equation (9.13) with[61]

$$\rho(\vec{x}, \tau) = \frac{m}{a^3} \int d^3p\; f(\vec{x}, \vec{p}, \tau) \,, \qquad \delta(\vec{x}, \tau) = \frac{\rho(\vec{x}, \tau) - \bar{\rho}(\tau)}{\bar{\rho}(\tau)} \,. \tag{9.21}$$

Because (9.13) and (9.20) now form a coupled non-linear system of partial differential equation in $3 + 3 + 1$ dimensions they are very challenging to solve even with numerical techniques. A standard approach to tackle the problem is to first re-write the Vlasov equation in terms of an infinite set of partial differential equations in $\vec{x}$ by decomposing the $\vec{p}$ dependence of f in some basis. The coefficients of this decomposition are often called *moments*. For example, the first few moments can be conveniently defined by

$$\rho(\vec{x}, \tau) = \frac{m}{a^3} \int d^3p\; f(\vec{x}, \vec{p}, \tau), \tag{9.22}$$

$$v_i(\vec{x}, \tau) = \frac{\int d^3p\; \frac{p_i}{am} f(\vec{x}, \vec{p}, \tau)}{\int d^3p\; f(\vec{x}, \vec{p}, \tau)}, \tag{9.23}$$

$$\sigma_{ij}(\vec{x}, \tau) = \frac{\int d^3p\; \frac{p_i}{am}\frac{p_j}{am} f(\vec{x}, \vec{p}, \tau)}{\int d^3p\; f(\vec{x}, \vec{p}, \tau)} - v_i(\vec{x})v_j(\vec{x}). \tag{9.24}$$

[60] This equation in known under many names, including Liouville equation, Gibbs equation and collisionless Boltzmann equation.

[61] The factor of $1/a^3$ in the definition of ρ indicates that ρ is a mass density in physical volume rather than comoving volume, $\int d^3r \rho = N$.

Here we recognise ρ as the (mass) density, v_i as the average velocity and σ_{ij} as the average deviation from a coherent velocity, i.e. a *velocity dispersion*. Infinitely many other moments can be defined by taking the average over f of higher powers of p_i. The Vlasov equation then becomes an infinite set of differential equations in $\vec{x}$ and τ that couple the infinitely many moments. The problem becomes tractable only when higher moments are increasingly smaller and the infinite set of equations can be truncated to a finite number, which might depend on the desired accuracy. The equations for the moments are often called a Boltzmann *hierarchy*.

An important example of such a hierarchical solution is given by *hydrodynamics*. One assumes that particles interact very efficiently with other nearby particles, in such a way that, on scales much larger than the mean free path between a collision and the next, the system is in local equilibrium. This means that all particles move approximately in the same direction at the same velocity. In this case, one can set the velocity dispersion σ_{ij} and all other higher moments to zero, which provides a consistent solution to the Boltzmann hierarchy.
It turns out that hydrodynamics provides a good description for dark matter too, but for a completely different reason. Indeed, we just stated that dark matter is collisionless and so nearby particles do not interact efficiently enough to ensure local equilibrium (the mean free path is infinite). So why would nearby dark matter particles move together? Surprisingly, the answer is: because of the finite age of the universe! This comes about as follows. The initial conditions for structure formation are such that all velocities vanish on superHubble scales. Velocities are then generated only when modes re-enter the Hubble radius and things start moving under the force of gravity. As time proceed, particles acquire a velocity because they are pulled towards overdensity by the gradients of the gravitational potential. Since the initial distribution is very smooth, nearby particle feel approximately the same potential and accelerate in the same direction. It is only when structures become very large that streams of dark matter particles cross each other, for example those coming from the opposite sides of a collapsed overdensity. To be more quantitative, recall that we estimated typical velocities to be $v/c \sim \mathcal{O}(10^{-2.5})$ (see (9.1)). So the maximum displacement from the beginning of time can be estimated by $vt_0 \sim 10$ Mpc where $t_0 \sim 13.7$ Gy is the age of the universe. This is the distance that substitutes the mean free path when applying hydrodynamics to collisionless dark matter. Now that we have understood structure formation in broad stokes, let's get quantitative.

9.3 The fluid equations

The equations of motion for the moments of the phase-space distribution, namely $\{\rho, v_i, \sigma_{ij}, \dots\}$ can now be obtained by "taking moments" of the Vlasov equation (9.20), namely

$$\int d^3p \, (p_{i_1} p_{i_2} \dots p_{i_n}) \, \frac{df}{d\tau} = 0 \, . \tag{9.25}$$

As an example, let's consider the zeroth moment with a convenient pre-factor[62]:

$$\frac{m}{a^3}\int d^3p\frac{df}{d\tau} = \frac{m}{a^3}\int d^3p\left(\frac{\partial f}{\partial\tau}+\frac{\vec{p}}{ma}\cdot\frac{\partial f}{\partial\vec{x}}-am\vec{\nabla}\phi\cdot\frac{\partial f}{\partial\vec{p}}\right) \tag{9.26}$$

$$= \frac{m}{a^3}\partial_\tau\left(\rho\frac{a^3}{m}\right)+\nabla(v\rho) \tag{9.27}$$

$$= \rho'+3\mathcal{H}\rho+\nabla(v\rho)=0\,. \tag{9.28}$$

This is known as *continuity equation* and enforces the conservation of dark matter particles: a change ρ' in the mass density at some point is either caused by the expansion of the universe, namely the term $3\mathcal{H}\rho$, or by the fact that particles move to a nearby point, namely the term $\nabla(v\rho)$.
We can remove the background solution $\bar{\rho}$ by noticing that for non-relativistic matter $\bar{\rho}\propto a^{-3}$ and we are left with

$$\boxed{\delta'+\vec{\nabla}\cdot\left[(1+\delta)\vec{v}\right]=0}\,, \tag{9.29}$$

Taking the the first moment of the Vlasov equation and using the continuity equation yields

$$\begin{aligned}
0 &= \int d^3p\,p_i\frac{df}{d\tau} = \int d^3p\,p_i\left[\partial_\tau f+\frac{p_j}{ma}\nabla_j f-am\nabla_j\phi\frac{\partial f}{\partial p_j}\right]\\
&= \left(a^4\rho v_i\right)'+\frac{m}{a}\nabla_j\left[\left(\sigma_{ij}+v_iv_j\right)\frac{a^5\rho}{m}\right]+am\nabla_j\phi\frac{a^3p}{m}\\
&= (a^3\rho)\left[(av_i)'+\frac{(a^3\rho)'}{a^3\rho}av_i+\frac{a}{\rho}\nabla_j(\sigma_{ij}\rho)+\frac{a}{\rho}v_i\nabla_j(\rho v_j)+v_j\nabla_j(av_j)+a\nabla_i\phi\right]
\end{aligned}$$

where in the third line we integrate $\partial f/\partial p_j$ by parts and throughout we used the definitions in (9.24). Using the continuity equation the second and third terms cancel each other out leaving our final result

$$\boxed{v_i'+\mathcal{H}v_i+\vec{v}\cdot\vec{\nabla}v_i=-\nabla_i\phi-\frac{1}{\rho}\nabla_i(\rho\sigma_{ij})}\,. \tag{9.30}$$

This equation enforces conservation of momentum and is known as *Euler equation.* In principle the hierarchy of equations continues and the equation for the n-th moment of the Vlaslov equation involves the $(n+1)$-th moment of f. To close the system of equation we postulate that all moments beyond the velocity are vanishing, starting with $\sigma_{ij}=0$. This assumption is sometimes referred to a pressureless perfect fluid assumption. Naively it seems a reasonable assumption and in the linear regime on large scales agrees well with simulations. However, as we will see, it is not a consistent assumption when interactions are take in to account and the initial conditions are assigned statistically, as opposed to deterministically. We will amend this in Sec. 11.

It is always useful to decompose our variables into irreducible representation of the rotation group. The velocity can be decomposed into a longitudinal and a transverse vector part $\vec{v}=\vec{v}_\parallel+\vec{v}_\perp$, where

$$\vec{\nabla}\times\vec{v}_\parallel=0\,,\qquad\qquad \vec{\nabla}\cdot\vec{v}_\perp=0\,. \tag{9.31}$$

[62]The third term in the first line vanishes upon integration by part in $\vec{p}$ because the density f vanishes for very large momenta.

The velocity field can thus be described by its vorticity $\vec{w}$ and its divergence θ,

$$\vec{w} \equiv \vec{\nabla} \times \vec{v} = \vec{\nabla} \times \vec{v}_\perp \,, \qquad\qquad \theta \equiv \nabla \cdot \vec{v} = \nabla \cdot \vec{v}_\parallel \,. \tag{9.32}$$

Vorticity It is easiest to start discussing the vorticity section. Taking the curl of the Euler equation and setting to zero the velocity dispersion σ_{ij} leads to

$$\vec{w}' + \mathcal{H}\vec{w} + \vec{\nabla} \times (\vec{v} \times \vec{w}) = 0. \tag{9.33}$$

This equation tells us that if there is no initial vorticity then evolution won't generate it. Moreover, when perturbations are small we can neglect the last term, which is quadratic, in favour of the first two terms, which are linear. Then we find the linear order equation

$$\vec{w}' + \mathcal{H}\vec{w} = 0 \quad \text{(linear order)}\,. \tag{9.34}$$

This tells us simply that $\vec{w} \propto a^{-1}$, i.e., that any vorticity that might be present initially decays at in the linear regime. Combining these two observations we can conclude that vorticity should be negligible throughout. Yet, there is evidence from simulations that at late times vorticity is generated, especially in high density regions. How can we reconcile this observation with out theoretical model? The issue is that we set to zero the velocity dispersion in the Euler equation. This assumption is clearly violated in regions where many streams of particles have collapsed onto overdense regions and are crossing each other with different velocities. A fully consistent theory should account for this phenomenon too and this can be achieved within the effective field theory of large scale structure, which we will introduce in Sec. 11. For the purpose of these notes, we simply notice that velocity dispersion and the associated vorticity are small on large scales and we will neglect them in the following. This simplification means that the solutions we will find are only valid on large scales, where vorticity is negligible.
Neglecting vorticity we have the following relation in Fourier space between velocity and velocity divergence,

$$\boxed{\vec{v}(\vec{k}) = -i\frac{\vec{k}}{k^2}\theta(\vec{k}).} \tag{9.35}$$

Using this we can re-write the Euler equation in (9.30) as

$$\theta' + \mathcal{H}\theta + \nabla^2\phi = -\nabla_i \left(v_j \nabla_j v_i\right)\,. \tag{9.36}$$

where v is given by (9.35).

9.4 Linear evolution

Neglecting all non-linear terms in the continuity and Euler equations we find the linearized system of equations

$$\delta' + \theta = 0 \tag{9.37}$$

$$\vec{v}' + \mathcal{H}\vec{v} = -\vec{\nabla}\phi\,, \tag{9.38}$$

$$\nabla^2\phi = \frac{3}{2}\mathcal{H}^2\Omega_{\mathrm{m}}(a)\delta\,. \tag{9.39}$$

The system can be solved straightforwardly after rewriting the Euler equation in terms of the velocity divergence θ while setting the vorticity to zero, $\vec{w} = 0$,

$$\theta' + \mathcal{H}\theta = -\nabla^2\phi\,. \tag{9.40}$$

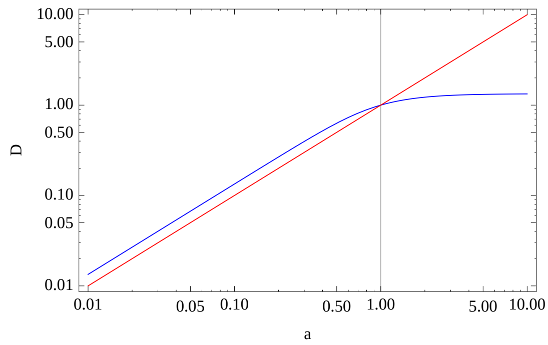


Figure 15: Linear growth factor $D = D_+$ for our fiducial ΛCDM (blue) and a matter only EdS Universe (red).

Density To solve the scalar equation, we take the time derivative of Eq. (9.37) and replace θ' with Eq. (9.40). In the resulting equation, we can replace θ using Eq. (9.40) and $\nabla^2\phi$ using the Poisson Eq. (9.13). We obtain the *linear growth equation*

$$\boxed{\delta''(\vec{x},\tau) + \mathcal{H}(\tau)\delta'(\vec{x},\tau) - \frac{3}{2}\Omega_{\rm m}(\tau)\mathcal{H}^2(\tau)\delta(\vec{x},\tau) = 0\,.} \tag{9.41}$$

It is sometimes convenient to rewrite this in terms of derivatives w.r.t. the scale factor,

$$-a^2\mathcal{H}^2\partial_a^2\delta + \frac{3}{2}\mathcal{H}^2\big[\Omega_{\rm m}(a) - 2\big]a\partial_a\delta + \frac{3}{2}\Omega_{\rm m}\mathcal{H}^2\delta = 0\,. \tag{9.42}$$

This solution of (9.41) has a growing and a decaying solution, which we label D_+ and D_-, respectively,

$$\delta^{(1)}(\vec{k},\tau) = D_+(\tau)\delta_{+,0}(\vec{k}) + D_-(\tau)\delta_{-,0}(\vec{k})\,. \tag{9.43}$$

The decaying solution behave as (as always we choose the normalization $a(\tau_0) = a_0 = 1$)

$$D_-(\tau) = D_{-,0}\,H = D_{-,0}\frac{\mathcal{H}}{a}\,. \tag{9.44}$$

The growing mode solution can then be obtained as

$$D_+(\tau) = D_{+,0}H(\tau)\int_0^{a(\tau)}\frac{da'}{\mathcal{H}^3(a')}, \tag{9.45}$$

where $D_{\pm,0}$ are integration constants that we will fix with the normalization $D_\pm(a = 1) = 1$.

Let us first discuss the solution in a the particular but very relevant case in which the universe is exclusively filled with matter (and is spatially flat). The corresponding FLRW solution is known as an *Einstein-de-Sitter* (EdS) universe. Since matter has equation of state parameter $w = 0$, from (1.21) we find the solution $a \propto t^{2/3}$ so that $H = a^{-3/2}$. In terms of conformal time we have

$$\mathcal{H} = 2/\tau\,, \qquad \Omega_m(\tau) = 1\,, \qquad a = (\tau/\tau_0)^2\,. \tag{9.46}$$

Thus matter inhomogeneity to linear order evolve as

$$D_+ = a\,, \qquad D_- = a^{-3/2} \qquad \text{(Einstein-de-Sitter EdS)}\,. \tag{9.47}$$

In passing, notice that this means that $\phi \propto \mathcal{H}^2\delta$ is constant in time. This might seem surprising because perturbations grown and form very dense objects like dark matter halos and galaxies. The resolution is that matter density decreases because of the expansion, $\bar{\rho} \sim 1/a^3$. At linear other these two effects cancel each other out in such a way that the Newtonian potential remains unchanged. This is not the case anymore beyond linear order.

In the left panel of Fig. 15 we compare this special case and the numerical solution of (9.45) for a model with 30% matter and 70% cosmological constant today (i.e. ΛCDM neglecting radiation). As expected, the two solutions agree at early times, when the cosmological constant is much smaller than ρ_m. Then the growth in the ΛCDM Universe stalls at late times, in contrast with the continues growth $D_+ \propto$ in an EdS universe. Even for generic initial conditions the decaying mode soon becomes negligible. Therefore, in what follows, we will concentrate on the growing mode solutions and use $D \equiv D_+$ unless otherwise stated.

Velocity Now that we know the linear evolution of density perturbations, we can straightforwardly find the linear evolution of velocity perturbations. When vorticity vanishes, we can write $\nabla \cdot \vec{v} = \theta$ in Fourier space and invert it as (9.35). From the linearized continuity equation in Fourier space we have

$$\theta(\vec{k},\tau) = i\vec{k}\cdot\vec{v}(\vec{k},\tau) = -\delta'(\vec{k},\tau) = -\mathcal{H}\frac{d\ln D}{d\ln a}\delta(\vec{k},\tau), \tag{9.48}$$

For the linear growing mode defined above in Eq. (9.45), we have

$$\frac{d\ln H}{d\ln a} + \frac{a}{(aH)^3}\frac{1}{\int_0^a da'\,[a'\,H(a')]^3}, \tag{9.49}$$

which is unity for EdS.

9.5 Initial conditions and the transfer function

Linear density perturbations $\delta^{(1)}$ are related to the primordial power spectrum for $\mathcal{R}$ from inflation. To linear order, it is common to express this relation first as a relation between $\mathcal{R}$ and ϕ, and then one can use Poisson's equation to obtain $\delta^{(1)}$. At linear order for the Newtonian potential we have

$$\phi(\mathbf{k}) = -\frac{27}{50}\mathcal{R}(\mathbf{k})\,T(k/k_{\rm eq})\,\frac{D_+(\tau)}{a}\,. \tag{9.50}$$

Let's unpack this expression. First, the time dependence D_+/a is such that during matter domination ϕ is constant[63] for any k (at linear order), in agreement with our discussion below (9.47). Second, we have introduced the so-called *transfer function* T and $k_{\rm eq} \simeq 0.01\ {\rm Mpc}^{-1}$ is the comoving wavenumber that enters the Hubble radius at matter-radiation equality (i.e. when $\rho_{mat} = \rho_{rad}$, around $z \sim 3300$), $k_{\rm eq} = (aH)_{\rm eq}$. Third, the numerical factor $-27/50$ appears because T is normalized to $T(0) = 1$, and can be computed from our solution for adiabatic modes by changing to Newtonian gauge

[63]This agrees with our finding in Sec.7.1 but to show it explicitly we have to convert from the comoving gauge we used there to Newtonian gauge, which the one that connects to the Newtonian limit.

(see [71]). When analyzing data, T is computed numerically, but for our discussion we will use the ugly but functional Bardeen-Bond-Kaiser-Szalay fitting function [16]

$$T(x) = \frac{\log(1+0.17x)}{0.17x}\left[1+0.28x+(1.18x)^2+(0.4x)^3+(0.49x)^4\right]^{-0.25}, \qquad (9.51)$$

which is accurate to a few percent. It's useful to study the two limits of the transfer function for very short and very large scales respectively

$$k \ll k_{\rm eq} \quad \Rightarrow \quad T(k/k_{\rm eq}) = 1 + \mathcal{O}(k/k_{\rm eq})\,, \qquad (9.52)$$

$$k \gg k_{\rm eq} \quad \Rightarrow \quad T(k/k_{\rm eq}) = \frac{12k_{\rm eq}^2}{k^2}\log\left[\frac{k}{k_{\rm eq}}\right] + \mathcal{O}(k_{\rm eq}^3/k^3)\,. \qquad (9.53)$$

This behaviour can be understood as follows. SuperHubble modes of ϕ are constant, as implied by our adiabatic mode argument. Modes with $k < k_{\rm eq}$ enter the Hubble radius during matter domination and there we noticed ϕ remains constant. Conversely, modes with $k > k_{\rm eq}$ enter during radiation domination. Since radiation does not clump because it has large pressure, $p_{\rm rad} = \rho_{\rm rad}/3$, ϕ decayed during that period. The larger $k/k_{\rm eq}$, the longer a mode spent inside the Hubble radius and the more it decayed, as in (9.53).

We can derive a similar expression for the initial conditions of δ by using the Poisson equation in (9.39) and (9.50). We find

$$\delta^{(1)}(\mathbf{k}) = \frac{\nabla^2\phi}{\frac{3}{2}\mathcal{H}^2\Omega_m} = -\frac{9}{25}\frac{k^2}{\Omega_m H_0^2}\mathcal{R}(\mathbf{k})\,T(k/k_{\rm eq})\,D_+(\tau)\,. \qquad (9.54)$$

This directly gives us a connection between the primordial power spectrum of $\mathcal{R}$ and the matter power spectrum at linear order $P_{\rm lin}$,

$$\langle\delta^{(1)}(\mathbf{k},\tau)\delta^{(1)}(-\mathbf{k},\tau)\rangle' \equiv P_{\rm lin}(k,\tau) = \left[\frac{9}{25}\frac{k^2}{\Omega_m H_0^2}\,T(k/k_{\rm eq})\,D_+(\tau)\right]^2 P_{\mathcal{R}}(k)\,. \qquad (9.55)$$

Measurements of the matter power spectrum together with a best fit are shown in Fig. 16. Since $P_{\mathcal{R}} \sim k^{-4+n_s}$, we find the following limits

$$k \ll k_{\rm eq} \quad \Rightarrow \quad P_{\rm lin}(k) \sim k^4 \times 1 \times \frac{1}{k^{3+(1-n_s)}} \sim k^{n_s} \sim k^1\,, \qquad (9.56)$$

$$k \gg k_{\rm eq} \quad \Rightarrow \quad P_{\rm lin}(k) = k^4 \times \frac{(\log k)^2}{k^4} \times \frac{1}{k^{3+(1-n_s)}} \sim \frac{(\log k)^2}{k^3}\,. \qquad (9.57)$$

The fact that on large scales $P_{\rm lin} \sim k^{n_s}$ is in fact the historical reason why we parameterize the deviation of the spectral tilt from scale invariance as $k^3 P_{\mathcal{R}} = k^{n_s-1}$, which in the power spectrum of $\mathcal{R}$ does not look like a natural choice. In the following we will use these statistical initial conditions to compute non-linear density correlators at late times.

10 Standard Perturbation Theory (SPT) ref

After having obtained some intuition on the solutions in the linear regime, where the quadratic terms are negligible, we will now return to the full equations. We will set up a formalism, known as Standard Perturbation Theory (SPT), to solve the equations

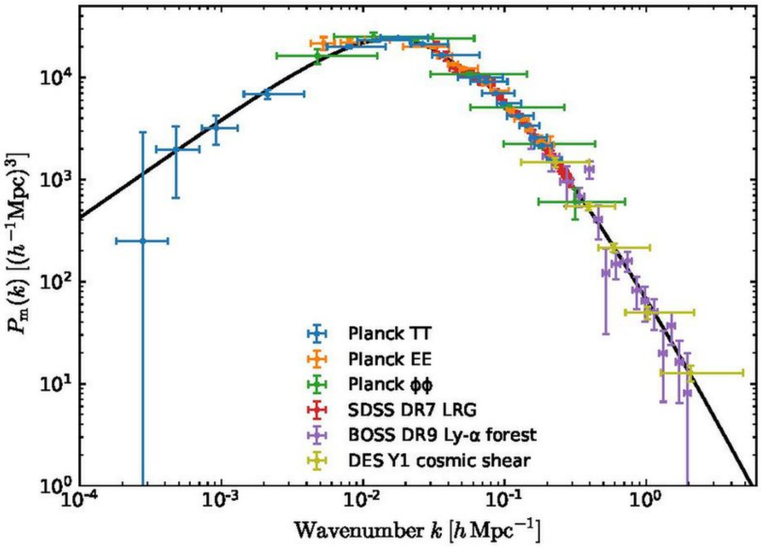


Figure 16: The matter power spectrum measured with a variety of cosmological probes. The large scale linear growth with k is clearly visible.

perturbatively in $\delta, \theta \ll 1$ to any desired order. As concrete examples, we will study the density power spectrum and the bispectrum.

To facilitate the analysis, we will work in Fourier space, where the Euler, (9.36), and continuity, (9.29), equations read as

$$\delta'(\vec{k}) + \theta(\vec{k}) = -\int_{\vec{q},\vec{q}'} (2\pi)^3 \delta^{(\mathrm{D})}(\vec{k} - \vec{q} - \vec{q}')\alpha(\vec{q}, \vec{q}')\theta(\vec{q})\delta(\vec{q}'), \quad (10.1)$$

$$\theta'(\vec{k}) + \mathcal{H}\theta(\vec{k}) + \frac{3}{2}\Omega_{\mathrm{m}}(a)\mathcal{H}^2\delta(\vec{k}) = -\int_{\vec{q},\vec{q}'} (2\pi)^3 \delta^{(\mathrm{D})}(\vec{k} - \vec{q} - \vec{q}')\beta(\vec{q}, \vec{q}')\theta(\vec{q})\theta(\vec{q}'). \quad (10.2)$$

Here α and β are so-called coupling *kernels* and are defined as

$$\alpha(\vec{k}_1, \vec{k}_2) = \frac{\vec{k}_1 \cdot \left(\vec{k}_1 + \vec{k}_2\right)}{k_1^2}, \quad (10.3)$$

$$\beta(\vec{k}_1, \vec{k}_2) = \frac{1}{2}\left(\vec{k}_1 + \vec{k}_2\right)^2 \frac{\vec{k}_1 \cdot \vec{k}_2}{k_1^2 k_2^2} = \frac{1}{2}\frac{\vec{k}_1 \cdot \vec{k}_2}{k_1 k_2}\left(\frac{k_2}{k_1} + \frac{k_1}{k_2}\right) + \frac{\left(\vec{k}_1 \cdot \vec{k}_2\right)^2}{k_1^2 k_2^2}. \quad (10.4)$$

Note that $\alpha(\vec{k}_1, \vec{k}_2)$ is not symmetric in its arguments but $\beta(\vec{k}_1, \vec{k}_2)$ is. The fluid Eqs. (10.1) and (10.2) are non-linear coupled differential equations for the density and velocity divergence. A closed form solution does not exist in general. One can however try to solve them perturbatively in the regime, where $\delta \ll 1$ and $\theta \ll 1$.

By taking linear combination of the continuity and Euler equations we can diagonalized the linear part of these equations and rewrite them as the following second order differential equations,

$$
\begin{aligned}
&\mathcal{H}^2\left[-a^2\partial_a^2+\frac{3}{2}\left(\Omega_\mathrm{m}(a)-2\right)a\partial_a+\frac{3}{2}\Omega_\mathrm{m}(a)\right]\delta=\mathcal{S}_\beta-\mathcal{H}\,\partial_a\left(a\,\mathcal{S}_\alpha\right)\,,\\
&\mathcal{H}\left[a^2\partial_a^2+\left(4-\frac{3}{2}\Omega_\mathrm{m}(a)\right)a\partial_a+(2-3\Omega_m)\right]\theta=\partial_a\left(a\,\mathcal{S}_\beta\right)-\frac{3}{2}\Omega_\mathrm{m}(a)\mathcal{H}\mathcal{S}_\alpha\,.
\end{aligned}
\tag{10.5}
$$

with source terms given by

$$
\begin{aligned}
\mathcal{S}_\alpha(\vec{k},\tau)&=-\int\frac{d^3q}{(2\pi)^3}\,\alpha(\vec{q},\vec{k}-\vec{q})\,\theta(\vec{q},\tau)\delta(\vec{k}-\vec{q},\tau)\,,\\
\mathcal{S}_\beta(\vec{k},\tau)&=-\int\frac{d^3q}{(2\pi)^3}\,\beta(\vec{q},\vec{k}-\vec{q})\,\theta(\vec{q},\tau)\theta(\vec{k}-\vec{q},\tau)\,.
\end{aligned}
\tag{10.6}
$$

The advantage of this formulation is that now δ and θ talk to each other only through the non-linear interactions $\mathcal{S}_\alpha$ and $\mathcal{S}_\beta$. In other words, we have diagonalized the linear theory, for which we found general solutions in Sec. 9.4.
The retarded Green's functions for δ and θ in ΛCDM are given by

$$
\begin{aligned}
&G_\delta(a,a')=\Theta(a-a')\,\frac{2}{5}\frac{1}{\mathcal{H}_0^2\Omega_m^0}\frac{D_+(a')}{a'}\left\{\frac{D_-(a)}{D_-(a')}-\frac{D_+(a)}{D_1(a')}\right\}\,,\\
&G_\theta(a,a')=-\mathcal{H}\,f(a)\,G_\delta(a,a')\,.
\end{aligned}
\tag{10.7}
$$

We could solve these equations numerically order by order. We can however gain more insights into the structure of the solutions by studying a power series ansatz in a matter-only Einstein-de Sitter (EdS) universe.

10.1 Series Ansatz and Coupling Kernels

Standard Perturbation Theory (SPT) aims to solve the fluid equations perturbatively. The expansion parameter is the linear density field $\delta^{(1)}$, which is small because the initial conditions after inflation involved fluctuations of order 10^{-5}. Then, the i-th order contribution to the fully non-linear δ is $\delta^{(n)}=\mathcal{O}([\delta^{(1)}]^n)$. It turns out that calculations simplifies significantly if we work in an EdS Universe, where $D=a$. Hence we will study this case first and discuss the generalization to ΛCDM later. We will furthermore neglect the decaying mode.

Time dependence The good news is that, remarkably, in EdS we will not need to worry about time dependence. The key insight is that the n-th order solution $\delta^{(n)}$ has the simple time dependence $\delta^{(n)}\propto a^n$, irrespectively of the spatial dependence. To see why this is the case, recall that in EdS we have

$$
\mathcal{H}=2/\tau\,,\qquad\Omega_m(\tau)=1\,,\qquad a=(\tau/\tau_0)^2\,,\qquad D=a\,. \tag{10.8}
$$

Let's now go back to (10.2) and (10.1) and let's rewrite them as

$$
\partial_\tau\delta+\theta\sim\alpha\theta\delta\,,\qquad\partial_\tau\theta+\mathcal{H}\theta+\frac{3}{2}\mathcal{H}^2\delta\sim\beta\theta\theta\,. \tag{10.9}
$$

Provided that we attribute the scaling $\theta \sim \mathcal{H} \sim 1/\tau$, these two equations are homogenous in τ and are hence also homogeneous in $a \propto \tau^2$. Homogeneous equations amid a "power law" perturbative ansatz of the form

$$\delta(\vec{k},\tau) = \sum_{n=1}^{\infty} a^n(\tau)\tilde{\delta}^{(n)}(\vec{k}) \qquad \theta(\vec{k},\tau) = -\mathcal{H}(\tau)\sum_{n=1}^{\infty} a^n(\tau)\,\tilde{\theta}^{(n)}(\vec{k}), \tag{10.10}$$

where time and space dependence has been separated and the time dependence is completely solved! Because of this separation, we can think of the time-independent $\tilde{\delta}^{(n)}$ and $\tilde{\theta}^{(n)}$ as the present time value of the n-th order density and velocity (since $a_0 = 1$).

Because of the cosmological constant, time evolution in our ΛCDM universe is not the same as in EdS, even though the two are very close in the redshift range $3000 \ll z \gg 0.5$. The fluid equations in ΛCDM do not allow for a separation of time and space dependence as in Eq. (10.11). However, one can still attempt to approximate the exact solution with a factorized series ansatz that generalize Eq. (10.10) to

$$\delta(\vec{k},\tau) = \sum_{i=1}^{\infty} D^i(\tau)\tilde{\delta}^{(i)}(\vec{k}) \qquad \theta(\vec{k},\tau) = -\mathcal{H}(\tau)f(\tau)\sum_{i=1}^{\infty} D^i(\tau)\,\tilde{\theta}^{(i)}(\vec{k}), \tag{10.11}$$

where the time dependence is now give by powers of the linear growth function D. This ansatz deviates from the exact solution only at the sub-percent level (see e.g. [94] and Fig. 20). Because of this, we will stick to the approximation in (10.11), which is sufficiently accurate for our purpose.

Space dependence The bad news is that we will have to do some work to figure out the spatial distribution. Without loss of generality, we can write the n-th order solutions as convolutions of linear density fields

$$\tilde{\delta}^{(n)}(\vec{k}) = \left\{\prod_{m=1}^{n}\int \frac{d^3q_m}{(2\pi)^3}\delta^{(1)}(\vec{q}_m)\right\} F_n(\vec{q}_1,\ldots,\vec{q}_n)(2\pi)^3\delta^{(\mathrm{D})}(\vec{k}-\vec{q}|_1^n) \tag{10.12}$$

$$\tilde{\theta}^{(n)}(\vec{k}) = \left\{\prod_{m=1}^{n}\int \frac{d^3q_m}{(2\pi)^3}\delta^{(1)}(\vec{q}_m)\right\} G_n(\vec{q}_1,\ldots,\vec{q}_n)(2\pi)^3\delta^{(\mathrm{D})}(\vec{k}-\vec{q}|_1^n) \tag{10.13}$$

with some unknown density kernels F_n and velocity kernels G_n, where $\vec{q}|_i^j = \sum_{l=i}^{j}\vec{q}_l$. A diagrammatic representation of this expansion is shown in Fig. 18. By constructions the kernels are fully symmetric under permutation of their arguments.
Working by induction, we can derive the following recursion relations for the convolution kernels:

$$\begin{aligned} F_n(\vec{q}_1,\ldots,\vec{q}_n) = \sum_{m=1}^{n-1}\frac{G_m(\vec{q}_1,\ldots,\vec{q}_m)}{(2n+3)(n-1)}\Big[&(2n+1)\alpha\left(\vec{q}|_1^m,\vec{q}|_{m+1}^n\right)F_{n-m}\left(\vec{q}_{m+1},\ldots,\vec{q}_n\right)\\ &+2\beta\left(\vec{q}|_1^m,\vec{q}|_{m+1}^n\right)G_{n-m}\left(\vec{q}_{m+1},\ldots,\vec{q}_n\right)\Big] \end{aligned} \tag{10.14}$$

$$\begin{aligned} G_n(\vec{q}_1,\ldots,\vec{q}_n) = \sum_{m=1}^{n-1}\frac{G_m(\vec{q}_1,\ldots,\vec{q}_m)}{(2n+3)(n-1)}\Big[&3\alpha\left(\vec{q}|_1^m,\vec{q}|_{m+1}^n\right)F_{n-m}\left(\vec{q}_{m+1},\ldots,\vec{q}_n\right)\\ &+2n\beta\left(\vec{q}|_1^m,\vec{q}|_{m+1}^n\right)G_{n-m}\left(\vec{q}_{m+1},\ldots,\vec{q}_n\right)\Big]. \end{aligned} \tag{10.15}$$

with $F_1 = G_1 = 1$ such that $\tilde{\delta}^{(1)} = \delta^{(1)}$.
Instead of proving the above relations in general, let's see how to derive F_2 and G_2. A similar method then works to arbitrary n. For this purpose we use first order terms from the power series in the non-linear coupling terms on the right hand side and the second order terms on the left hand side

$$[\delta^{(2)}(\vec{k},\tau)]' + \theta^{(2)}(\vec{k},\tau) = -\int_{\vec{q}} \alpha(\vec{q},\vec{k}-\vec{q})\theta^{(1)}(\vec{q},\tau)\delta^{(1)}(\vec{k}-\vec{q},\tau) \tag{10.16}$$

$$[\theta^{(2)}(\vec{k},\tau)]' + \mathcal{H}\theta^{(2)}(\vec{k},\tau) + \frac{3}{2}\mathcal{H}^2\delta^{(2)}(\vec{k},\tau) = -\int_{\vec{q}} \beta(\vec{q},\vec{k}-\vec{q})\theta^{(1)}(\vec{q},\tau)\theta^{(1)}(\vec{k}-\vec{q},\tau) \tag{10.17}$$

Using the time dependence of the series ansatz, e.g. $\mathcal{H}' = -\mathcal{H}^2/2$, we find that all factors of a and $\mathcal{H}$ cancel out as expected. What's left takes the form

$$2\tilde{\delta}^{(2)}(\vec{k}) - \tilde{\theta}^{(2)}(\vec{k}) = \int_{\vec{q}} \alpha(\vec{q},\vec{k}-\vec{q})\tilde{\theta}^{(1)}(\vec{q})\tilde{\delta}^{(1)}(\vec{k}-\vec{q}) \tag{10.18}$$

$$-\frac{5}{2}\tilde{\theta}^{(2)}(\vec{k}) + \frac{3}{2}\tilde{\delta}^{(2)}(\vec{k}) = -\int_{\vec{q}} \beta(\vec{q},\vec{k}-\vec{q})\tilde{\theta}^{(1)}(\vec{q})\tilde{\theta}^{(1)}(\vec{k}-\vec{q}) \tag{10.19}$$

Since the kernels are defined to be symmetric under all permutations of their arguments (any non-symmetric part would integrate to zero), it is convenient to work with the symmetrized version of α, namely

$$\alpha_s(\vec{q}_1,\vec{q}_2) \equiv \frac{1}{2}(\vec{q}_1+\vec{q}_2)\cdot\left(\frac{\vec{q}_1}{q_1^2}+\frac{\vec{q}_2}{q_2^2}\right) = 1 + \frac{\vec{q}_1\cdot\vec{q}_2(q_1^2+q_2^2)}{2q_1^2q_2^2}\,. \tag{10.20}$$

Using the definition of the kernels in (10.12) and (10.13) this becomes

$$2F_2(\vec{q}_1,\vec{q}_2) - G_2(\vec{q}_1,\vec{q}_2) = \alpha_s(\vec{q}_1,\vec{q}_2)\,, \tag{10.21}$$

$$-\frac{5}{2}G_2(\vec{q}_1,\vec{q}_2) + \frac{3}{2}F_2(\vec{q}_1,\vec{q}_2) = -\beta(\vec{q}_1,\vec{q}_2)\,, \tag{10.22}$$

which is a linear system that can be solved for F_2 and G_2 in terms of α and β. The solution is simply

$$\begin{aligned} F_2(\vec{k}_1,\vec{k}_2) &= \frac{5}{7}\,\alpha_s(\vec{k}_1,\vec{k}_2) + \frac{2}{7}\,\beta(\vec{k}_1,\vec{k}_2) \\ &= \frac{5}{7} + \frac{1}{2}\frac{\vec{k}_1\cdot\vec{k}_2}{k_1k_2}\left(\frac{k_2}{k_1}+\frac{k_1}{k_2}\right) + \frac{2}{7}\frac{\left(\vec{k}_1\cdot\vec{k}_2\right)^2}{k_1^2k_2^2}\,. \end{aligned} \tag{10.23}$$

$$\begin{aligned} G_2(\vec{k}_1,\vec{k}_2) &= \frac{3}{7}\alpha_s(\vec{k}_1,\vec{k}_2) + \frac{4}{7}\beta(\vec{k}_1,\vec{k}_2)\,, \\ &= \frac{3}{7} + \frac{1}{2}\frac{\vec{k}_1\cdot\vec{k}_2}{k_1k_2}\left(\frac{k_2}{k_1}+\frac{k_1}{k_2}\right) + \frac{4}{7}\frac{\left(\vec{k}_1\cdot\vec{k}_2\right)^2}{k_1^2k_2^2}\,. \end{aligned} \tag{10.24}$$

There are words that go with these expressions. For example let's rewrite F_2 as

$$F_2(\vec{k}_1,\vec{k}_2) = \underbrace{\frac{17}{21}}_{\text{growth}} + \underbrace{\frac{1}{2}\frac{\vec{k}_1\cdot\vec{k}_2}{k_1k_2}\left(\frac{k_2}{k_1}+\frac{k_1}{k_2}\right)}_{\text{advection}} + \underbrace{\frac{2}{7}\left[\frac{\left(\vec{k}_1\cdot\vec{k}_2\right)^2}{k_1^2k_2^2} - \frac{1}{3}\right]}_{\text{tidal}}\,. \tag{10.25}$$

Each of these three terms represents a different physical effect at second order. The first term, without any k dependence, is a "growth" term that accounts for the fact that if we are in an overdense region then gravitational collapse is enhanced and density grows faster. The second is an "advection" term and accounts for the fact that matter is moving and so the density at a fixed point $\vec{x}$ is related to the density at an earlier time at a different point $\vec{x} - \vec{v}t$. Finally, the third is a "tidal" term describing the coupling of the non-spherical distribution of matter to the local tidal field s_{ij}.

The explicit expressions for F_3 and G_3 are

$$\begin{aligned} F_3(\vec{q}_1,\vec{q}_2,\vec{q}_3) =& \frac{1}{18}\Big[7\alpha(\vec{q}_1,\vec{q}_2+\vec{q}_3)F_2(\vec{q}_2,\vec{q}_3) + 2\beta(\vec{q}_1,\vec{q}_2+\vec{q}_3)G_2(\vec{q}_2,\vec{q}_3)\Big] \\ &+ \frac{G_2(\vec{q}_1,\vec{q}_2)}{18}\Big[7\alpha(\vec{q}_1+\vec{q}_2,\vec{q}_3) + 2\beta(\vec{q}_1+\vec{q}_2,\vec{q}_3)\Big] \end{aligned} \tag{10.26}$$

$$\begin{aligned} G_3(\vec{q}_1,\vec{q}_2,\vec{q}_3) =& \frac{1}{18}\Big[3\alpha(\vec{q}_1,\vec{q}_2+\vec{q}_3)F_2(\vec{q}_2,\vec{q}_3) + 6\beta(\vec{q}_1,\vec{q}_2+\vec{q}_3)G_2(\vec{q}_2,\vec{q}_3)\Big] \\ &+ \frac{G_2(\vec{q}_1,\vec{q}_2)}{18}\Big[3\alpha(\vec{q}_1+\vec{q}_2,\vec{q}_3) + 6\beta(\vec{q}_1+\vec{q}_2,\vec{q}_3)\Big] \end{aligned} \tag{10.27}$$

The above formulae are not symmetrized over the arguments yet. Upon integration over three equivalent density fields $\delta(\vec{q}_1)\delta(\vec{q}_2)\delta(\vec{q}_3)$ we have to symmetrize, accounting both for the cyclic and odd permutations of the arguments in the kernels.
The kernels obey some general properties [49]:

- All kernels F_n and G_n are rational functions of the momentum of homogenous degree zero, i.e. for every $\lambda \neq 0$ we have

$$F_n(\lambda\mathbf{q}_1,\lambda\mathbf{q}_2,\ldots,\lambda\mathbf{q}_n) = F_n(\mathbf{q}_1,\mathbf{q}_2,\ldots,\mathbf{q}_n) \tag{10.28}$$

  and similarly for G_n.

- *Infra-Red (IR) limit:* When one of the momenta of F_n goes to zero, say $\mathbf{q}_1 \to 0$, then

$$\lim_{\mathbf{q}_1\to 0} F_n = \frac{q_a^i W^i(\mathbf{q}_2,\ldots,\mathbf{q}_n)}{q_a^2}\,, \tag{10.29}$$

  where the index is contracted with some non-universal combination W^i of the other wavenumbers. Naively this suggests an IR-divergence, but in fact, after summing over all contributions to a give order, all IR-divergences must cancel [58,85]. This cancellation is a consequence of the equivalence principle of general relativity, which here reduces to a time-dependent extension of Galilean invariance.

- *Ultra-Violet (UV) limit:* When $\mathbf{k} = \sum_{a=1}^n \mathbf{q}_a$ stays constant, but two momenta diverge, then the kernels vanish,

$$\lim_{q_n\to\infty} F_n(\mathbf{q}_1,\ldots,\mathbf{q}_{n-2},\mathbf{q}_n,-\mathbf{q}_n) \propto \frac{k^2}{q_n^2} \to 0\,, \tag{10.30}$$

  where $k = |\vec{k}_1 + \ldots \vec{k}_{n-2}|$. This is not a trivial property and in fact requires precise cancellation among the different terms. For example, to leading order in $q \to \infty$ the three terms of F_2 in (10.25) are $17/21$, -1 and $4/21$, respectively. It can be show that this follows from mass momentum conservation.

10.2 Diagrammatica

Once the kernels F_n and G_n are found, there is a handy set of diagrammatic rules to compute any correlator to the desired order. Here we state the general rules that we will use them in the following to compute the power spectrum and bispectrum to leading and next-to-leading orders.

Connected correlators The diagrammatic rules are simplest when computing *connected* correlators of fields with a vanishing one-point function, so let's say a few words about those. First, note that in cosmology we usually already work with fields that have vanishing one-point functions, i.e. $\langle\delta(\mathbf{k})\rangle = 0$. If we ever encounter a field that violates this, we can always define a related field that doesn't. For example, if $\langle\phi\rangle \neq 0$, then we define $\tilde{\phi} = \phi - \langle\phi\rangle$, which obeys $\langle\tilde{\phi}\rangle = 0$. Throughout these lectures we assume this has been done and we only concern ourselves with correlators of two or more fields. Second, recall that a correlator is connected if it is proportional to a single momentum-conserving delta function. Any correlator can be written as the sum of products of connected correlators. In this sense connected correlators are the building blocks of generic correlators. A simple non-trivial example is a four-point correlator $\langle\phi_1 \dots \phi_4\rangle$, where we use the shorthand notation $\phi_a = \phi(\mathbf{x}_a, t_a)$. The connected part of this correlator is obtained by subtracting off all possible sums of products of lower-point correlators

$$\langle\phi_1\phi_2\phi_3\phi_4\rangle_c = \langle\phi_1\phi_2\phi_3\phi_4\rangle - [\langle\phi_1\phi_2\rangle\langle\phi_3\phi_4\rangle + 5 \text{ perm's}] \,. \tag{10.31}$$

Here we did not have to bother subtracting terms like $\langle\phi_1\rangle\langle\phi_2\phi_3\phi_4\rangle$ because they vanish. More formally, we could introduce a generating function of disconnected and diagrams, where the former is the exponential of the ladder, but we will not need to this here (see e.g. [18] for a discussion).

Since in large scale structure the time and space dependence factorizes up to very small correlations, we will focus on equal time correlators for convenience, but generalization to unequal times are straightforward. The contribution to an n-point correlator to order p in perturbation theory is given by the following sum of terms

$$\langle\prod_a^n \delta(\mathbf{k}_a, \tau)\rangle|_p = D(\tau)^p \sum_{m_1,\dots,m_n} \langle\prod_a^n \tilde{\delta}^{(m_a)}(\mathbf{k}_a)\rangle\, \delta_{\sum_b m_b, p} \,, \tag{10.32}$$

where the Kronecker delta ensures that all terms in the sum are of the desired order. Each of the m_a-th order fields are a convolution of the kernel F_{m_i} with m_i powers of the linear density, as given in (10.12). The simplest situation is that the linear field is Gaussian. We will assume this here and discuss primordial non-Gaussianity in Sec. 10.3. For Gaussian initial conditions we use Wick's theorem to compute the right-hand side of (10.32) in terms of products of linear power spectra

$$\left\langle \delta^{(1)}(\vec{q})\delta^{(1)}(\vec{q}\,') \right\rangle = (2\pi)^3 \delta^{(\mathrm{D})}(\vec{q} + \vec{q}\,') P_{\mathrm{lin}}(q) \,. \tag{10.33}$$

In figure 17 we show a cartoon summary of how correlators are computed in perturbation theory. The time and spatial evolutions starting from some fixed but arbitrary initial condition $\delta^{(1)}$ are computed deterministically solving the fluid equations in perturbation theory. Then one averages over all possible initial conditions using the classical statistical distribution determined by inflation. In the simplest case of Gaussian initial conditions the only non-vanishing connected correlator of the initial condition is the two point function in (10.33).

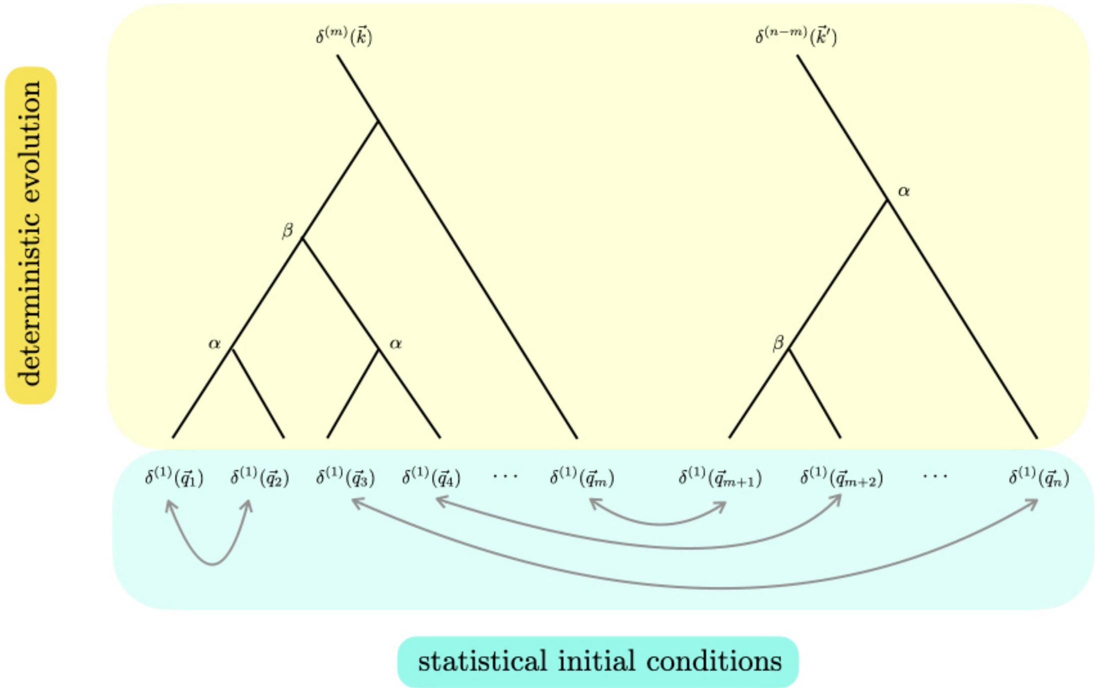


Figure 17: A summary of how correlators are computed in standard perturbation theory for some Gaussian initial conditions.

Diagrammatic rules Because of the many possible contractions at higher order, it is useful to associate to each contribution a diagram according to the following rules:

1. For each field $\tilde{\delta}^{(m_a)}(\mathbf{k}_a)$ in the correlator draw a square corresponding to the kernel $F_{m_a}(\mathbf{q}_1, \ldots, \mathbf{q}_{m_a})$ where $\sum_i^{m_a} \mathbf{q}_i = \mathbf{k}_a$. For correlators of the velocity divergence $\tilde{\theta}^{(m_a)}(\mathbf{k}_a)$ use the kernel G_{m_a} instead.

2. To each square attach one external line carrying momentum $\mathbf{k}_a$ and m_a internal lines carrying momenta $\{\mathbf{q}_1, \ldots, \mathbf{q}_{m_a}\}$, as in Fig. 18. All internal lines must connect either two vertices or one vertex to itself. Because of this, the total order $\sum_a m_a$ is always an even number, as expected from the fact that $\delta^{(1)}$ is a Gaussian field.

3. Decorate each internal line with a dot representing the linear power spectrum in (10.33). The delta function in the linear power spectrum tells us that the momentum flowing in an internal line from one vertex is equal and opposite to the momentum flowing from the other vertex.

4. Integrate over all the momenta of internal lines, $\int d^3q_i/(2\pi)^3$. A tree-level diagram is on for which all these integrals are fixed by momentum conservation in the kernels (squares) and in the power spectrum (dots) in terms of external momenta $\mathbf{k}_a$. Otherwise one has a loop diagram.

5. Multiply by the appropriate symmetry factors that count the number of different ways in which internal lines can be pairwise attached while leading to the same diagram

Example of such diagrams are given in Fig. 22 and 19.

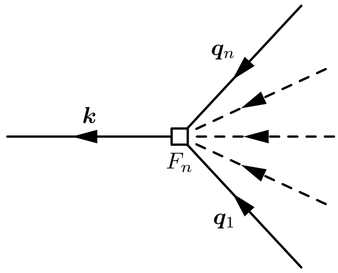


Figure 18: Diagrammatic representation of the series expansion of the density field in Eq. (10.12). The square represents the kernel F_n (or G_n for $\theta^{(n)}$). The line flowing out of it towards the left represents $\delta^{(n)}$, while the n lines flowing into it from the right each represent one initial density field $\delta^{(1)}(\mathbf{q}_i)$.

Two comments are in order. First, these diagrams serve the same purpose as Feynman diagrams, namely to automatize perturbative calculations. However, they are distinct from Feynman diagrams both in the details and in the interpretation. In particular, every internal line represents an averaging over some statistical initial conditions (here assumed to be Gaussian) rather than the propagation of a virtual particle. Second, these rules assume (i) that the system is described by Standard Perturbation Theory (SPT), with its kernels and (ii) that initial conditions are Gaussian, so that the $\delta^{(1)}$'s are only joint pairwise via internal lines connecting kernels. We will see shortly how to deal with primordial non-Gaussianity (see 10.47).

10.3 Non-linear power spectrum and bispectrum

Let's use these graphic rules to compute the leading contributions to the simplest correlators of δ: the two-point function, or power spectrum, and the three-point function or bispectrum. We will first assume Gaussian initial conditions and discuss later in Sec. 10.3 the signature of primordial non-Gaussianity in the matter bispectrum.

Power spectrum The leading order contribution to the power spectrum is simply the linear power spectrum in (10.33). This is represented by the diagram in the left panel of Fig. 19. Since the leading contribution is order $(\delta^{(1)})^2$, the next-to-leading contribution must be fourth order[64]. There are two ways to get this: correlating two second-order density fields or one linear and one third-order density field

$$\left\langle \delta(\vec{k})\delta(\vec{k}') \right\rangle = \left\langle \delta^{(1)}(\vec{k})\delta^{(1)}(\vec{k}') \right\rangle + 2\left\langle \delta^{(1)}(\vec{k})\delta^{(3)}(\vec{k}') \right\rangle + \left\langle \delta^{(2)}(\vec{k})\delta^{(2)}(\vec{k}') \right\rangle + \mathcal{O}((\delta^{(1)})^6)\,. \tag{10.34}$$

The diagrams representing the two next-to-leading contributions are in the center and right-hand panel of Fig. 19. Notice that these diagrams require integrating over an arbitrary internal momentum and are therefore loop diagrams. To simplify the notation,

[64] Recall that here we are assuming Gaussian initial conditions so the correlator of an odd number of $\delta^{(1)}$ vanishes. Later we will see the effect of primordial non-Gaussianity on the matter bispectrum.

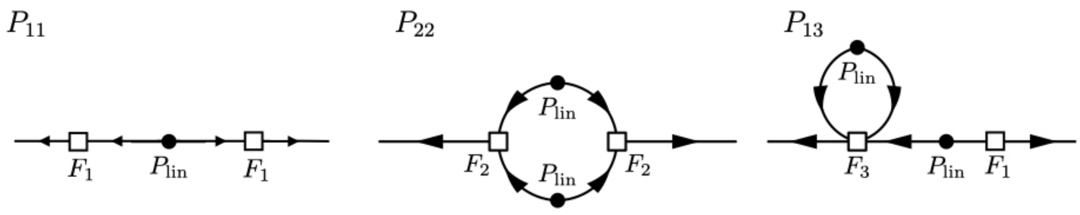


Figure 19: Diagrammatic representation of the one loop matter power spectrum in Eq. (10.37) where $P_{11} = P_{\rm lin}$.

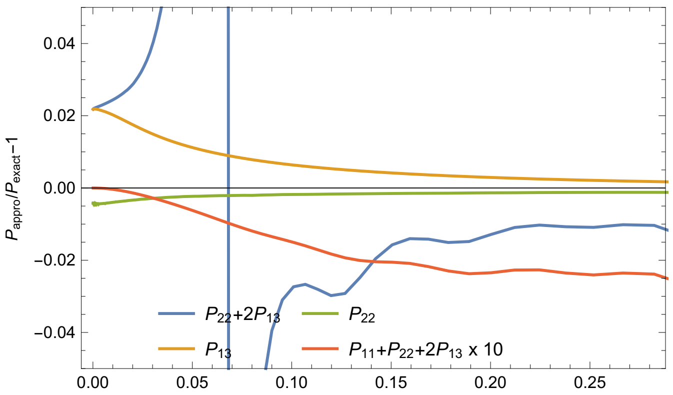


Figure 20: Comparison of contributions to the one-loop matter power spectrum in the EdS approximation and the exact ΛCDM solution. Note that the fractional contribution for the full one-loop result has been multiplied by a factor of 10.

we drop the omnipresent delta function and simply write

$$\left\langle \delta^{(m)}(\vec{k})\delta^{(m')}(\vec{k}')\right\rangle' = P_{mm'}\,, \tag{10.35}$$

$$\left\langle \prod_a^n \delta^{(m_a)}(\vec{k}_a)\right\rangle' = B_{m_1\ldots m_n}\,. \tag{10.36}$$

So we can write more compactly the power spectrum up to one-loop as

$$P_{\text{1-loop}}(k) = P_{11}(k) + P_{13}(k) + P_{31}(k) + P_{22}(k), \tag{10.37}$$

where P_{11} is simply $P_{\rm lin}$ and $P_{13} = P_{31}$. It's worth stressing that these loops have nothing to do with quantum mechanics or virtual particles. We are simply solving some classical equations in perturbation theory. Non-linear interactions in position space become convolutions in Fourier space (see e.g. (10.1)) which in turn generate loop corrections.

Let's inspect these loop diagrams. For P_{22} we have two F_2 kernels, two linear power spectra (the dots) and one loop integral

$$P_{22}(k) = 2\int \frac{d^3q}{(2\pi)^3} P_{\rm lin}(q) P_{\rm lin}(|\vec{k}-\vec{q}|)\left|F_2(\vec{q},\vec{k}-\vec{q})\right|^2 . \tag{10.38}$$

The symmetry factor of 2 counts the two ways in which the internal lines can connect the two vertices. To compute this integral numerically, it is convenient to parametrize

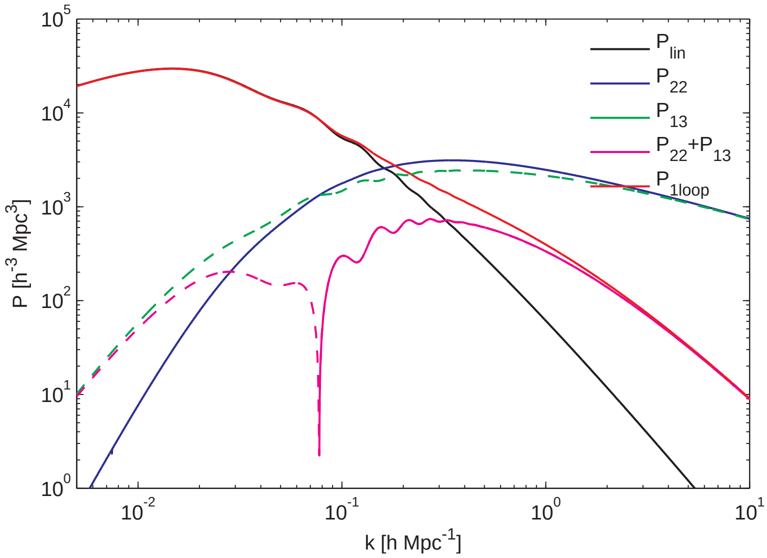


Figure 21: The one-loop power spectrum and its contributions are plotted as function of the wavenumber. Dashed lines indicate that a contribution is negative. The cancellation between $2P_{13}$ and P_{22} for high momenta is evident.

the momenta as $\vec{k} = (0, 0, k)$ and $\vec{q} = rk(\sqrt{1-\mu^2}, 0, \mu)$

$$P_{22} = \frac{k^3}{2\pi^2} \int dr\ r^2 \int d\mu\ P_{\rm lin}(rk) P_{\rm lin}(\psi(r,\mu)k)\, |F_{2,\rm d}(r,\mu)|^2\ , \tag{10.39}$$

where $\psi(r,\mu) = \sqrt{1+r^2-2r\mu}$ and

$$F_{2,\rm d}(r,\mu) = \frac{7\mu + \left(3 - 10\mu^2\right) r}{14r\left(r^2 - 2\mu r + 1\right)}\,. \tag{10.40}$$

The second one-loop contribution P_{13} has one F_3 one $F_1 = 1$ and two power spectra. Notice however that one of the power spectra is not in the loop and so can be moved out of the loop integral. The diagrammatic rules give

$$P_{13}(k) = 3P_{\rm lin}(k) \int \frac{d^3q}{(2\pi)^3} P_{\rm lin}(q) F_3(\vec{k}, \vec{q}, -\vec{q}). \tag{10.41}$$

One realises immediately that this term is a product of a linear power spectrum and a k-dependent correction. The angular part of the integral can be calculated analytically leaving a single radial integral to be computed numerically

$$\begin{aligned} P_{13}(k) =& \frac{k^3}{252(2\pi)^2} P_{\rm lin}(k) \int dr\, r^2 P_{\rm lin}(kr) \times \\ &\times \left[\frac{12}{r^4} - \frac{158}{r^2} + 100 - 42r^2 + \frac{3}{r^5}\left(7r^2+2\right)\left(r^2-1\right)^3 \log\left(\frac{r+1}{r-1}\right)\right]\,. \end{aligned} \tag{10.42}$$

It is now clear how things will continue to higher order as well. To next-to-next-to-leading order we have the two-loop contributions from P_{15}, P_{24} and P_{33}, and so on.

In Fig. 20 we show that the factorized power ansatz we started from deviates from the non-factorized exact solution only at the sub-percent level in the power spectrum. In practice almost all theoretical discussion and numerical analysis use this approximation.

Bispectrum from gravitational evolution Gaussian random fields have vanishing odd-point correlators and so the B_{111} bispectrum vanishes. For Gaussian initial conditions, the leading contribution comes from non-linearities from the gravitational evolution. These are conceptually similar to the non-Gaussianities we computed during inflation but are generated at late times, during structure formation, as opposed to the first fraction of a second after the big bang. To leading order these are

$$\left\langle \delta(\vec{k}_1)\delta(\vec{k}_2)\delta(\vec{k}_3)\right\rangle = \left\langle \delta^{(2)}(\vec{k}_1)\delta^{(1)}(\vec{k}_2)\delta^{(1)}(\vec{k}_3)\right\rangle + 2 \text{ perm's} + \mathcal{O}((\delta^{(1)})^6)\,. \tag{10.43}$$

Introducing the notation

$$\left\langle \delta^{(m_1)}(\vec{k}_1)\delta^{(m_2)}(\vec{k}_2)\delta^{(m_3)}(\vec{k}_3)\right\rangle = (2\pi)^3\delta_D^{(3)}(\textstyle\sum \vec{k})\, B_{m_1 m_2 m_3}\,, \tag{10.44}$$

and using our diagrammatic rules and Fig. 22 we recognize this as a tree-level contribution equal to

$$\begin{aligned} B(\vec{k}_1,\vec{k}_2,\vec{k}_3) &= B_{112} + B_{121} + B_{211} && (10.45)\\ &= 2F_2(\vec{k}_1,\vec{k}_2)P_{\text{lin}}(k_1)P_{\text{lin}}(k_2) + 2 \text{ perm's}\,. && (10.46) \end{aligned}$$

This is a tree-level contribution and we don't have any loop integrals. The one loop contribution to the bispectrum arise at sixth order and are B_{222}, B_{114} and B_{123}, but we will not discuss them here. As anticipated, even in the absence of primordial non-Gaussianity from inflation, the statistic of matter perturbations is non-Gaussian because the fluid equations are non-linear; this provides a crucial source of noise when looking for primordial non-Gaussianity from inflation.

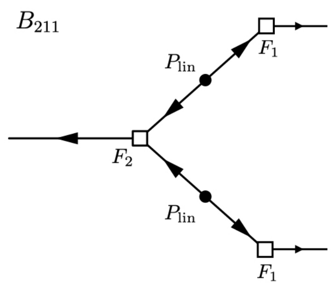


Figure 22: Tree level bispectrum.

Bispectrum from primordial non-Gaussianity In the presence of primordial non-Gaussianity the correlators of the product of three $\delta^{(1)}$'s does not vanish and so $B_{111}^{(NG)}$ is the leading contribution (where the label "(NG)" reminds us that it comes from primordial Non-Gaussianities)

$$\begin{aligned} B_{111}^{NG} &= D_+^3 \,\langle \prod_{a=1}^{3} \tilde{\delta}^{(1)}(\mathbf{k}_a)\rangle' && (10.47)\\ &= -D_+^3\left[\frac{9}{25}\frac{k^2}{\Omega_m H_0^2}T(k/k_{\text{eq}})\right]^3 B(k_1,k_2,k_3)\,, && (10.48) \end{aligned}$$

where $B(k_1,k_2,k_3)$ is the primordial bispectrum of $\mathcal{R}$ and we used 9.54 to convert $\delta^{(1)}$ into $\mathcal{R}$. Depending on the type of primordial non-Gaussianity, this bispectrum may be

	UV-finite	IR-finite
P_{13}	$n<-1$	$n>-1$
P_{22}	$n<1/2$	$n>-1$
$P_{1-\text{loop}}$	$n<-1$	$n>-3$

Table 1: Convergence properties of the one-loop power spectrum and its components P_{22} and P_{13}.

more or less similar to the bispectrum B_{211} generated by late time evolution. Given the current bounds on primordial non-Gaussianity, we already know that B_{111} is at least two orders of magnitude small than B_{112}. We refer the reader to [4] for an in-depth critical discussion of the prospects of detecting primordial non-Gaussianity in large scale structures.

10.4 UV and IR effects

When you meet a loop integral for the first time, it is often useful to ask what happens when the loop momentum is very large and very small. These are called the Ultra Violet (UV) and Infra Red (IR) contributions and often display remarkable properties rooted in general physical principles. As an example we briefly discuss here the one-loop power spectrum.

To begin with, let's us take a look at the various contributions to the power spectrum computed numerically and plotted in Fig. 21. The tree-level power spectrum (black line) grows linearly on large scales, peaks at matter radiation equality, $k\sim k_{\text{eq}}\sim 10^{-2}h\text{Mpc}^{-1}$ and then decays eventually asymptoting $\log(k)^2/k^{-3}$. The very small oscillations are known as Baryon Acoustic Oscillations (BAO) and are a precious source of information, but we will not discuss them here. The one-loop corrections P_{22} (blue line) and P_{13} (green line) are small on large scales, where linear evolution is an excellent approximation and become important somewhere in the range $(0.1-1)\,h\text{Mpc}^{-1}$. The combination of these two terms is significantly smaller than each of them individually indicating a remarkable cancellation.

The IR Let's consider the IR contributions to the loop integrals, when $q\ll k$. We can use the explicit expression for the kernels or the IR limits discussed around (10.29). First P_{22}:

$$P_{22}(k)\supset 2\int_0^{|\mathbf{q}|\ll k}\frac{d^3q}{(2\pi)^3}P_{\text{lin}}(q)P_{\text{lin}}(|\vec{k}-\vec{q}|)\left|F_2(\vec{q},\vec{k}-\vec{q})\right|^2 \tag{10.49}$$

$$\simeq\frac{1}{3}k^2P_{\text{lin}}(k)\int_0^{|\mathbf{q}|\ll k}\frac{dq\,q^2}{(2\pi)^2}\frac{P_{\text{lin}}(q)}{q^2}+\dots \tag{10.50}$$

For P_{13} we find

$$2P_{13}(k)\supset 6P_{\text{lin}}(k)\int_0^{|\mathbf{q}|\ll k}\frac{d^3q}{(2\pi)^3}P_{\text{lin}}(q)F_3(\vec{k},\vec{q},-\vec{q}). \tag{10.51}$$

$$\simeq-\frac{1}{3}k^2P_{\text{lin}}(k)\int_0^{|\mathbf{q}|\ll k}\frac{dq\,q^2}{(2\pi)^2}\frac{P_{\text{lin}}(q)}{q^2}+\dots \tag{10.52}$$

We immediately see that $2P_{13}$ exactly cancels P_{22} confirming analytically what we observed in Fig. 21. This cancellation of IR divergences is a general consequence of the equivalence principle and will continue to hold to all order in perturbation theory [58,85]. The physical picture is that long wavelength velocity with wavenumber k effectively displace short scales with wavenumber q with a uniform amount and hence cannot have any physical effect by the equivalence principle[65]. The next to leading order term in the expansion $q \ll k$ does not cancel and the surviving IR contribution to the power spectrum takes the form

$$P_{22} + 2P_{13} \supset C P_{\rm lin}(k) \int_0^{|\mathbf{q}| \ll k} dq\, q^2 P_{\rm lin}(q) + \dots\,, \tag{10.53}$$

for some numerical constant C. For perturbation theory to make sense we need this integral to converge. For example, assuming a simple power law scaling for $P_{\rm lin}(q) \propto q^n$ as $q \to 0$, we find the IR-finiteness demands $n > -3$, as summarized in Table 1. This is amply satisfied in our universe where on large scale we found $n = k^4 P_{\mathcal{R}}(k) = n_s \simeq 0.96$, with n_s the scalar spectral tilt that would be exactly one in the scale invariant limit.

The UV In the opposite limit of very large loop momentum, $q \gg k$, can again use explicit expression or the general UV limits in (10.30). We find the following UV contributions

$$P_{22}(k) \xrightarrow{k \ll q} \frac{9}{98} k^4 \int_{|\mathbf{q}| \gg k}^{\infty} \frac{P^2_{\rm lin}(q)}{q^4}\,, \tag{10.54}$$

$$P_{13}(k) \xrightarrow{k \ll q} -\frac{1}{3} k^2 P_{\rm lin}(k) \int_{|\mathbf{q}| \gg k}^{\infty} \frac{P_{\rm lin}(q)}{q^2} \left(\frac{61}{210} - \frac{2}{35} \frac{k^2}{q^2} + \dots \right) \tag{10.55}$$

While the exact numerical prefactor needs to be calculated from taking the limit of the F_2 kernel, the k^4 scaling can be inferred from the general asymptotic properties of the F_n kernels discussed in (10.30) above. Whatever the actual numerical value of the integrals, there is something concerning about these expressions: they are saying that arbitrarily short perturbations (i.e. arbitrarily large q) can affect the power spectrum at order k^4 and $k^2 P_{\rm lin}$ respectively. However many of our assumptions at the beginning of this sections, including the collisionless and pressureless fluid description and the smallness of perturbations break down on short scales where matter collapses and forms very dense structures. Because these problems display a separation of scales, they are begging for a resolution within effective field theories.

11 Effective Field Theory Approach ref

We might feel quite good about ourselves: we have a nifty diagrammatic to find any correlator to any order and a recursion relation to compute the kernels that appear in the calculation. And yet there were a few unsatisfactory elements of our treatment that deserve a closer look.

[65] If the velocity was constant in time, this would follow from Galilean invariance of the non-relativistic equations. Here a stronger cancellation takes place, even for arbitrary time-dependent velocities. See [30] and [59] for a nice discussion.

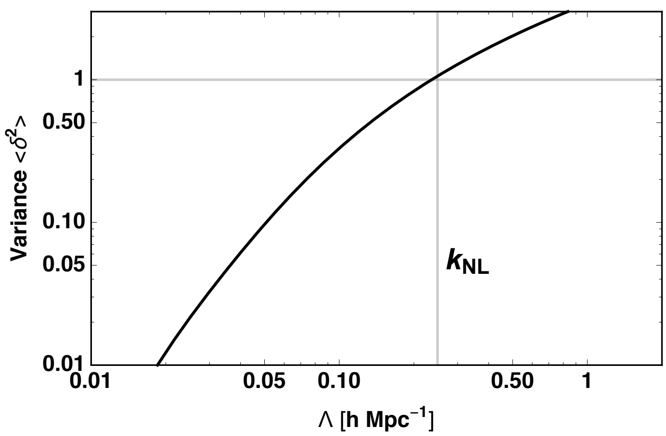


Figure 23: Variance of the density field as a function of wavenumber cutoff Λ corresponding to smoothing scale $R \sim 1/\Lambda$.

Expansion parameter We have been a bit cavalier about what our expansion parameter actually is. We said we would expand in small δ and θ, but how small are these actually in the real world? The expectation value of δ itself vanishes, but we can calculate the variance, i.e. $\langle \delta(\mathbf{x})^2 \rangle$. To get more information we will consider the contribution to the variance σ_Λ^2 of only those modes with wavenumbers below some cutoff wavenumber, $k < \Lambda$. Intuitively this can be thought of as the real space variance of the density field after it has been smoothed on a scale $R = 1/\Lambda$. We will formalize smoothing later, for the moment let's just look at

$$\sigma_\Lambda^2 \equiv \int^{|\mathbf{q}|<\Lambda} \frac{d^3q}{(2\pi)^3} P_{\rm lin}(q) = \frac{1}{2\pi^2} \int_0^\Lambda d\ln q \; q^3 P_{\rm lin}(q) \approx \frac{\Lambda^3 P_{\rm lin}(\Lambda)}{2\pi^2} \tag{11.1}$$

which is shown in Fig. 23 as function of Λ. A perturbative expansion is clearly not warranted once the typical size of fluctuations exceeds unity. The variance is a growing function of Λ and the wavenumber at which it crosses unity gives us an estimate of the scales at which perturbation theory should break down[66]. We will call this scale the *non-linear wavenumber* $k_{\rm NL}$. For the currently favored ΛCDM model this happens at $k_{\rm NL} \approx 0.3\ h\mathrm{Mpc}^{-1}$.

Fluid approximation As particles collapse into overdensities, there comes a moment when flows of particles from different directions cross each other at some point. This phenomenon is known as shell crossing and is depicted in Fig. 24 for a simple one-dimensional case. As the figure shows (see caption for more details) when two shells with different velocities cross the density profile has a peak (formally a divergence). This leads to two problems: first the density at the shell crossing point is formally infinite, which seems to invalidate perturbation theory. Second, at shell-crossing points the fluid approximation clearly breaks down because there is no single fluid velocity. Conversely, the velocity dispersion becomes large and out truncation of the Boltzmann hierarchy seems inconsistent.

Garbage integrals Loop integrals receive contributions from infinitely short scales with $q \to \infty$. From the fact that you and I (and our planet and our galaxy) are very

[66]In fact, perturbation theory could break down even earlier, but no later

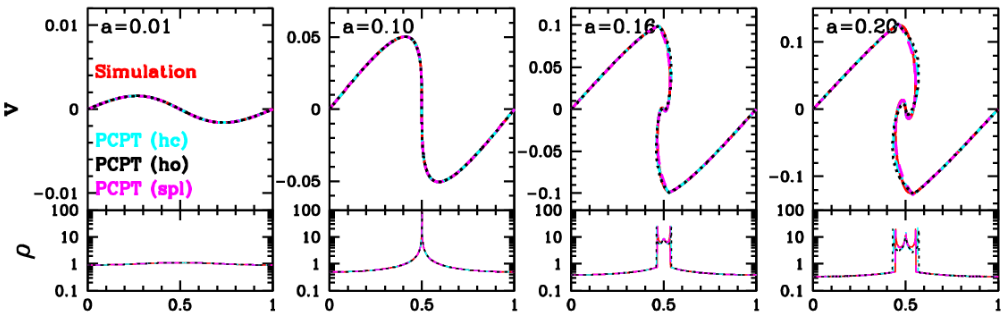


Figure 24: The figure from [95] shows four snapshots of the phase space of a 1-dimensional system of particle collapsing under gravity. In the first panel, an initial inhomogeneous distribution is setup in such a way that the density varies little from point to point. In the second panel the first shell crossing takes pace as many particles with different velocities (i.e. different height) occupy the same spatial position $\mathbf{x} = 0$. The density displays the characteristic diverge at the shell crossing point. In the third and four panels the collapse continues with additional shell crossing and additional spikes in the density profile.

large fluctuations compared to the comic average, we already know that perturbation theory must break down on short scales and in fact on any scale shorter than $k_{NL} \sim 0.3$ Mpc^{-1}. So these contributions to our predictions for correlators are completely garbage, total non-sense. Similar contributions arise in QFT all the time when considering loop corrections. In QFT these are not only garbage, but are often infinite. In our case, loop integrals would be nominally finite because the linear power spectrum decays just barely enough, but this does not matter. Whatever comes from $q \gg k_{NL}$ has not reason to be right and should not be trusted, finite or infinite that it might be. How can we rid ourselves of this garbage? The answer is renormalization.

In the following we will see that all these problems are addressed in a new systematic and self-consistent approach to the problem that goes under the name of Effective Field Theory of Large Scale Structures (EFTofLSS) [17, 23].

11.1 Coarse graining

We realized above, that the density $\delta(\mathbf{x})$ is not a good small parameter to organize our perturbation theory because it is not small at short distances. However, the fact that δ is small at large distances suggests an angle of attack: if we are only interested in questions about distances $1/k$ much larger that the non-linear scale $1/k_{NL}$, then we could try to formulate a theory that only involves the long distance part of δ. Because this problem involved a separation of scales, $k \ll k_{NL}$ it lends itself to an EFT approach. In this section we will introduce *coarse graining* as a tool to precisely separate short scales from long scales and we will re-derive the fluid equations for the coarse grained fields.

We will remove short scales by spatially smoothing all spatially varying fields with a window function W_Λ with characteristic length $1/\Lambda \gg 1/k_{\rm NL}$. Since our fields $\delta(\mathbf{x})$ and $\theta(\mathbf{x})$ both came from the phase space distribution f, we directly smooth that

$$f_l(\vec{x}, \vec{p}, \tau) = \int d^3x' \; W_\Lambda(\vec{x} - \vec{x}') f(\vec{x}', \vec{p}, \tau) \, . \tag{11.2}$$

Here the subscript "l" reminds that that we have selected the "long-distance" part, while we set to zero the short-distance part. More generally, given any spatially varying function $X(\vec{x})$, the coarse graining procedure defines the long wavelength part of fluctuations of arbitrary fields as

$$X_{\mathrm{l}}(\vec{x}) = [X]_\Lambda(\vec{x}) = \int d^3x' W_\Lambda(|\vec{x}-\vec{x}'|)X(\vec{x}') \tag{11.3}$$

and we define the short wavelength part through $X_{\mathrm{s}} = X - X_{\mathrm{l}}$. While the exact functional form of the smoothing function W_Λ is not extremely important, a Gaussian is sometimes convenient

$$W_\Lambda(\vec{x}) = \left(\frac{\Lambda}{\sqrt{2\pi}}\right)^3 \exp\left\{\left[-\frac{1}{2}\Lambda^2 x^2\right]\right\}, \qquad W_\Lambda(\vec{k}) = \exp\left\{\left[-\frac{1}{2}\frac{k^2}{\Lambda^2}\right]\right\}.$$

A hand property of coarse graining by a convolution in position space, as in (11.3), is that it reduces to a simple multiplication in Fourier space

$$X_{\mathrm{l}}(\mathbf{k}) = X(\mathbf{k})W_\Lambda(\mathbf{k})\,. \tag{11.4}$$

The coarse grained density is

$$\rho_{\mathrm{l}} = \int d^3p\; f_{\mathrm{l}}(\vec{x},\vec{p},\tau)\,, \tag{11.5}$$

We could define the coarse grained velocity similarly, but it will turn out more convenient to instead define v_{l} as

$$\vec{v}_{\mathrm{l}} = \frac{1}{\rho_{\mathrm{l}}}\frac{m}{a^3}\int d^3p\; \frac{\vec{p}}{ma} f_{\mathrm{l}}(\vec{x},\vec{p},\tau)\,. \tag{11.6}$$

Hence keep in mind that we are abusing our notation, and in particular $v_{\mathrm{l}} \neq [v]_\Lambda$.

The effective stress tensor The coarse grained fluid equations are obtained by taking the first and second moment of the coarse grained Vlasov equation $[df/d\tau]_\Lambda = 0$. The result is

$$\delta_{\mathrm{l}}' + \partial_j\left[(1+\delta_{\mathrm{l}})v_{\mathrm{l},j}\right] = 0\,, \tag{11.7}$$

$$v_{\mathrm{l},i}' + \mathcal{H}v_{\mathrm{l},i} + \partial_i\phi_{\mathrm{l}} + v_{\mathrm{l},j}\partial_j v_{\mathrm{l},i} = -\frac{1}{\rho_{\mathrm{l}}}\partial_j\left[\tau_{ij}\right]_\Lambda\,. \tag{11.8}$$

The last term in the Euler equation constitutes the noteworthy novelty that is interpreted as an *effective stress-energy tensor*. It a spectacularly messy term involving the smoothing of products of short modes and of long modes. The full expression can be found in App. A of [23], but here we will only quote two of the most iconic terms

$$\partial_j\left[\tau_{ij}\right]_\Lambda \supset \partial_j\left[v_{\mathrm{l}}^{[i}(v-v_{\mathrm{l}})^{j]}\rho\right]_\Lambda + \left[\rho_s\partial_i\phi_s\right]_\Lambda + \dots \tag{11.9}$$

The dots include a bunch other terms which are either quadratic in long perturbations or the smoothing of terms quadratic in short perturbations such as the terms displayed above. The key point is that to really calculate the effective stress tensor we would need to know f on short scales, where the fluid approximation breaks down and we have no perturbative handle on the problem[67]. Indeed this issue was to be expected: for a non-linear equation there would always be terms of the form $[\text{short} \times \text{short}]_\Lambda$ contributing to long perturbations because given two arbitrary short modes, $q_1, q_2 \gg k_{NL}$, one can always build a long mode by choosing them to be almost equal and opposite, $|\mathbf{q}_1 - \mathbf{q}_2| \ll k_{NL}$.

11.2 The effective field theory expansion

To make progress we will appeal to EFT logic. We will follow the general and by now familiar EFT steps:

- Determine the relevant degrees freedom. For us, these are the smoothed density and velocity fields. In fact, for simplicity we will drop the vorticity and focus on δ_l and θ_l (see [68]).
- Include all possible interactions in the degrees of freedom compatible with the symmetries of the problem. For us the relevant symmetries are translation and rotation invariance, the equivalence principle, which manifests itself in a time-dependent generalization of Galilean symmetry, and finally the conservation of mass and momentum on short scales.
- Develop a consistent *power counting scheme* and truncate to the necessary order for the desired precision. We will postpone this to the next section.

Following these steps we can *formally* expand $\partial_j [\tau_{ij}]_\Lambda$ in powers of δ_l and θ_l with arbitrary unknown coefficients. Homogeneity and isotropy enforce the coefficients to be rotation invariant tensors, such as scalars or powers of δ_{ij}, that are independent of position. Conversely we cannot exclude that the EFT coefficients are time dependent because we don't have any symmetry under time translations. Because of the separation of scale $k \ll k_{NL}$, we are allowed to keep only a finite number of terms. The more accurate predictions we want the more terms we need to keep. We postpone a more quantitative discussion of how many terms we actually need to the next section where we will develop a *power counting scheme*.
Finally, we write down the *effective stress tensor* as an infinite sum of powers of δ_l, θ_l and their derivatives[68]

$$\begin{aligned}
[\tau_{ij}]_\Lambda =& \bar{p}\delta_{ij} + \bar{\rho}\tilde{c}_s^2 \delta_{ij}\delta_l - \bar{\rho}\frac{\tilde{c}_{bv}^2}{\mathcal{H}}\delta_{ij}\partial_m v_{l,m} \\
&- \frac{3}{4}\bar{\rho}\frac{\tilde{c}_{sv}^2}{\mathcal{H}}\left[\partial_i v_{l,j} + \partial_j v_{l,i} - \frac{2}{3}\delta_{ij}\partial_m v_{l,m}\right] + \Delta\tau_{ij} + \mathcal{O}((\delta_l, v_l)^2)
\end{aligned} \tag{11.10}$$

[67] In [23] these terms were computed using numerical simulations to have non-perturbative results about the short scales and it was shown that the final set of equations agree with the EFT expansion we now introduce in the limit of separation of scales $k \ll k_{NL}$.

[68] Here are glossing over a few subtleties. First of all, the expansion is better organized in powers of the Newtonian potential ϕ, but the equivalence principle dictates that it must appear with at least two derivatives. Second, one should in principle include terms that are non-local in time integrating fields over the past history of a fluid elements. As shown in [13] in perturbation theory this can alway be re-written as a sum of terms that are local in time at the cost of introducing a controlled number of inverse spacial Laplacians. Finally, different interactions may be degenerate at low order and become distinct only to higher order.

Here we introduced the background pressure $\bar{p}$, the speed of sound $\tilde{c}_s^2$ as well as the bulk- and shear-viscosities $\tilde{c}_{bv}^2$ and $\tilde{c}_{sv}^2$. All the EFT coefficients must be independent of position by homogeneity and isotropy, but will in general depend on time. Finally, $\Delta\tau_{ij}$ describes terms that cannot be written as any power of long modes. These arise because two different concrete realization of short mode perturbations can have the same statistical properties but in general differ from each by some statistical fluctuation. This difference has nothing to do with long modes and would be present even if we set all long modes to zero. We will refer to this term as the *stochastic noise*. We will see that it is in part constrained by mass and momentum conservation on short scales.

To compute the equation of motion for the velocity divergence we have to take the divergence of the Euler equation (we will neglect vorticity). This gives rise to a term with two spatial derivatives of the effective stress tensor, which provides a source term in the divergence of the Euler equation

$$\theta'(\vec{k}) + \mathcal{H}\theta(\vec{k}) + \frac{3}{2}\Omega_{\rm m}(a)\mathcal{H}^2\delta(\vec{k}) = -\int_{\vec{q}}\beta(\vec{q},\vec{k}-\vec{q})\theta(\vec{q})\theta(\vec{k}-\vec{q}) + \tau_\theta(\vec{k})\,, \tag{11.11}$$

where

$$\begin{aligned}\tau_\theta = \partial_i\partial_j\tau_{ij} &= \bar{\rho}\left[\tilde{c}_{\rm s}^2\partial^2\delta_{\rm l} - \frac{\tilde{c}_{bv}^2}{\mathcal{H}}\partial^2\theta_{\rm l} - \frac{3}{4}\frac{\tilde{c}_{sv}^2}{\mathcal{H}}\partial^2\theta_{\rm l}\right] + \Delta J + \ldots \\ &= \bar{\rho}\left[\tilde{c}_{\rm s}^2\partial^2\delta_{\rm l} - \frac{\tilde{c}_{v}^2}{\mathcal{H}}\partial^2\theta_{\rm l}\right] + \Delta J + \ldots\end{aligned} \tag{11.12}$$

Here we defined the stochastic noise by $\Delta J = \partial_i\partial_j\Delta\tau_{ij}$ and we combined shear and bulk viscosities into a single effective term because they contribute in the same way to θ (but they would contribute differently to vorticity). We have neglected infinitely many terms with more powers for $\delta_{\rm l}$ and $\theta_{\rm l}$ and more derivatives, such as $\delta_{\rm l}^2$, $\partial_i\delta_{\rm l}\partial_i\theta_{\rm l}$ and so on, because these are expected to be higher order in our power counting, i.e. in $k/k_{NL} \ll 1$. In summary, the EFT fluid equations can again be written as in (10.5) with the only modification compared to SPT that $S_\beta \to S_\beta + \tau_\theta$.

Let's contrast these fluid equations with the ones in the previous sections. Before we had stated that we can truncate the Boltzmann hierarchy of equations by setting to zero the higher moments of the phase space distribution. That amounted to (i) neglect the effect of short modes and (ii) impose that long modes are well described by a perfect, pressureless fluid. Both assumptions are clearly incorrect to describe the actual system and can at most be useful as *approximations*. In contrast here we have seen that we can model the effect of short modes by an expansion in long modes (that is local in space). In turn this taught us that long modes do have a non-vanishing pressure, resulting in a finite speed of propagation, and are furthermore described by a dissipative rather than perfect fluid, as seen from the appearance of viscosities. As long as we can ensure a separation of scales, $k \ll k_{NL}$, the EFT approach has upgraded what was just an approximation to a systematic *expansion*, which can deliver, in principle, arbitrary precise results.

EFT corrections Having calculated the stress tensor, we now need to solve the equations again in presence of this correction. For this purpose it is good to develop a notion of the size of various terms. As we will motivate in more detail below, we will assume $c_{\rm s}^2 = \mathcal{O}([\delta^{(1)}]^2)$ and $\Delta J = \mathcal{O}([\delta^{(1)}]^2)$. Since from now on we will only exclusively discuss long modes, we drop the label "l" from all fields.

The inclusion of the source terms is best performed using the Greens functions of the linear differential operators on the left-hand side of (10.5), namely those in (10.7). We will focus on the corrections that are linear in $\tilde{c}^2_{s,v}$ and ΔJ, since these are of the same order as the one-loop corrections we computed in the power spectrum. For example, to leading order in $\tilde{c}^2_{s,v}$ and ΔJ, the solution for δ can be obtained using the Green's function

$$\delta_{c_s^2}(\vec{k}, a) = \int da' G_\delta(a, a') k^2 \left[\tilde{c}_s^2(a') + \tilde{c}_v^2(a')\right] \delta^{(1)}(\vec{k}, a') \,, \tag{11.13}$$

$$\delta_J(\vec{k}, a) = \int da' G_\delta(a, a') \Delta J(\vec{k}, a') \,. \tag{11.14}$$

We don't really know how to perform these time integrals, but we know that, to the order we are working, the c_s term can be written as

$$\delta_{c_s^2} \equiv c_s^2 k^2 \delta^{(1)}(\vec{k}) \tag{11.15}$$

where we introduced a new c_s^2, without the tilde, which accounts for the collective effect of $\tilde{c}^2_{s,v}$ and their time integral. In summary, our solution for δ to order $(\tilde{\delta}^{(1)})^3$ is

$$\delta(\vec{k}, \tau) = \delta^{(1)}(\vec{k}, \tau) + \delta^{(2)}(\vec{k}, \tau) + \delta^{(3)}(\vec{k}, \tau) - k^2 c_{\rm s}^2(\tau) \delta^{(1)}(\vec{k}, \tau) + \delta_J(\vec{k}, \tau) \,. \tag{11.16}$$

Not only we have found new corrections by treating short modes more systematically, but, as we will see now, we have formulated a theory in which one can make sense of garbage contributions from UV integrals via renormalization.

11.3 Renormalization, counterterms and the EFT power spectrum

We are now in the position to calculate the EFT power spectrum, which includes the terms P_{SPT} appearing in SPT and the EFT corrections $\Delta P_{\Delta EFT}$ we just derived,

$$P = P_{STP} + \Delta P_{\Delta EFT} \,. \tag{11.17}$$

The contributions coming from Standard Perturbation Theory (SPT) take the same form as before with the crucial difference that now the fields have all been smoothed on a scale $1/\Lambda$,

$$P_{SPT}(k) = P_{\rm lin}(k) + P_{22,\Lambda}(k) + 2P_{13,\Lambda}(k) + \dots \,, \tag{11.18}$$

where we have neglected the explicit dependence of P_{11} on the cutoff Λ as it is negligible for $k \ll \Lambda$. This goes a long way addressing the first and third issue we pointed out at the beginning of this section: now we have a better expansion parameter, namely δ_l, which for sufficiently small Λ is certainly small. Also, the garbage we were getting from the UV part of the integral is now to a large extent removed because the loop integrals quickly converge for $q \gg \Lambda$ where the smoothed fields vanish, $\delta_l(q) \propto W_\Lambda(q) \sim e^{-q^2/\Lambda^2} \to 0$. For example

$$P_{13,\Lambda}(k) = 3P_{\rm lin}(k) \int_0^\infty \frac{d^3q}{(2\pi)^3} F_{3,s}(\vec{k}, \vec{q}, -\vec{q}) P_{\rm lin}(q) W_\Lambda(q)^2 \tag{11.19}$$

$$\sim 3P_{\rm lin}(k) \int_0^\Lambda \frac{d^3q}{(2\pi)^3} F_{3,s}(\vec{k}, \vec{q}, -\vec{q}) P_{\rm lin}(q) \,. \tag{11.20}$$

One the flip side, we have introduced a dependence of our result on the arbitrary scale Λ, which should cancel out of any physical observable (see Sec. 5.1). To understand what cancels the Λ dependence, we have to add the EFT corrections

$$\Delta P_{EFT} = -2c^2_{\mathrm{s},\Lambda} k^2 P_{\mathrm{lin}}(k) + P_{JJ,\Lambda}(k) + \dots\,, \tag{11.21}$$

where we made use of the fact that, by definition, ΔJ is the part of the stress-energy tensor that is not correlated with the long modes and so $\langle \Delta J \delta^{(1)} \rangle = 0$. The speed of sound c_s^2 is not predicted by the EFT, but rather is a free parameter. It needs to be measured in an experiment or computed in the full theory to which the EFT provides a long-wavelength expansion. The latter procedure is called *matching* and it is in general hard to perform explicitly because the full theory is usually too difficult to solve[69]. The former procedure requires comparing the EFT prediction $P_{SPT} + \Delta P_{EFT}$ to data or to numerical simulations to obtain the best fitting value for c_s^2 and $P_{JJ,\Lambda}(k)$. Since P_{SPT} depends on Λ, the best fit value obtained by this procedure will also depend on Λ and so we should write

$$P(k) = P_{\mathrm{lin}}(k) + P_{22,\Lambda}(k) + 2P_{13,\Lambda}(k) - 2c^2_{\mathrm{s},\Lambda} k^2 P_{\mathrm{lin}}(k) + P_{JJ,\Lambda}(k) + \dots \tag{11.22}$$

This is the archetypal structure of renormalization: perturbative corrections depend on an arbitrary cutoff Λ, but we can choose the unknown EFT coefficients to also depend on Λ in precisely such a way that the total is Λ independent. The fact that this cancellation is always possible is ensured by having allowed all possible interactions compatible with the symmetry of the problem (the second step for building generic EFT's). Let's see that this works out in our case.

Counterterms Let's go back to the full $P_{13} = P_{13,\Lambda=\infty}$ with all its UV garbage, and let's split it into a contribution from modes below Λ and another contributions from modes above it[70],

$$\begin{aligned} P_{13,\infty}(k) &= 3P_{\mathrm{lin}}(k) \left[\int_0^\Lambda \frac{d^3q}{(2\pi)^3} F_{3,s}(\vec{k},\vec{q},-\vec{q}) P_{\mathrm{lin}}(q) + \int_\Lambda^\infty \frac{d^3q}{(2\pi)^3} F_{3,s}(\vec{k},\vec{q},-\vec{q}) P_{\mathrm{lin}}(q) \right] \\ &= P_{13,\Lambda}(k) - k^2 P_{\mathrm{lin}}(k) \frac{61}{210}\frac{1}{6\pi^2} \int_\Lambda^\infty dq P_{\mathrm{lin}}(q)\,. \end{aligned} \tag{11.23}$$

Clearly $P_{13} = P_{13,\infty}$ does not depend on Λ. Let's hypothetically imagine finding the best fit of c_s^2 for this $\Lambda = \infty$ case, which we'll call $c^2_{s,\infty}$. Clear $c^2_{s,\infty}$ does not depend on Λ either. Then we can write

$$P_{13,\Lambda}(k) - c^2_{s,\Lambda} k^2 P_{\mathrm{lin}}(k) = P_{13,\infty}(k) - c^2_{s,\infty} k^2 P_{\mathrm{lin}}(k)\,, \tag{11.24}$$

where we defined

$$c^2_{s,\infty} = c^2_{s,\Lambda} - \frac{61}{210}\frac{1}{6\pi^2} \int_\Lambda^\infty dq\ P_{\mathrm{lin}}(q)\,. \tag{11.25}$$

[69] For the EFT of LSS a preliminary matching has been studied in [23] using N-body simulations for the full theory, but we will not discuss it here.

[70] Here we have used that the effect of the smoothing function $W_\Lambda(k)$ is roughly equivalent to cutting off the integral at Λ and have used the low-k limit of the integral for the contributions coming from $q > \Lambda \gg k$. In practice different concrete choices of W_Λ give slightly different results. None of this matters for this discussion.

By construction the combination $P_{13,\Lambda} - c^2_{s,\Lambda} k^2 P_{\rm lin}$ does not depends on Λ. In the theory of renormalization, the term $c^2_{s,\Lambda}$ is called a *counterterm*. The role of counterterms is to remove the garbage coming from loop integrals over modes that are outside the control of the EFT. Most often in QFT one finds that the loop integrals are UV divergent and one has to subtract an infinite amount of garbage with infinitely large counterterms. This is OK because neither the counterterm nor the UV part of the integral are physical. They always cancel each other out to leave a finite, cutoff independent and accurate result. Notice that the cancellation in (11.24) required that the EFT correction has exactly the same $k^2 P_{\rm lin}$ dependence as the UV-part of the loop integral. As long as one has written down all possible operators, it must be true to all loop orders that there is always an appropriate counterterm for each garbage contribution from the UV.

Since the EFT prediction does not depend on Λ, we can choose Λ to be whatever is convenient for the calculation. A convenient choice is $\Lambda = \infty$. Then, the residual $c^2_{s,\infty} k^2 P_{\rm lin}(k)$ is a correction to the standard SPT result. Since $c^2_{s,\infty}$ is a free parameter of the effective theory, a so called low-energy constant, it has to be fitted to the data. As one can imagine the same procedure can be carried out again at two loops, which requires including higher order terms in the effective stress tensor and the associated additional counterterms. The result of this procedure is shown in Fig. 25. The plot shows the ratio of the power spectrum $P(k)$ to the linear power spectrum $P_{\rm lin} = P_{11}$ for STP (dashed lines) and the EFT of LSS (continuous lines) at one (red) and two (blue) loops. The data points come from a numerical N-body simulation of dark matter. Several observations are in order. We notice that already after $k \simeq 0.1\, h$ Mpc^{-1} the linear prediction makes a mistake of order a few percent, which grows to 50% by $k \simeq 0.3\, h$ Mpc^{-1}. This was to be expected given that we estimated the non-linear scale around $k_{NL} \simeq 0.3\, h$ Mpc^{-1}. The one-loop of SPT improves agreement with data only marginally compared to linear order and two-loop STP hardly improves the agreement at all. This is a manifestation that STP by itself does *not* provide a systematic expansion that approaches the real answer (not even asymptotically). Conversely, the EFT prediction for the power spectra perform much better. Partially this is due to the fact that the EFT has more fitting parameters. Partially, it relies on the EFT corrections having just the right k dependence to be able to correctly match the data. When performing data analysis, care must be taken not to overfit the data.

The stochastic contribution Let us now discuss the remaining stochastic contribution. So far we haven't said anything about its behaviour, just that it arises from fluctuations in the short modes. These fluctuations can be understood as the reshuffling of matter at short distance changing short scales from the original linear configuration $\delta_s^{(1)}(\mathbf{x})$ to a highly non-linear configuration $\tilde{\delta}_s(\mathbf{x})$. This is a positions space picture and we would like to translate into Fourier space. Imagine that the initial and final configurations $\delta_s^{(1)}(\mathbf{x})$ and $\delta_s(\mathbf{x})$ differ from each other only in a small region of linear size $\Delta_B \ll 1/k_{NL} \ll 1/k$, that is centered around some position $\mathbf{x}_B$. Outside that region the two configurations are identical. Now we can Taylor expand the expression for the stochastic noise

$$\begin{aligned}\delta_J(\vec{k}) &= \int d^3x e^{-i\vec{k}\cdot\vec{x}} \big[\delta_s(\vec{x}) - \delta_s^{(1)}(\vec{x})\big] \qquad (11.26)\\ &= e^{-i\vec{k}\cdot\vec{x}_B} \int_{\Delta_B} d^3y \left[1 + i\vec{k}\cdot\vec{y} - \frac{1}{2}(\vec{k}\cdot\vec{y})^2 + \mathcal{O}((ky)^3)\right] \big[\delta_s(\vec{x}_B+\vec{y}) - \delta_s^{(1)}(\vec{x}_B+\vec{y})\big].\end{aligned}$$

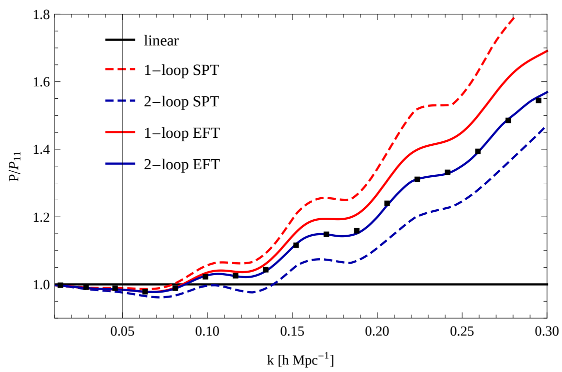


Figure 25: Ratio of the non-linear and linear matter power spectrum. We show one-loop (red) and two-loop (blue) results for SPT (dashed) and EFT (solid). Clearly in SPT there is no notion of convergence as one goes to higher loops. EFT fares much better, providing a good fit to the data up to $k = 0.3\ h\mathrm{Mpc}^{-1}$.

Here we invoke that the spatial reshuffling inside the small region B must obey mass and momentum conservation. Mass conservation ensures that the integral over $\delta_s(\vec{x})$ is the same as that over $\delta_s^{(1)}(\vec{x})$, since they both measure the total mass. Similarly the integrals of $\mathbf{x}$ times $\delta_s(\vec{x})$ is the same as $\mathbf{x}$ times $\delta_s^{(1)}(\vec{x})$ because they both measure the position of the center of mass, which is determined by momentum conservation. This tells us that for $k \ll k_{NL} \ll 1/\Delta_B$, the stochastic noise δ_J starts only at order k^2. We can thus infer that the power spectrum of the stochastic term as $k \to 0$ scales as

$$P_{JJ}(k) \propto k^4 \,. \tag{11.27}$$

The fact that this is the same momentum dependence as that of P_{22} for small wavenumbers (see (10.54)) is both great news and no coincidence. It's great news because it ensures that the counterterm $\delta_{J,\Lambda}$ will cancel out the leading Λ-dependence of $P_{22}(k)$, just as we saw for $c_{s,\Lambda}$ and $P_{13,\Lambda}$,

$$P_{22,\Lambda}(k) + P_{JJ,\Lambda}(k) = P_{22,\infty}(k) + P_{JJ,\infty}(k) \,. \tag{11.28}$$

It is also no coincidence. The reason why the Λ dependence of $P_{22}(k)$ goes as k^4 for small k^4 is that SPT too must conserve mass and momentum.

Finally, we can choose any convenient value of Λ and, just like before, it is often convenient to send it to infinity, i.e., calculate the usual SPT loop integrals and then add the EFT corrections with best-fit coefficients. In the observationally favoured ΛCDM model these integrals are formally convergent, but the fact that they are running over non-perturbative wavenumbers requires a counterterm to capture the unphysical contributions. So our final expression for the one-loop matter power spectrum in the EFT of LSS is

$$P(k) = P_{11}(k) + P_{22}(k) + 2P_{13}(k) - 2c_{s^2,\infty} k^2 P_{11}(k) + P_{JJ,\infty}(k) \,. \tag{11.29}$$

11.4 Power counting and scaling universes

A crucial step in any EFT is to establish a power counting that allows us to determine how many and which operators/interactions should be included to achieve the desired precision for a given observable. We will do this here for the EFT of LSS following [75]).

Scaling universes To begin with we observe that the ΛCDM model is too complex to handle. First, the power spectrum has different power law behaviors at different scales. Second, the dynamics changes dramatically after the cosmological constant being to dominate. To gain more insight we will instead consider a matter-only Einstein-de Sitter universe with a power law initial power spectrum, $P_{\rm lin} = Ak^n$. The SPT equations for a perfect pressureless fluid in EdS are invariant under arbitrary independent rescalings of space and time, $\tau, \mathbf{x} \to \{\lambda_\tau \tau, \lambda_x \mathbf{x}\}$, as long as the fields are also rescaled according to their dimension. In other words, given a solution $\{\delta, \mathbf{v}, \phi\}$ one can always generate infinitely many new solutions $\{\tilde{\delta}, \tilde{\mathbf{v}}, \tilde{\phi}\}$ by

$$\tilde{\delta}(\mathbf{x}, \tau) = \delta(\lambda_x \mathbf{x}, \lambda_\tau \tau)\,, \quad \tilde{\mathbf{v}}(\mathbf{x}, \tau) = \frac{\lambda_\tau}{\lambda_x} \mathbf{v}(\lambda_x \mathbf{x}, \lambda_\tau \tau)\,, \quad \tilde{\phi}(\mathbf{x}, \tau) = \left(\frac{\lambda_\tau}{\lambda_x}\right)^2 \phi(\lambda_x \mathbf{x}, \lambda_\tau \tau)\,. \tag{11.30}$$

This double symmetry exists because there are no scales in the problem with units of time, length or time over length. We used this fact to motivate the factorized power-law ansatz for SPT. This symmetry has also been used in the literature to benchmark the accuracy of N-body simulations of dark matter.
In general the new solution has a different initial power spectrum from the original solution. However, for the specific choice of a power-law initial power spectrum one can choose $\lambda_x = \lambda_\tau^{4/(n+3)}$ in such a way that

$$\tilde{\Delta}^2(k, \tau) = \Delta(k/\lambda_x, \lambda_\tau \tau) = \frac{k^{3+n} A \tau^4}{2\pi^2} \frac{\lambda_\tau^4}{\lambda_x^{3+n}} = \Delta(k, \tau)\,, \tag{11.31}$$

where we use that the linear power spectrum evolves as $P_{\rm lin} \propto a^2 \propto \tau^4$. This is just saying that time evolution acts in the same way as a particular rescaling of coordinates, irrespectively of perturbation theory, which is a very powerful result. Given that our definition of the non-linear wavenumber $k_{\rm NL}$ was $\Delta^2(k = k_{NL}) = 1$, we can write[71]

$$\Delta^2_{\rm lin} \equiv \frac{k^3 P_{\rm lin}}{2\pi^2} = \frac{k^{3+n} A \tau^4}{2\pi^2} = \left(\frac{k}{k_{\rm NL}}\right)^{3+n}\,, \tag{11.32}$$

where $k_{NL} = k_{NL}(\tau) \propto \tau^{-4/(3+n)}$. Because everything is a power law, it is easy to estimate all loops. First we imagine divergences in loop integrals are cancelled by counterterms and we focus on the finite part. Since all loop-integrals are simple power laws, it is straightforward to show that the finite part of loop-integrals simply counts the number of power spectra present. For instance at one-loop we have two power spectra. More generally at l-loops we have [75]:

$$\Delta^2_{l-\rm loop} = \left(\frac{k}{k_{\rm NL}}\right)^{(3+n)(1+l)}\,. \tag{11.33}$$

Now let's discuss the EFT corrections. One might think that they break the scaling symmetry discussed above because now the effective fluid has pressure, a speed of sound

[71] We will consider initial power spectra $P_{\rm lin}(k) = k^n$ with $n > -3$ such that $\Delta^2(k)$ is a growing function of k.

and viscosity. However, these corrections only emerged because we integrated out short scales. The short scales obey the same scaling symmetries as the long scales and so the terms that they generate are also symmetric in this sense. For example the effective coefficients $\{c_s, c_{bv}, c_{sv}\}$ all have units of length over time and so their scaling is fixed. This is particular useful because it fixes their otherwise arbitrary time dependence

$$c_s^2, c_v^2 \propto \left(\frac{\lambda_x}{\lambda_\tau}\right)^2 \propto \tau^{(2-2n)/(n+3)} . \tag{11.34}$$

Using these arguments we can express all EFT corrections as powers of k/k_{NL}. We have seen above that the speed of sound leads to $k^2 P_{\text{lin}}$ and the stochastic term to $P_{JJ} \propto k^4$. Translating these contributions to dimensionless power spectra $\Delta^2 = k^3 P/2\pi^2$ we obtain

$$\Delta^2_{\text{stoch}} = \left(\frac{k}{k_{\text{NL}}}\right)^7 , \qquad \Delta^2_{c_s^2} = \left(\frac{k}{k_{\text{NL}}}\right)^{5+n} . \tag{11.35}$$

This makes it manifest that the expansion parameter of perturbation theory is k/k_{NL} and that all corrections are proportional to this parameter to some power. Higher order corrections have largers powers and are suppressed on very large scales, where k/k_{NL} is very small. This is very similar to the suppression of higher derivative operators in EFT by power of E/E_0, which we discussed in Sec. 5.1. If one desires prediction with a given precision at some particular value of k/k_{NL} one should just include terms up to some maximum finite order in k/k_{NL}.

Relative importance and ΛCDM To make this more concrete, the exponents of k/k_{NL} for all the terms we have discussed so far is shown in Fig. 26, as a function of the power-law slope n in the initial conditions. The lines emanating from the point $n = -3$ with exponent 0 are the tree level and loop corrections, which have increasing exponents for increasing loop order for any $n > -3$. The other lines are the higher derivative corrections (parallel to tree level) and the stochastic noise (horizontal). For a given n one should include all the terms up to some maximum exponent, without skipping any interaction. For example, for $n = 0$ it would be inconsistent to include the two loop correction, with exponent 9 and not including the stochastic noise, with exponent 7.

Does this discussion of scaling universe have any relevance for the real world of ΛCDM? It turns out that the answer is yes. In particular, the prescription is that we should choose n to correspond to the effective slope of the linear power spectrum at the non-linear wavenumber k_{NL}, which is roughly $n = -3/2$, and use that to estimate the relative size of terms. The reason for this is that, after renormalization, the net contribution form higher order terms comes approximately from around the non-linear scale and so the slope there is the most relevant. Considering this, the ordering of importance in our universe would be:

$$\text{linear} > \text{1-loop} > c_s^2 > \text{2-loop} > \text{higher } \partial\text{'s} > \text{stochastic term} \tag{11.36}$$

For practical purposes it is thus safe to ignore P_{JJ}.

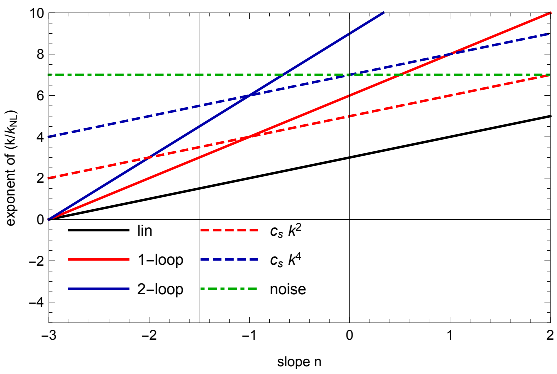


Figure 26: Effective slopes of the contributions to the power spectrum in a scaling universe with initial slope n.

Part III
Cosmic Microwave Background

The Cosmic Microwave Background is a sort of baby picture of the Universe, that allows us a first glimpse of the young Universe when it was only 380.000 years old.[72] Up to last scattering, the Universe was opaque due to the high density of free electrons. Once the temperature had fallen sufficiently (to about 3000 K), electrons (re)combined with the nuclei and formed neutral hydrogen (and helium). Thus, fairly suddenly the photons were able to propagate freely all the way until they reach our telescopes today. During the free-streaming the photons conserve the blackbody spectrum, with temperature decreasing as $1/a$ until today's value of $T_{CMB} = 2.7255$ K.

We will first derive an evolution equation for the photon temperature. It will turn out that we will mostly be interested in the monopole and dipole of the photons, i.e., we will obtain the equations of motion that resemble those of a fluid. We will then have to relate the temperature that the observer sees in direction $\hat{n}$ to the intrinsic temperature of the CMB photons at the time of last scattering. We will see that the photon perturbations undergo acoustic oscillations as they enter the horizon and the phase of these oscillations at the time of last scattering will yield the oscillatory fluctuation spectrum. We will then go to the next-to-leading order in the tight coupling approximation to derive the damping of fluctuations on small scales. Having derived the temperature fluctuations, we will then consider the role of polarization and calculate the effect of gravitational lensing on the intrinsic CMB fluctuations.
We will be considering the following effects:

- primordial anisotropies (Sachs-Wolfe effect, Doppler effect, polarization, damping - linear)
- secondary anisotropies (integrated Sachs-Wolfe effect, Sunayev-Zel'dovich effect, gravitational lensing - mostly non-linear)

As a reference for this Chapter, see the textbook by Dodelson [36] and for a more detailed treatment the book by Durrer [41]. The review article [24] is based on the CMB chapters of Advanced Cosmology as previously taught by Prof. Challinor. Pedagogical explanations are also provided in the review article [54].

12 Boltzmann Equation

We shall use kinetic theory to describe the transition from the fluid-like, pre-recombination tightly coupled regime to the free-streaming radiation. The CMB photons can be described by the one-particle distribution function $f(x^\mu, p^\mu)$, which is a function of the spacetime position and four-momentum. The distribution function is Lorentz-invariant and, in the absence of scattering, it is conserved along the photon path in phase space, just like it was for collisionless dark matter.
The main difference now is that photons scatter off charged particles and the universe is filled with electrons and protons. Now photons can move from one position in phase space before a collision to another position after the collision. As a consequence, the Boltzmann equation changes and features a collision term C as a source on the right-hand side. The collision term is a functional of the photon distribution function f and

[72] If the Universe was one year old at the time of last scattering, it would soon be celebrating it's 35th birthday.

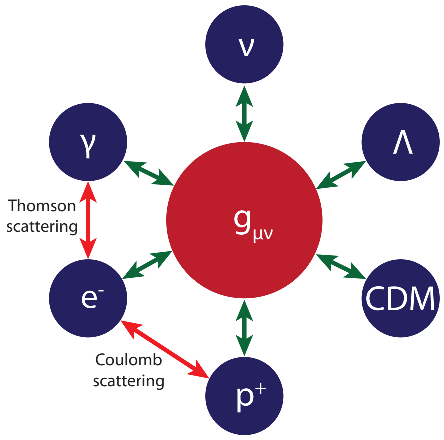


Figure 27: Interaction between the perturbations in the various species in the universe. All of the species are coupled to the metric. Coulomb-scattering couples the electrons and protons, and the photons are Thomson scattering on free electrons before recombination. Adapted from [36].

the electron distribution function f_e

$$\frac{df}{d\eta} = C[f, f_e]\,. \tag{12.1}$$

In the following we will study the left- and right-hand side of this equation in turn. To massage the left-hand side into something manageable we will follow the same strategy as for dark matter. We will first study the motion of a single photon and we will then use this solution to simplify the terms obtained when expanding $df/d\eta$ into partial derivatives.

12.1 Collisionless evolution

Here we derive how to compute the left-hand side of the Boltzmann equation for photons in a linearly perturbed universe described by general relativity. We will use the metric in Newtonian gauge[73]

$$ds^2 = -a^2(1+2\Psi)d\eta^2 + a^2(1-2\Phi)\delta_{ij}dx^i dx^j\,, \tag{12.2}$$

where we will refer to Φ and Ψ as Newtonian potentials[74]. Throughout we will work to linear order in Φ and Ψ. The Einstein Equations in Newtonian gauge are given by

$$-k^2\Phi - 3\mathcal{H}(\Phi' + \mathcal{H}\Phi) = 4\pi G a^2 \delta\rho_{\rm N}, \tag{12.3}$$

$$\Phi' + \mathcal{H}\Psi = 4\pi G a^2\left(\bar\rho + \bar p\right) v_{\rm N}, \tag{12.4}$$

$$k^2(\Phi - \Psi) = 8\pi G a^2 \Sigma. \tag{12.5}$$

[73]This gauge is sometimes called "conformal" Newtonian gauge to remind the reader that the Newtonian potentials Φ and Ψ have been defined with a factor of a^2 in front.

[74]Note that $\Phi_{\rm here} = -\Phi_{\rm Dodelson}$, while $\Psi_{\rm here} = \Psi_{\rm Dodelson}$.

Here Σ is called anisotropic stress and is negligibly small for the components of our universe[75]. Assuming $\Sigma = 0$ the last equation implies $\Phi = \Psi$.

Photon's geodesics Now we want to derive the equations of motion for the free propagation of a photon in an expanding FLRW universe that is perturbed as in (12.2). The photon 4-momentum $P^\mu = dx^\mu/d\lambda$ is subject to the on-shell constraint

$$g_{\mu\nu}P^\mu P^\nu = -a^2(1+2\Psi)(P^0)^2 + p^2 = 0\,, \tag{12.6}$$

where

$$p^2 = g_{ij}P^iP^j = a^2(1-2\Phi)\delta_{ij}P^iP^j\,. \tag{12.7}$$

This tells us that the temporal component P^0 is not an independent quantity but can be expressed as

$$P^0 = \frac{p}{a}(1-\Psi)\,. \tag{12.8}$$

We can now factorize the spatial part of the momentum into a normalized direction $\hat{p}^i$ satisfying $\delta_{ij}\hat{p}^i\hat{p}^j = 1$ and an amplitude that is fixed by the constraint in (12.7). Finally, we obtain for the 4-momentum in terms of the *comoving energy* $\epsilon = pa$

$$P^\mu = \frac{\epsilon}{a^2}\left(1-\Psi,(1+\Phi)\hat{p}^i\right)\,. \tag{12.9}$$

To know how ϵ and $\hat{p}^i$ evolve with time we need to solve the *geodesic equation*

$$\frac{dP^\mu}{d\lambda} + \Gamma^\mu_{\alpha\beta}P^\alpha P^\beta = 0\,. \tag{12.10}$$

We can express the derivative with respect to the affine parameter λ as

$$\frac{dP^\mu}{d\lambda} = \frac{dP^\mu}{d\eta}\frac{d\eta}{d\lambda} = P^0\frac{dP^\mu}{d\eta}\,. \tag{12.11}$$

Using the Christoffel symbols[76] for the metric Eq. (12.2), we find that the $\mu = 0$ component of the geodesic equations becomes

$$\begin{aligned}\frac{\epsilon}{a^2}(1-\Psi)\partial_\eta\left[\frac{\epsilon}{a^2}(1-\Psi)\right] =& -(\mathcal{H}+\Psi')\frac{\epsilon^2}{a^4}(1-\Psi)^2 - 2\partial_i\Psi\hat{p}^i\frac{\epsilon^2}{a^4}(1-\Psi)(1+\Phi)\\ &-\left[\mathcal{H}-\Phi'-2\mathcal{H}(\Psi+\Phi)\right]\frac{\epsilon^2}{a^4}(1+\Phi)^2\end{aligned} \tag{12.13}$$

To zeroth order this is solved by $\partial_\eta\epsilon = 0$, which is simply saying that a photon's energy and momentum in an unperturbed FLRW universe redshift with the expansion $p \propto 1/a$. Also, the i component of the geodesic equation implies $d\hat{p}^i/d\eta = 0$ to zeroth order,

[75] An interesting exception are neutrinos. This has relevant implication for the propagation of tensor modes [101]. The effects on scalars are relatively less important and will be neglected here.

[76] These are

$$\Gamma^0_{00} = \mathcal{H}+\Psi' \qquad \Gamma^0_{0i} = \partial_i\Psi \qquad \Gamma^0_{ij} = \left[\mathcal{H}-\Phi'-2\mathcal{H}(\Psi+\Phi)\right]\delta_{ij}\,. \tag{12.12}$$

which is expected since there is no preferred direction for the photon to turn into. To first order we get the remarkably short equation[77]

$$\partial_\eta \ln \epsilon = -\hat{p}^i \partial_i \Psi + \Phi' \,. \tag{12.14}$$

Note that the change of $\ln \epsilon$ does not depend on ϵ, which in word tells us that gravitational redsfhit is a-chromatic (independent of frequency). This leads to the important simplification that the first order Boltzmann equation does not depend on ϵ.
The result in (12.14) has a nice interpretation. First, notice that if Ψ (and Φ) are evaluated along the trajectory $\vec{x}(\eta)$ of a photon, we can introduce a total "convective" (sometimes also "advective" or "material") derivative by

$$\frac{d\Psi}{d\eta} = \Psi' + \hat{p}^i \partial_i \Psi \,. \tag{12.15}$$

In terms of the convective derivative the above solutions becomes

$$\partial_\eta \ln \epsilon = -\frac{d\Psi}{d\eta} + \left(\Psi' + \Phi'\right) \,, \tag{12.16}$$

where $\Psi = \Psi(\eta, \vec{x}(\eta))$ and similarly for Φ. This tells us that the photon's energy changes due to changes in the gravitational potential along the photon's trajectory (first term) and the time dependence of the gravitational potentials (last two terms). The latter effect is suppressed during matter domination when the potentials are constant on large scales, but contributes at early times during the matter-radiation transition and at late times when dark energy starts to dominate the Universe.

Phase space evolution As we did for dark matter, we can rewrite the total time derivative in the Boltzmann equation in terms of the partial derivatives with respect to the dependencies of the distribution function

$$\frac{df}{d\eta} = \frac{\partial f}{\partial \eta} + \frac{\partial f}{\partial x^i}\frac{dx^i}{d\eta} + \frac{\partial f}{\partial \ln \epsilon}\frac{d \ln \epsilon}{d\eta} + \frac{\partial f}{\partial \hat{p}^i}\frac{d\hat{p}^i}{d\eta} = C[f] \,. \tag{12.17}$$

Note that here we have made the choice to consider $f(\ln \epsilon)$ rather than $f(\epsilon)$ because, as we noticed, gravitational redshift is achromatic. We notice that the last term is second order in perturbations because in an unperturbed universe the photon's direction does not change with time and f does not depend on the direction of the photon. Since we limit our discussion to first order, this term can be dropped. The comoving time derivative of the photon's position can be calculated from the momentum as

$$\frac{dx^i}{d\eta} = \frac{dx^i}{d\lambda}\frac{d\lambda}{dx^0} = \frac{P^i}{P^0} = (1 + \Phi + \Psi)\hat{p}^i \,. \tag{12.18}$$

Since $\partial_{x^i} f$ is already first order, we can ignore the metric perturbations in the above equation, and simply use $\partial_\eta x^i = \hat{p}^i$. Furthermore, since comoving energy is conserved at zeroth order and $d \ln \epsilon / d\eta$ starts at first order, we can evaluate $\partial f / \partial \ln \epsilon$ at zeroth order, $f = \bar{f}$. Finally, we obtain for the left-hand side of the Boltzmann equation up to first-order terms as

$$\frac{\partial f}{\partial \eta} + \hat{p} \cdot \vec{\nabla} f + \frac{\partial \bar{f}}{\partial \ln \epsilon}\frac{d \ln \epsilon}{d\eta} = C \,. \tag{12.19}$$

[77] To first order the direction changes as $\partial_\eta \hat{p}^i = -\left(\partial_i - \hat{p}^i\, \hat{p} \cdot \partial\right)(\Phi + \Psi)$. This is important to compute the lensing of the CMB from the intervening matter, but we will not need it to compute the primary CMB anisotropies.

Box 12.1 Lorentz-invariance of the phase-space density. Let us establish Lorentz-invariance of the phase space density by considering a number of massive particles occupying a phase-space volume $dN = f d^3x d^3p$. Another observer will see the *same number* of particles in a different phase-space volume $dN = f' d^3x' d^3p'$. We will establish that the phase-space volume is Lorentz-invariant and thus $f(x,p) = f'(x',p')$.
As you will be well aware from special relativity or QFT, the on-shell, Lorentz-invariant measure for momentum integrals can be written as

$$\int d^4p\ \delta^{(\mathrm{D})}\left(E^2 - \vec{p}^2 - m^2\right) = \int \frac{d^3p}{2E(p)}\,, \tag{12.20}$$

where we have used

$$\delta^{(\mathrm{D})}\left(E^2 - \vec{p}^2 - m^2\right) = \frac{1}{2E}\delta^{(\mathrm{D})}\left(E - \sqrt{p^2 + m^2}\right)\,. \tag{12.21}$$

Under a Lorentz transformation we have due to length contraction that $d^3x' = d^3x/\gamma$ and the energy scales as $E' = \gamma E$. Thus Ed^3x is Lorentz-invariant. Multiplying the Lorentz-invariant quantities Ed^3x and d^3p/E, we see that d^3xd^3p has to be Lorentz-invariant. Since the density of particles in phase space must also remain the same, we conclude that the phase space density transforms as a scalar, i.e. $f(x,p) = f'(x',p')$.

Zeroth order and the CMB dipole Now we would like to solve this equation up to first order. Let's start at zeroth order, i.e. the distribution of photons in a homogeneous and isotropic FLRW background. As we will show later in Sec. 12.2, the collision term starts at first order, so does not contribute to zeroth order. In the absence of conserved charges, we already know that the thermodynamic equilibrium distribution of non-interacting bosonic particles is the *Bose-Einstein distribution* with homogeneous temperature $\bar{T}$ and vanishing chemical potential, sometimes also referred to as black body distribution,

$$\bar{f}(\epsilon) = \frac{1}{e^{p/\bar{T}(\eta)} - 1} = \frac{1}{e^{\epsilon/a\bar{T}(\eta)} - 1} \quad \text{(zeroth order)}\,, \tag{12.22}$$

Plugging this into the zeroth order Boltzmann equation one find that it is solved if

$$\bar{T}(\eta) = T_0/a(\eta)\,. \tag{12.23}$$

Before discussing first order, it is worth mentioning an interesting effect visible in the CMB already at zeroth order. The solar system and the earth within it are moving with respect to the CMB rest frame. This motion introduces the largest anisotropy in the CMB in the form of a dipole modulation along the direction of relative motion. More quantitatively, the speed of the solar system's barycenter with respect to the CMB rest frame is about $v_\odot \simeq 3.68 \times 10^5$m/s. This introduces a dipolar modulation of the CMB temperature of order v/c, which is about 0.1%. Since today's CMB temperature is $T_0 \sim 2.725$ K we expect a dipole of order 3 mK (milli-Kelvin). Let's work this out.

We expect a dipole of order 10^{-3}, while anisotropies from primordial perturbations are or order 10^{-5}. Therefore, we can discuss the dipole to zeroth order in perturbations, i.e. for $\Theta = 0$. The idea of the calculation is to consider a Lorentz boost from the CMB reference frame, where the CMB is isotropic, to the earth's reference labelled by "$\odot$".

This changes a photons comoving energy and direction from $\{\epsilon, \hat{p}\}$ to $\{\epsilon_\odot, \hat{p}_\odot\}$ with

$$\epsilon = \gamma\epsilon_\odot(1 + \hat{p}_\odot \cdot \vec{v}_\odot)\,, \tag{12.24}$$

where $\gamma = (1-v_\odot^2)^{-1/2}$ is the Lorentz factor. Since the distribution function f transforms as a scalar (see derivation in the box), we can find $f_\odot$ for a boosted observer simply by evaluating $f(x(x_\odot), p(p_\odot))$. As observed above it suffices to consider the zeroth order unperturbed distribution function f, which actually does not depend on $\mathbf{x}$ or on $\hat{p}$ but just on the photon's energy ϵ. Then the boosted distribution is

$$f_\odot(\epsilon_\odot, \hat{p}_\odot) = \bar{f}(\epsilon(\epsilon_\odot, \hat{p}_\odot)) \tag{12.25}$$

$$= \left[e^{\frac{\epsilon}{a\bar{T}(\eta)}} - 1\right]^{-1} = \left[e^{\frac{\gamma\epsilon_\odot(1+\hat{p}_\odot\cdot\vec{v}_\odot)}{T_0}} - 1\right]^{-1}\,. \tag{12.26}$$

It follows that the effective temperature of the CMB photons is related to the underlying homogeneous temperature by

$$T_\odot = \frac{T_0}{\gamma(1 + \hat{p}_\odot \cdot \vec{v}_\odot)} \approx T_0(1 - \hat{p}_\odot \cdot \vec{v}_\odot)\,. \tag{12.27}$$

This describes a temperature monopole T_0 superimposed to a temperature dipole where the CMB temperature on earth depends on the angle of arrival of photons.

First order At first order in perturbations, the distribution f can in principle depend on both position x^i, and on the direction $\hat{p}^i$ and the energy ϵ of photons at that point. Because gravitational redshift of $(\ln\epsilon)$ is acromatic, (12.14), we can actually choose variables such that the ϵ-dependence is solved automatically from the get to (this simplification arises only at first order). To see this, let's introduce the *temperature perturbation* $\Theta(x^i, p^i, t)$ by

$$T(x^i, p^i, \eta) = \bar{T}(\eta)\left[1 + \Theta(x^i, p^i, \eta)\right]\,, \tag{12.28}$$

where for $\Theta = 0$ we recover the zeroth order distribution $\bar{f}$ and we will work to linear order in Θ. Expanding the distribution function around the equilibrium distribution function $\bar{f}$ we get

$$f(\eta, \vec{x}, \hat{p}, \epsilon) = \frac{1}{e^{\frac{\epsilon}{a\bar{T}(1+\theta)}} - 1} \simeq \frac{1}{e^{\frac{\epsilon(1-\theta)}{a\bar{T}}} - 1} \simeq \bar{f}(\epsilon)\left[1 - \Theta(\eta, \vec{x}, \hat{p})\frac{\partial \ln \bar{f}}{\partial \ln \epsilon}\right]\,. \tag{12.29}$$

This tells us that our choice of parameterization of perturbations to f We thus have for the left-hand side of the Boltzmann equation

$$-\frac{\partial \bar{f}}{\partial \ln \epsilon}\left[\frac{\partial \Theta}{\partial \eta} + \hat{p}\cdot\vec{\nabla}\Theta - \frac{d\ln\epsilon}{d\eta}\right] = C\,. \tag{12.30}$$

The time dependence of the photon energy at first order, namely $d\ln\epsilon/d\ln\eta$, was found in (12.14) by solving the geodesic equations.

Box 12.2 Spectral Distortions* The spectrum of the CMB has been observed by the FIRAS instrument on the COBE satellite in 1989/90. The measurements are shown in Fig. 28 together with the Planck blackbody spectral distribution function [44]

$$I_\nu = \frac{2h\nu^3}{c^2}\frac{1}{e^{h\nu/k_\mathrm{B}T}-1}\,. \tag{12.31}$$

While the CMB looks like an almost perfect blackbody, there can be small residual deviations that originate either from the physics of the early Universe or from the propagation of the photons between the last scattering surface to the observer. In the early Universe the blackbody is maintained by Compton scattering ($e^- + \gamma \leftrightarrow e^- + \gamma$), double Compton scattering ($e^- + \gamma \leftrightarrow e^- + \gamma + \gamma$) and bremsstrahlung ($e^- \to e^- + \gamma$ in the presence of an electro-magnetic field). Due to the expansion of the Universe, these effects become more and more inefficient. Around redshift $z \approx 2 \times 10^6$, the rate of double Compton scattering and bremsstrahlung drops below the expansion rate, while Compton scattering is still very efficient. Since the only processes that can change the number of photons are too slow to maintain equilibrium, the number of photons becomes effectives conserved (very soft photons can always be created, but here we focus on photons whose energy is of the order of the temperature of the plasma). Any energy injections after $z \approx 2 \times 10^6$ hence lead to a spectra distortion. If the energy injections takes place when Compton scattering is still efficient at redistributing photons across different energies, namely in the window $5 \times 10^4 < z < 2 \times 10^6$, then the spectrum relaxes to theormodynamical equilibrium with a conserved charge, namely to

$$I_\nu = \frac{2h\nu^3}{c^2}\frac{1}{e^{h\nu/k_\mathrm{B}T+\mu}-1}\,. \tag{12.32}$$

where the so-called μ-distortions can be interpreted as a dimensionless chemical potential for the approximately conserved number of photons. If the energy injections takes place after $z \approx 5 \times 10^4$, Compton scattering is not sufficiently efficient to maintain kinetic equilibrium and the photon distribution can in principle be modified arbitrarily. When the photons receive energy from another sector that is approximately thermalised, but at a different temperature, then one finds a so-called y-type distortions. An example of this is the thermal Sunyaev-Zel'dovich effect, which describes the scattering of CMB photons off hot electrons within galaxy clusters. Spectral distortions are extremely tightly constrained, with limits $|\mu| < 9 \times 10^{-5}$ for the chemical potential and $|y| < 1.5 \times 10^{-9}$ for Compton y-distortions. Furthermore, galactic foregrounds contaminate measurements of the primary CMB. Since the foregrounds and secondary effects have a spectral distribution that differs from the primary CMB blackbody, current CMB experiments observe in a number of spectral bands and combine those measurements to remove contaminations.

12.2 The collision term

The rate at which a unit flux of (unpolarized) photons scatters off a charged particle at rest is the differential cross section and is given by

$$\frac{d\sigma}{d\Omega} = \frac{3\sigma_\mathrm{T}}{16\pi}\left[1+\cos\theta^2\right]\,. \tag{12.33}$$

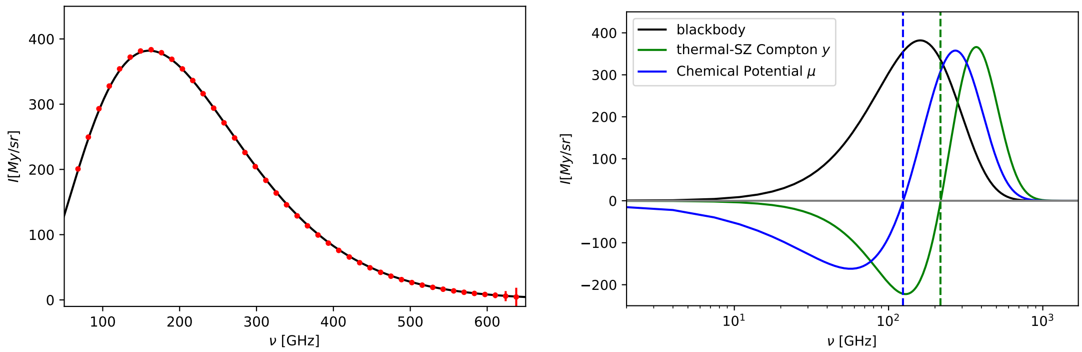


Figure 28: *Left panel:* Spectral density of the CMB monopole as measured by COBE/FIRAS and the black body spectrum corresponding to $T_{\rm CMB} = 2.7255$ K. Error bars have been enhanced by a factor of 50 for clearer visibility. *Right panel:* Spectral distortions arising from energy injection in the early universe (μ-distortions) or the thermal Sunayev-Zel'dovich effect (Compton y-distortion).

where θ is the scattering angle experienced by the photons. Here $\sigma_{\rm T}$ is the Thomson cross section[78]

$$\sigma_{\rm T} = \int d\Omega \frac{d\sigma}{d\Omega} = \frac{8\pi}{3}\left(\frac{q}{4\pi\epsilon_0 mc^2}\right)^2 . \tag{12.34}$$

with q the charge and m the mass. The charged particles around much after the QCD phase transition are mostly electrons and protons. Since the Thomson cross section goes as $1/m^2$ and the electron's mass is 2000 times small than the protons, we can focus on just scattering off electrons, for which we find $\sigma_{\rm T} \simeq 6.65 \times 10^{-29}{\rm m}^2$. The number density of *free* electrons n_e is given by the product of baryon density and ionization fraction $n_{\rm e} = n_{\rm b} x_{\rm e}$. For future convenience we define the *scattering rate*

$$\Gamma = a\sigma_{\rm T} n_e \, . \tag{12.35}$$

Because the scattering rate depends on the number density of free electrons, we have to consider carefully how it changes. Recombination is the time when the free electrons combine with free protons to form neutral hydrogen. Scattering off electrons is important before recombination and shuts off at recombination when the universe becomes neutral. After recombination, the photons propagate through a transparent universe and their dynamics is captured by free streaming, a solution of the collisionless Boltzmann equation, which will be discussed later. Here we derive collision term implied by photons scattering off electrons. We will ignore the angular dependent term in (12.33) as it is not necessary unless percent level accuracy is required. Furthermore, we will assume photons are unpolarized. This is a reasonable approximation when discussing temperature anisotropies. We will re-introduce the polarization dependence in Sec. **??**, when considering CMB polarization. Now we proceed by first working in the electron rest frame, and then in a general frame where electrons may be moving.

[78]Thompson scattering is the low-energy limit of the more general Compton scattering captured by the Klein-Nishina formula. When the photon's energy E is much smaller than the charged particle's mass, typically an electron, the momentum transfer is of order $E/m \ll 1$ and so the scattering is approximately elastic and the photon does not change its energy.

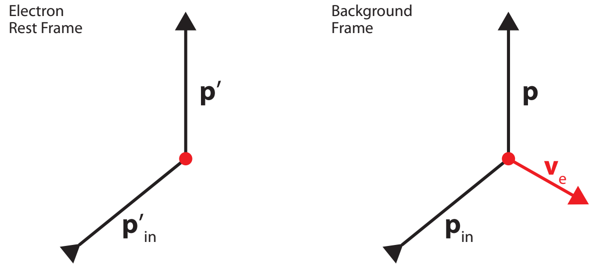


Figure 29: Thomson scattering in the electron rest frame (left) and background frame (right).

Electron rest frame We would like to consider the effect of the following collision on the Boltzmann equation

$$e^-(\vec{q}_{\rm in}) + \gamma(\vec{p}_{\rm in}) \leftrightarrow e^-(\vec{q}) + \gamma(\vec{p})\,. \tag{12.36}$$

We will start considering the rest frame of the electron and we will denote with a prime any 4-momentum evaluated in this frame (not to be confused with a derivative with respect to conformal time). Then let's denote the ingoing and outgoing photon's direction by

$$\vec{p}^{\,\prime}_{\rm in} = \epsilon'_{\rm in}\hat{p}'_{\rm in} \quad\to\quad \vec{p}^{\,\prime} = \epsilon'_{\rm in}\hat{p}'\,, \tag{12.37}$$

where we used that Thomson scattering does not change the energy of the photon, such that the energy remains the same after the collision. The collision term describes the difference between photons scattered into the phase space volume and the photons scattered out of the phase space volume. In its most general form it can be written as

$$C'[f_\gamma(p')] = \int \frac{d^3q_{\rm in}}{2E_e(q_{\rm in})}\int\frac{d^3q}{2E_e(q)}\int\frac{d^3p'_{\rm in}}{2E(p'_{\rm in})}|\mathcal{M}|^2\left[f_e(q'_{\rm in})f_\gamma(p'_{\rm in}) - f_e(q')f_\gamma(p')\right]$$

where the energy of the non-relativistic electron is $E_e(q) \simeq m_e + q^2/(2m_e)$. The dynamics of Thomson scattering is such that $f_e(p) \approx f_e(p')$ because the electron's energy is so much bigger than the photon's [36]. The dimensionless matrix element is $|\mathcal{M}|^2 = \sigma_{\rm T}m_e^2$. We have to average over all possible directions for the in-scattering

$$\begin{aligned} C'[f'(\epsilon',\hat{p}')] = \frac{df'}{d\tau'} =& n_e\int d\hat{p}'_{\rm in}\frac{d\sigma}{d\Omega}\left[f'(\epsilon',\hat{p}'_{\rm in}) - f'(\epsilon',\hat{p}')\right] \\ =& -n_e\sigma_{\rm T}f'(\epsilon',\hat{p}') + n_e\sigma_{\rm T}\int\frac{d\hat{p}'_{\rm in}}{4\pi}f'(\epsilon',\hat{p}'_{\rm in}) \end{aligned} \tag{12.38}$$

This is the scattering rate with respect to proper time in the electron rest frame.

Background frame The collision term in Eq. (12.1) describes the scattering per unit conformal time, whereas the collision term in the rest frame was computed for unit proper time. We thus have for their relation

$$C = \frac{df}{d\eta} = a\frac{df}{d\tau} = aC'\,. \tag{12.39}$$

We can now transform to a generic frame in which the electrons are moving with velocity $\vec{v}_e$ by performing a Lorentz boost

$$\vec{p}_{\rm in} = \epsilon_{\rm in}\hat{p}_{\rm in} \quad \rightarrow \quad \vec{p} = \epsilon\hat{p}\,. \tag{12.40}$$

with

$$\epsilon' = \gamma\epsilon(1 - \hat{p}\cdot\vec{v}_e)\,, \qquad\qquad \epsilon = \gamma\epsilon'(1 + \hat{p}'\cdot\vec{v}_e)\,. \tag{12.41}$$

We can use $\epsilon' = \epsilon'_{\rm in}$ to obtain

$$\begin{aligned}\epsilon_{\rm in} =&\gamma\epsilon'_{\rm in}\left(1 + \hat{p}'_{\rm in}\cdot\vec{v}_e\right) = \gamma\epsilon'\left(1 + \hat{p}'_{\rm in}\cdot\vec{v}_e\right)\\ =&\gamma^2\epsilon\left(1 + (\hat{p}'_{\rm in} - \hat{p})\cdot\vec{v}_e\right)\end{aligned} \tag{12.42}$$

Using the invariance of the phase space density and expanding around the blackbody distribution function we find

$$f'(\epsilon'_{\rm in},\hat{p}'_{\rm in}) = f(\epsilon_{\rm in},\hat{p}_{\rm in}) = \bar{f}(\epsilon) - \frac{d\bar{f}}{d\ln\epsilon}(\hat{p} - \hat{p}'_{\rm in})\cdot\vec{v}_e - \frac{d\bar{f}}{d\ln\epsilon}\Theta(\hat{p}_{\rm in}) \tag{12.43}$$

$$f'(\epsilon',\hat{p}') = f(\epsilon,\hat{p}) = \bar{f}(\epsilon) - \frac{d\bar{f}}{d\ln\epsilon}\Theta(\hat{p}) \tag{12.44}$$

We can now use the above result to evaluate the collision term

$$\begin{aligned}C[f(\epsilon,\hat{p})] =& -\Gamma f'(\epsilon',\hat{p}') + \Gamma\int\frac{d\hat{p}'_{\rm in}}{4\pi}f'(\epsilon',\hat{p}'_{\rm in})\\ =& -\Gamma\bar{f}(\epsilon) + \Gamma\frac{d\bar{f}}{d\ln\epsilon}\Theta(\hat{p}) + \Gamma\int\frac{d\hat{p}'_{\rm in}}{4\pi}\left[\bar{f}(\epsilon) - \frac{d\bar{f}}{d\ln\epsilon}(\hat{p} - \hat{p}'_{\rm in})\cdot\vec{v}_e - \frac{d\bar{f}}{d\ln\epsilon}\Theta(\hat{p}_{\rm in})\right]\\ =&\Gamma\left[\frac{d\bar{f}}{d\ln\epsilon}\Theta(\hat{p}) - \frac{d\bar{f}}{d\ln\epsilon}\hat{p}\cdot\vec{v}_e - \frac{d\bar{f}}{d\ln\epsilon}\int\frac{d\hat{p}'_{\rm in}}{4\pi}\Theta(\hat{p}_{\rm in})\right]\,,\end{aligned}$$

where $\Gamma = a\sigma_{\rm T}n_e$. In the last term we can replace $d\hat{p}'_{\rm in} \rightarrow d\hat{p}_{\rm in}$ at zeroth order. The angular average then yields the *monopole* Θ_0 of the temperature distribution

$$\Theta_0(\vec{x},\eta) = \int\frac{d^2\hat{p}}{4\pi}\Theta(\vec{x},\hat{p},\eta)\,. \tag{12.45}$$

In summary we have for the collision term

$$C = \Gamma\frac{d\bar{f}}{d\ln\epsilon}\left(\Theta - \Theta_0 - \hat{p}\cdot\vec{v}_{\rm e}\right)\,. \tag{12.46}$$

Had we kept the angular dependence the Thomson cross-section we would have obtained

$$C = \frac{d\bar{f}}{d\ln\epsilon}\Gamma\left\{\Theta - \hat{p}\cdot\vec{v}_{\rm e} - \frac{3}{16\pi}\int d^3\hat{p}_{\rm in}\Theta(\hat{p}_{\rm in})\left[1 + (\hat{p}_{\rm in}\cdot\hat{p})^2\right]\right\}\,. \tag{12.47}$$

Combining the collision term with the left hand side of the Boltzmann equation (12.57), we get

$$\frac{d\Theta}{d\eta} = \frac{d\ln\epsilon}{d\eta} - \Gamma\left[\Theta - \Theta_0 - \hat{p}\cdot\vec{v}_e\right]\,. \tag{12.48}$$

At early times, when the scattering rate is high $\Gamma \gg \mathcal{H}$, the second term needs to vanish. This means that the scattering makes the radiation isotropic in the electron rest frame and leads to a distribution with a monopole and dipole in a generic frame $\Theta \to \Theta_0 + \hat{p}\cdot\vec{v}_e$. Combining the geodesic equation (12.14) and collision term we obtain

$$\frac{\partial\Theta}{\partial\eta} + \hat{p}\cdot\vec{\nabla}\Theta = \Phi' - \hat{p}\cdot\vec{\nabla}\Psi - \Gamma\left[\Theta - \Theta_0 - \hat{p}\cdot\vec{v}_e\right] . \tag{12.49}$$

Restricting ourselves to curl-free velocities[79] $\vec{v}_e(\vec{k}) = i\hat{k}v_e(\vec{k})$ we can rewrite the above equation in Fourier space as

$$\Theta' + ik\mu\Theta = \Phi' - ik\mu\Psi + \Gamma\left[\Theta_0 - \Theta + i\mu v_{\rm e}\right] , \tag{12.50}$$

where $\mu = \hat{k}\cdot\hat{p}$.

12.3 Free streaming and the line-of-sight Solution

Solving Eq. (12.49) is challenging because it is a partial differential equation in six variables. Fortunately for us, there are two regimes in which we can find solutions analytically. The first regime is when photons interact very often with charged particles, which are electrons and protons in our case. In this regime all the constituents of the system combine together to form a charged fluid which is called the photon-electron-baryon plasma. This regime can be tackled using perturbation theory and hydrodynamic equations, as we will see in Section 13.2. The second regime is somewhat the opposite, namely when the photons don't interact at all but free stream along FLRW geodesics without collisions. The corresponding solution is called the line-of-sight solution and was introduced in [88]. The line-of-sight solution captures the evolution of CMB photons after recombination, which is when the universe becomes almost completely transparent and photons travel undisturbed. To derive this formal solution we have to first introduce two quantities, the optical depth and visibility function.

The *optical depth* is defined by the integral

$$\tau = \int_{\eta}^{\eta_0} d\eta'\ \Gamma(\eta') , \tag{12.51}$$

where $\Gamma = an_e\sigma_T$ is the Thomson scattering rate defined in (12.35), which is simply the inverse of the mean free path of photon. Note that Γ has units of inverse time, as appropriate for a rate, and so τ is dimensionless. The physical interpretation is that $e^{-\tau}$ corresponds to the probability for a photon not to scatter between conformal time η and η_0, where η_0 refers to today. To see why this is the case, let $\Delta\eta$ be an infinitesimal interval of time and $\text{Prob}(\eta)$ the probability that a photon has not scattered between η and η_0. Working to linear order in $\Delta\eta$ we can say that $\text{Prob}(\eta+\Delta\eta)$ is bigger than $\text{Prob}(\eta)$ because of the fraction $\Gamma(\eta)\Delta\eta$ of photons that do scatter in that interval of time,

$$\text{Prob}(\eta+\Delta\eta) = \text{Prob}(\eta)\left(1 + \Gamma(\eta)\Delta\eta\right) . \tag{12.52}$$

Rearranging terms and taking the $\Delta\eta \to 0$ limit gives the differential equation

$$\frac{d\text{Prob}(\eta)}{d\eta} = \Gamma(\eta)\text{Prob}(\eta) , \tag{12.53}$$

[79] Note that $v_{\rm b,Dodelson} = iv_{\rm b,here}$ and that $v_e = \theta/k$.

which is solved by $\text{Prob}(\eta) = e^{-\tau}$ as claimed (with boundary condition $\text{Prob}(\eta) = 1$). Note that the integral bounds in the above definition are set up such that the conformal time derivative of the optical depth is given by

$$\tau' = -\Gamma = -a\sigma_{\mathrm{T}} n_e \,, \tag{12.54}$$

where the number density of *free* electrons n_e can be related to the number density of baryons n_b using the *free electron fraction* $n_e = x_e n_{\mathrm{b}}$ (or ionization fraction). Note that the baryon number density n_{b} counts both free protons as well as protons that combined with an electron to form neutral hydrogen, $n_{\mathrm{b}} = n_p + n_H$.

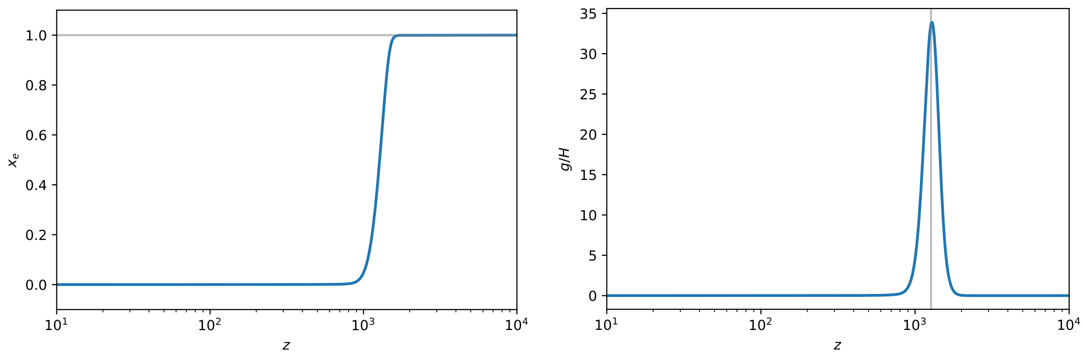


Figure 30: *Left panel:* Ionization fraction x_e calculated using RECFAST. *Right panel:* visibility function g.

The *visibility function* encodes the probability that a photon last scattered at η. This can be obtained as the time derivative of the probability $e^{-\tau}$ that a photon does not scatter, since the change in time of $e^{-\tau}$ at time η must be caused by the fraction of electrons that do scatter at that time. In formulae

$$g = \partial_\eta [e^{-\tau(\eta)}] = -\tau' e^{-\tau} = \Gamma e^{-\tau} \,. \tag{12.55}$$

By definition, the visibility function integrates to

$$\int_\eta^{\eta_0} d\eta \; g = -e^{-\tau}\Big|_{\eta_0}^{\eta} = 1 - e^{-\tau(\eta)} \,, \tag{12.56}$$

which becomes unity for very early times, $\eta \to 0$, since the probability that a photon does not scatter during the entire hot big bang goes to zero very fast. In other words, we can use that at early times $\tau \to \infty$ and at late times $\tau \to 0$. The ionization fraction $x_{\mathrm{e}} = n_{\mathrm{e}}/(n_{\mathrm{e}} + n_H)$ encodes the ratio of free electrons and the total number of electrons (free and bound in hydrogen) and the visibility function g are shown in Fig. 30. The visibility function is peaked at the time of recombination with at width of $\Delta z \approx 10$ or $\Delta\eta \approx 10\ h^{-1}$Mpc.
Having defined $\tau(\eta)$ and $g(\eta)$ we can now move on to the derivation of the line-of-sight solution. Our starting point is the first-order Boltzmann equation we derived in (12.49). If we imagine evaluating $\Theta(\eta, \vec{x}, \hat{p})$ along the trajectory $\vec{x}(\eta)$ of a photon moving in the $\hat{p}$ direction (recall $\hat{p}$ is time-independent to zeroth order in perturbations), we can re-write the first two terms in the Boltzmann equation in (12.49) in terms of the convective

derivative in (12.15). Using the solution of the geodesic equation in (12.16) we arrive at the linear Boltzmann equation in the form

$$\frac{d\Theta}{d\eta} = \Psi' + \Phi' - \frac{d\Psi}{d\eta} - \Gamma \left[\Theta - \Theta_0 - \hat{p} \cdot \vec{v}_e\right] . \tag{12.57}$$

While this expression is equivalent to the one we started from, (12.49), all the spatial gradients have been absorbed into convective derivatives and we are now in front of an ordinary, rather than partial, differential equation! Before trying to solve it, we perform two trivial operations. First, we take the perspective of someone observing the photon after it has travelled a certain distance. A photon that is propagating in direction $\hat{p}$ is detected by an observer as coming from the direction $\hat{n} = -\hat{p}$ (see Figure 31). Second, we multiply Eq. (12.48) by $e^{-\tau}$ and rewrite

$$\begin{aligned} \frac{d\left(e^{-\tau}\Theta\right)}{d\eta} = e^{-\tau}\left[\frac{d\Theta}{d\eta} + \Gamma\Theta\right] &= e^{-\tau}(\Psi' + \Phi') - e^{-\tau}\frac{d\Psi}{d\eta} + \Gamma e^{-\tau}\left[\Theta_0 - \hat{n}\cdot\vec{v}_e\right] \\ &= e^{-\tau}(\Psi' + \Phi') - \frac{d\left(e^{-\tau}\Psi\right)}{d\eta} + g\left[\Theta_0 + \Psi - \hat{n}\cdot\vec{v}_e\right] . \end{aligned} \tag{12.58}$$

Multiplying by $e^{-\tau}$ is useful because the last term above is now proportional to the visibility function g, which, as saw previously, is highly peaked around recombination (see Figure 30). We can bring the metric perturbation Ψ to the left-hand side and write

$$\frac{d}{d\eta} e^{-\tau}(\Theta + \Psi) = \hat{S}(\eta, \vec{x}(\eta), \hat{n}) , \tag{12.59}$$

with the source $\hat{S}$ defined by

$$\hat{S}(\eta, \vec{x}, \hat{n}) = e^{-\tau}\left(\Phi' + \Psi'\right) + g\left[\Theta_0 + \Psi - \hat{n}\cdot\vec{v}_e\right] . \tag{12.60}$$

In the form (12.59) the Boltzmann equation has the very simple line-of-sight solution

$$e^{-\tau}(\Theta + \Psi)\Big|_0^{\eta_0} = \int_0^{\eta_0} d\eta' \; \hat{S}(\eta', \vec{x}_0 + (\eta_0 - \eta')\hat{n}, \hat{n}) , \tag{12.61}$$

where we used that, for a photon moving in the direction $\hat{p} = -\hat{n}$ and arriving at $\vec{x}_0$ at η_0, the trajectory is $\vec{x}(\eta) = \vec{x}_0 + (\eta_0 - \eta')\hat{n}$. We can now use the fact that at present time $\tau(\eta_0) = 0$, and that at early times we have $\tau(0) = \infty$. This implies that the contribution to the left-hand-side of (12.61) from the upper boundary vanishes. Furthermore, the present day potential $\Psi(\eta_0, \vec{x}_0)$ leads to an unobservable monopole and can thus be dropped. Hence, the line-of-sight solution reduces to

$$\boxed{\Theta(\eta_0, \vec{x}_0, \hat{n}) = \int_0^{\eta_0} d\eta' \; \hat{S}(\eta', \vec{x}_0 + (\eta_0 - \eta')\hat{n}, \hat{n}) .} \tag{12.62}$$

This line-of-sight solution can be further simplified by assuming that recombination happened instantaneously at $\eta_\star \simeq 144$ Mpc. In formulae, we will approximate the visibility function $g(\eta)$ and the probability of not scattering $e^{-\tau(\eta)}$ as follows

$$g(\eta) = \delta^{(\mathrm{D})}(\eta - \eta_\star) , \qquad e^{-\tau(\eta)} = \begin{cases} 1 \text{ for } \eta > \eta_\star , \\ 0 \text{ for } \eta < \eta_\star . \end{cases} \tag{12.63}$$

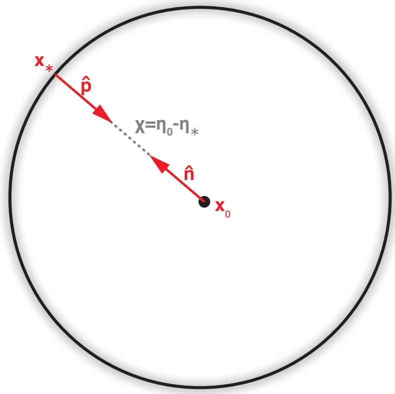


Figure 31: Photons released at the last scattering surface travel to the observer fairly uninhibitedly.

Substituting these approximation into the line-of-sight solution, we are immediately able to perform the line-of-sight integral (12.62) over the part of source $\hat{S}$ that is proportional to g. We find

$$\Theta(\eta_0, \vec{x}_0, \hat{n}) = \underbrace{(\Theta_0 + \Psi)_\star}_{\text{SW}} - \underbrace{(\hat{n} \cdot \vec{v}_\text{e})_\star}_{\text{Doppler}} + \underbrace{\int_{\eta_\star}^{\eta_0} d\eta' \; (\Psi' + \Phi')}_{\text{ISW}} \,, \tag{12.64}$$

where the first two terms should be evaluated at the time $\eta_\star$ corresponding to instantaneous reheating and at the position $\vec{x}_0 + (\eta_0 - \eta_\star)\hat{n}$ corresponding to the last scattering surface of a CMB photon observed from $\vec{x}_0$ at η_0. In (12.64) we have identified three distinct physical effects:

Sachs-Wolfe: The first term in (12.64) tells us that the temperature fluctuations we observe today in a given direction is given by the monopole temperature Θ_0 on the last scattering surface in that direction *plus* a change in temperature equal to the difference of the gravitational potential Ψ between that point and here. This Ψ contributions describes the fact that a CMB photon might have climbed in or out of a gravitational potential on its path to us, which has red- or blue-shifted its temperature in an anisotropic way because of the inhomogeneity of the Newtonian potential $\Psi(\eta, \vec{x})$. As we will see later, the Sachs-Wolfe term is the largest contribution on large angular scales. Very naively, one might have expected that relatively "hotter" directions in the CMB correspond to relatively hotter regions on the last scattering surface and the terms Θ_0 in the Sachs-Wolfe effect might seem to substantiate this expectation (at least on large scales, where other contributions are subdominant). However, due to gravitational shift, things turn out to work in exactly the opposite way. To see this, note that on large scales during matter domination[80] we have $\Theta_0 = \delta_\gamma/4 = -2\Psi_\text{md}/3$. Then the Sachs-Wolfe term becomes

$$\Theta \supset (\Theta_0 + \Psi_\text{md})_\star = (\delta_\gamma/4 + \Psi_\text{md})_\star = (-\delta_\gamma/8)_\star \,. \tag{12.65}$$

[80]Recombination takes place around $z \sim 1100$. This is during matter domination, quite a bit after matter-radiation equality around $z \sim 3300$.

Thus, on large scales, overdensities in the photon distribution ($(\delta_\gamma)_\star > 0$) appear as cold spots in the CMB ($\Theta < 0$), because the gravitational redshift acquired from climbing out of the potential well dominates over the intrinsic temperature monopole of the photons.

Doppler: Because the electron-baryon-photon plasma moves around due to inhomogeneities in pressure, the last scattering of most photons is off of moving electrons. As we discussed, this effect is simply captured by performing a Lorentz boost from the rest frame of the electron to the rest frame of the observer. This boost leaves a signal in the distribution of scattered photons, which we call the Doppler contribution. Notice that it is only the component of the electron's velocity $\vec{v}_e$ along the line of sight, i.e. towards or away from us, that contributes. As we will see shortly, the Doppler effect goes to zero on large angular scales, where the velocities are small, but is of the same order as the Sachs-Wolfe contribution on all other scales.

Integrated Sachs-Wolfe: During Matter domination, the metric perturbations are constant, and so the last term in (12.65) vanishes. However, Φ and Ψ do evolve in time during the transition from radiation to matter domination and then again during the transition from matter to dark energy domination. While matter-radiation equality takes place around $z \sim 3300$, there is still a sizeable amount of radiation around by the time or recombination. This is easy to estimate because $\rho_r \propto a^{-4}$ while $\rho_m \propto a^{-3}$ and so we expect that at recombination, namely around $z \sim 1100$, $\rho_r/\rho_m \sim 1/3$. The ongoing red-shifting of radiation causes a time dependence of the Newtonian potential and hence a non-zero "early" Integrated-Sachs-Wolfe (ISW) contribution, which is most prominent on angular scales corresponding to wavelengths entering the sound horizon around matter-radiation equality. On the opposite end, at late times, around $z \sim 0.4$, dark energy starts to dominate over matter and the Newtonian potential begins to evolve again, even in linear theory on large scales. This leads to the so-called "late" ISW contribution, which is most prominent on the largest angular scales.

13 The Angular Power Spectrum

In this section, we compute the angular power spectrum C_l of CMB temperature anisotropies. To guide the reader, we summarize here the main steps of the derivation. First, consider the 2d map $\Theta(\vec{x}_\odot, \eta_0, \hat{)}$ of CMB anisotropies as seen here at $\vec{x}_\odot$ and now at η_0 in the direction $\hat{n} = -\hat{p}$. Without loss of generality we will choose our coordinates such that the earth sits at the origin, $\vec{x}_\odot = 0$. At fixed $\vec{x}_\odot$ and η_0, $\Theta(\hat{n})$ is a function on the two-sphere and hence can be decomposed into spherical harmonics,

$$\Theta(\vec{x}_\odot, \hat{n}) = \sum_{l,m} \Theta_{lm} Y_{lm}(\hat{n}) \ , \tag{13.1}$$

where Θ_{lm} are the spherical harmonic coefficients of the CMB temperature (sometimes also denoted a_{lm} or a^T_{lm}). These coefficients depend on where in the universe they are observed, $\Theta_{lm} = \Theta_{lm}(\vec{x}_\odot)$, but here we will always consider them as seen from earth. As far as we have been able to probe, the laws of nature are covariant under rotations and the initial conditions from inflation are invariant. As a consequence, all correlators should be *statistically isotropic*, which means that the n-point correlator of fields $f_a(\hat{n})$

observed in the directions $\hat{n}_a$ for $a = 1, \ldots, n$ must obey

$$\langle \prod_{a=1}^{n} f_a(\hat{n}_a) \rangle = \langle \prod_{a=1}^{n} f_a(R\hat{n}_a) \rangle \,, \tag{13.2}$$

where R is an arbitrary rotation. In spherical harmonic space, this invariance highly constrains the correlators of products of Θ_{lm}, just like statistical homogeneity and isotropy constrains the correlators of three-dimensional fields (see Section 7.3). For the two-point function one finds

$$\langle \Theta_{lm} \Theta_{l'm'} \rangle = C_l \delta_{ll'} \delta_{mm'} \,. \tag{13.3}$$

where C_l (sometimes also denoted C_l^{TT}) is the *angular power spectrum* of CMB temperature anisotropies. The angular power spectrum is the main rotation invariant observable related to the CMB and the main object of study of this section.

In the second step we use the line-of-sight solution to write C_l as an integral of the primordial power spectrum $P_{\mathcal{R}}$ from inflation times some appropriate *temperature transfer functions*, which play an analogous role to the transfer functions we encountered in the linear matter power spectrum. More in detail, in the next subsection we will derive the relation

$$C_l = 4\pi \int d\ln k \; \Delta^2_{\mathcal{R}}(k) T_l^2(\eta_*, k) \,, \tag{13.4}$$

where the transfer functions T_l are defined in (13.12) and the dimensionless power spectrum $\Delta_{\mathcal{R}}$ was introduced in (6.55). What remains to do in the third step is to compute these transfer function by solving the time evolution during the hot big bang to linear order. We will do this in two partially overlapping regimes, the tight-coupling approximation (before recombination, Subsection 13.2) and on sub-horizon scales (after recombination, Section **??**). The final solutions are given in (13.51).

13.1 Temperature transfer functions

This subsection contains a series of computational steps to write the angular power spectrum C_l in terms of an integral over the primordial power spectrum $P_{\mathcal{R}}$ times appropriate l-dependent *temperature transfer functions* T_l. The final result is (13.18).

Since our goal is to relate Θ to the primordial perturbations $\mathcal{R}$, which we studied in Fourier space, let us express the source term in Fourier space. Without loss of generality we can choose the observation point $\vec{x}_0$ to be the origin $\vec{x}_0 = \vec{0}$. Then we find

$$\hat{S}(\eta', (\eta_0 - \eta')\hat{n}, \hat{n}) = \int_{\mathbf{k}} e^{ik(\eta_0 - \eta')\hat{k}\cdot\hat{n}} \hat{S}(\eta', \vec{k}, \hat{n}) \,. \tag{13.5}$$

The source term depends on $\vec{k}$ through the initial conditions $\mathcal{R}(\vec{k})$ and on the photon propagation direction only $\hat{n}$ through $\mu = \hat{k} \cdot \hat{n}$. Because of this we will henceforth write $\hat{S} = \hat{S}(\eta', k, \mu)$. It will be useful to factorize the angular dependence μ, which will determine the multipole expansion of the observed fluctuations, and the dependence on the initial conditions through $\mathcal{R}$. To do this, first we neglect the integrated Sachs-Wolfe contribution and then multiply and divide the rest by $\mathcal{R}(\mathbf{k})$

$$\hat{S}(\eta', k, \mu) = e^{-\tau}\left(\Phi' + \Psi'\right) + g\left[\Theta_0 + \Psi - i\mu v_e\right] \tag{13.6}$$
$$\simeq g(\eta')\left[\Theta_0 + \Psi - i\mu v_e\right] \tag{13.7}$$
$$= g(\eta')\Big[\tilde{T}_{\rm SW}(\eta', k) - i\mu \tilde{T}_{\rm D}(\eta', k)\Big]\mathcal{R}(\vec{k})\,, \tag{13.8}$$

where we defined the Sachs-Wolfe and Doppler transfer functions by

$$\tilde{T}_{\rm SW}(\eta, k) = \frac{\Theta_0(\eta, k) + \Psi(\eta, k)}{\mathcal{R}(\vec{k})}\,, \qquad \tilde{T}_{\rm D}(\eta, k) = \frac{v_e(\eta, \vec{k})}{\mathcal{R}(\vec{k})}\,. \tag{13.9}$$

These two terms have a distinct μ dependence. This distinction is conveniently accounted for by rewriting

$$\hat{S}(\eta', k, \mu)e^{ik(\eta_0-\eta')\mu} = \mathcal{R}(\vec{k})\left[\tilde{T}_{\rm SW}(\eta', k) + \frac{\tilde{T}_{\rm D}(\eta', k)}{(\eta_0 - \eta')}\frac{d}{dk}\right]e^{ik(\eta_0-\eta')\mu}\,. \tag{13.10}$$

We can now use the Rayleigh expansion of the plane wave to expand the exponential,

$$e^{i\vec{k}\cdot\vec{r}} = 4\pi\sum_{lm} i^l j_l(kr)Y^*_{lm}(\hat{k})Y_{lm}(\hat{n}) = \sum_l (2l+1)i^l j_l(kr)\mathcal{P}_l(\hat{r}\cdot\hat{k})\,. \tag{13.11}$$

The derivative in the above equation only acts on the spherical Bessel function j_l giving

$$\begin{aligned}\hat{S}(\eta', (\eta_0 - \eta')\hat{n}, \hat{n}) &= \int_{\mathbf{k}} \hat{S}(\eta', k, \mu)e^{ik(\eta_0-\eta')\mu} \\ &= \int_{\mathbf{k}}\sum_l i^l(2l+1)\mathcal{P}_l(\mu)g(\eta')\mathcal{R}(\vec{k})T_l(\eta', k)\end{aligned}$$

where we have defined the combination of the spherical Bessel functions and the transfer functions as a new l-dependent transfer function

$$T_l(\eta', k) = \tilde{T}_{\rm SW}(\eta', k)j_l(k(\eta_0 - \eta')) + \tilde{T}_{\rm D}(\eta', k)j'_l(k(\eta_0 - \eta'))\,. \tag{13.12}$$

We can expand the Legendre polynomials in spherical harmonics using

$$P_l\left(\hat{k}\cdot\hat{n}\right) = \frac{4\pi}{2l+1}\sum_{m=-l}^{l} Y_{lm}(\hat{k})Y^*_{lm}(\hat{n})\,, \tag{13.13}$$

which gives

$$\begin{aligned}\Theta(\eta_0, \vec{x}_0 = 0, \hat{n}) &= \int_0^{\eta_0} d\eta' \int_{\mathbf{k}}\sum_l (-i)^l(2l+1)T_l(\eta', k)g(\eta')\mathcal{R}(\vec{k})P_l(\hat{k}\cdot\hat{n}) \\ &= \int_0^{\eta_0} d\eta'\, 4\pi\sum_{lm}\int_{\mathbf{k}}(-i)^l T_l(\eta', k)g(\eta')\mathcal{R}(\vec{k})Y^*_{lm}(\hat{k})Y_{lm}(\hat{n})\,.\end{aligned} \tag{13.14}$$

The spherical multipoles of the temperature map we observe today from earth are thus given by

$$\Theta_{lm} = \int d^2\hat{n}\ \Theta(\hat{n})Y^*_{lm}(\hat{n}) = 4\pi\int_0^{\eta_0} d\eta' \int_{\vec{k}} g(\eta')\mathcal{R}(\vec{k})i^l T_l(\eta', k)Y^*_{lm}(\hat{k})\,, \tag{13.15}$$

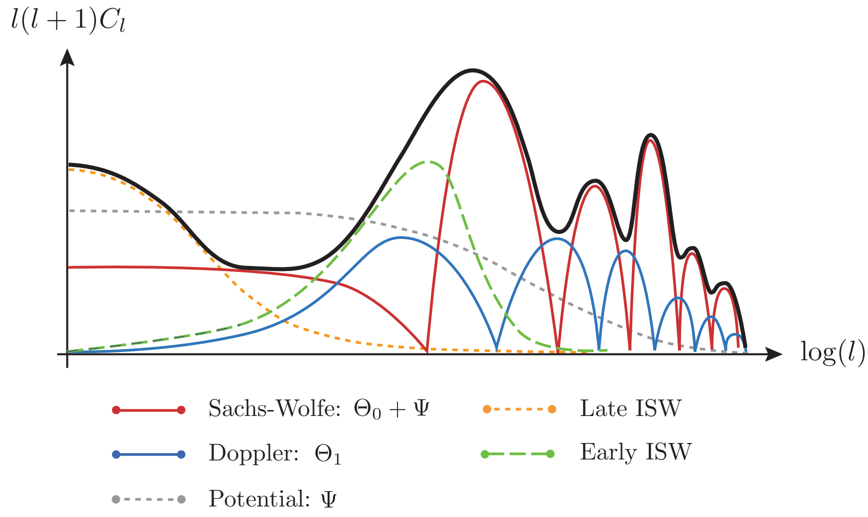


Figure 32: Overview of the contributions to the angular power spectrum [?].

where we used the orthonormality of the spherical harmonics, namely

$$\int d\Omega Y_{lm}(\hat{n}) Y^*_{l'm'}(\hat{n}) = \delta_{ll'}\delta_{mm'} \,. \tag{13.16}$$

Following from the definition of the angular power spectrum in (13.3), this leads to

$$C_l = 4\pi \int_0^{\eta_0} d\eta' \int_0^{\eta_0} d\eta'' \int d\ln k \; \Delta^2_{\mathcal{R}}(k) g(\eta') g(\eta'') T_l(\eta', k) T_l(\eta'', k) \,. \tag{13.17}$$

In the approximation of instantaneous recombination at η_*, we can use $g(\eta) = \delta(\eta - \eta_\star)$ and find

$$\boxed{C_l = 4\pi \int d\ln k \; \Delta^2_{\mathcal{R}}(k) T_l^2(\eta_\star, k) \,.} \tag{13.18}$$

13.2 Hydrodynamics: the tight-coupling expansion

It thus remains to calculate the transfer functions relating the Sachs-Wolfe and Doppler terms in terms of the primordial curvature perturbation $\mathcal{R}(\vec{k})$

$$\tilde{T}_{\mathrm{SW}}(\eta, k) = \frac{\Theta_0(\eta, \vec{k}) + \Psi(\eta, \vec{k})}{\mathcal{R}(\vec{k})} \,, \qquad \tilde{T}_{\mathrm{D}}(\eta, k) = \frac{v_e(\eta, \vec{k})}{\mathcal{R}(\vec{k})} \,. \tag{13.19}$$

To compute these we need to solve the Boltzmann equation for photons coupled to electrons and protons. To do this the line-of-sight formal solution is not very useful because before recombination the photon scatters many times and their trajectory is enormously complicated. Instead here we will take advantage of the fact that, before recombination photons electrons and protons combine into a plasma, i.e. a charged fluid, and can be described on distances longer than the mean free path using the hydrodynamics expansion.

Multiplying the evolution equation Eq. (12.50) by $\mathcal{P}_{l\geq 2}$ and integrating over the photon direction, which simply requires an integral over $d\mu$ most of the terms vanish. We finally obtain[81]

$$\Theta_l' + k\left(\frac{l+1}{2l+1}\Theta_{l+1} - \frac{l}{2l+1}\Theta_{l-1}\right) = -\Gamma\Theta_l\,. \tag{13.21}$$

We will now consider the limit of large optical depth $\tau \gg 1$ or scattering rate larger than the expansion rate $\Gamma \gg k \gtrsim \mathcal{H}$ and large scales $k\eta \approx 1$. Moments of order exceeding two $\Theta_l, l > 2$ are suppressed for $\Gamma \gg k$. To see this let us first drop the Θ_{l+1}-term (we'll justify this later) and replace the derivative in the first term by a η^{-1}

$$\Theta_l - \eta k\frac{l}{2l+1}\Theta_{l-1} + \eta\Gamma\Theta_l = 0 \tag{13.22}$$

With our approximations, we can drop Θ_l in comparison to $\eta\Gamma\Theta_l$ in the above equation. We thus have for the relation of Θ_l and Θ_{l-1}

$$\Theta_l \approx \frac{k}{\Gamma}\Theta_{l-1} \ll \Theta_{l-1} \tag{13.23}$$

where we used that $k \ll \Gamma$. The fact that $\Theta_l \ll \Theta_{l-1}$ obviously justifies dropping Θ_{l+1} above. Multiplying the evolution equations with $\mathcal{P}_0$ and $\mathcal{P}_1$ and integrating we obtain the continuity and Euler equations for photons in the tight-coupling-approximation

$$\Theta_0' = -k\Theta_1 + \Phi'\,, \tag{13.24}$$
$$3\Theta_1' = k\Theta_0 + k\Psi - \Gamma(3\Theta_1 + v_\mathrm{b})\,. \tag{13.25}$$

Moments of the stress-energy tensor

$$T^\mu_\nu = \int d^3p\; fP^\mu P_\nu \tag{13.26}$$

can be associated with the fluid overdensity and momentum, yielding $\delta_\gamma = 4\Theta_0$, $v_\gamma = -3\Theta_1$ and $\sigma_\gamma = -3\Theta_2$. For dark matter we have

$$\delta_\mathrm{dm}' = -kv_\mathrm{dm} + 3\Phi'\,, \tag{13.27}$$
$$v_\mathrm{dm}' = -\mathcal{H}v_\mathrm{dm} - k\Psi\,. \tag{13.28}$$

Electrons scatter off protons very often because the rate of Coulomb scattering is much larger than any other time scale in the problem. So we can effectively treat protons and electrons as a single entity, which we will awkwardly denote by "baryons" and indicate with a subscript "b". For the baryons we have

$$\delta_\mathrm{b}' = -kv_\mathrm{b} + 3\Phi'\,, \tag{13.29}$$
$$v_\mathrm{b}' = -\mathcal{H}v_\mathrm{b} - k\Psi - \frac{\Gamma}{R}(3\Theta_1 + v_\mathrm{b})\,. \tag{13.30}$$

[81] This can be seen using the recursion relation

$$(2l+1)\mu P_l(\mu) = (l+1)P_{l+1}(\mu) + lP_{l-1}(\mu) \tag{13.20}$$

Here we have introduced the *baryon-to-photon ratio*

$$R = \frac{\rho_\mathrm{b}}{\rho_\gamma + p_\gamma} = \frac{3}{4}\frac{\rho_\mathrm{b}}{\rho_\gamma} \propto a\,. \tag{13.31}$$

The baryon to photon ratio is zero early on an approaches unity at recombination. The presence of baryons makes the fluid heavier, thus reducing the speed of sound as

$$c_\mathrm{s}^2 = \frac{1}{3}\frac{1}{1+R}\,. \tag{13.32}$$

13.3 Acoustic oscillations

Let us slightly rewrite the baryon Euler equation (13.30)

$$v_\mathrm{b} = -3\Theta_1 - \frac{R}{\Gamma}\left[v_\mathrm{b}' + \mathcal{H}v_\mathrm{b} + k\Psi\right]\,. \tag{13.33}$$

At leading order $\Gamma \to \infty$ we find

$$v_\mathrm{b} = -3\Theta_1 \qquad\qquad \text{(leading order)}\,. \tag{13.34}$$

The correction term is suppressed by Γ^{-1}, but in the photon Euler equation (13.25) the baryon velocity is multiplied with Γ. We can solve the equation perturbatively by replacing $v_\mathrm{b} = -3\Theta_1$ in the correction term to obtain

$$v_\mathrm{b} = -3\Theta_1 + \frac{R}{\Gamma}\left[3\Theta_1' + 3\mathcal{H}\Theta_1 - k\Psi\right]\,. \tag{13.35}$$

Using this result in the photon Euler equation we obtain

$$\Theta_1' = \frac{k}{3(1+R)}\Theta_0 - \frac{\mathcal{H}R}{1+R}\Theta_1 + \frac{k}{3}\Psi\,. \tag{13.36}$$

Taking another time derivative of the photon continuity equation in Eq. (13.25) and using the above solution we finally have

$$\Theta_0'' + \frac{\mathcal{H}R}{1+R}\Theta_0' + c_s^2k^2\Theta_0 = -\frac{1}{3}k^2\Psi + \Phi'' + \frac{\mathcal{H}R}{1+R}\Phi'\,, \tag{13.37}$$

where the speed of sound c_s of the photon-baryon-electron plasma is

$$c_s^2 = \frac{1}{3(1+R)}\,. \tag{13.38}$$

This is only slightly decreased from the pressure of a general relativistic fluid, for which one has $p = -\rho/3$ and so $c_s^2 = \delta p/\delta\rho = 1/3$, implying a traceless energy-momentum tensor as appropriate for a conformal system. The difference from $1/3$ is due to the fact that the plasma contains the massive protons and electrons, which at these energies are both non-relativistic. On the left-hand side of (13.37) we have a friction term and a pressure term and on the right hand side a gravitational term and dilation terms. We will neglect the friction term Θ_0' because the time-scale of the oscillations is much faster than that of the dissipation generate by this term. This could be rectified using the WKB approximation but we will not do this here. Moreover, we neglect the time derivative of metric perturbations, which vanish during matter domination but not during radiation domination. In summary we are left with the much simpler equation

$$\Theta_0'' + c_s^2 k^2 \Theta_0 = -\frac{1}{3} k^2 \Psi \,. \tag{13.39}$$

This is a forced harmonic oscillator. Let's start with the solutions of the homogeneous equation:

$$\Theta_0(\eta) \supset A \sin(k r_s) + B \cos(k r_s)\,, \tag{13.40}$$

where we defined the *sound horizon* as

$$r_{\mathrm{s}}(\eta) = \int_0^\eta d\eta' \; c_{\mathrm{s}}(\eta')\,. \tag{13.41}$$

This is the distance travelled by a sound wave, which is moving at a speed $c_s(\eta)$, from be beginning of time, $\eta = 0$, until time η. If we neglect the small correction due to $(1+R)$ in the speed of sound we would find $r_s \approx c_{\mathrm{s}}\eta \sim \eta/\sqrt{3}$.
To find the inhomogeneous solution of (13.39) we can use the Green's function method.

$$\Theta_0(\eta, \vec{k}) = \left[\Theta_0(0,\vec{k}) + \Psi(0,\vec{k})\right] \cos(k r_{\mathrm{s}}) + \frac{1}{k c_{\mathrm{s}}} \Theta_0'(0,\vec{k}) \sin(k r_{\mathrm{s}}) - \Psi(\eta, \vec{k}) \tag{13.42}$$

For adiabatic initial conditions $\Theta_0'(0) = 0$, which means that the sine mode has to vanish. The tight coupling approximation entails that at leading order

$$\Theta_{0,\mathrm{SW}} = \Theta_0(\eta,\vec{k}) + \Psi = \left[\Theta_0(0,\vec{k}) + \Psi(0,\vec{k})\right] \cos(k r_{s,\star}) - R\Psi \tag{13.43}$$

$$v_b = -3\Theta_1 = \frac{3}{k}\left[\Phi' - \Theta_0'\right] \tag{13.44}$$

$$= 3 c_{\mathrm{s}} \left[\Theta_0(\eta,\vec{k}) + (1+R)\Psi\right] \sin(k r_{s,\star})\,, \tag{13.45}$$

where for $-\Theta_1$ we used (13.24). The comoving curvature perturbation $\mathcal{R}$ is related to the Newtonian gauge metric perturbations via

$$\mathcal{R} = -\Phi - \frac{\mathcal{H}(\Phi' + \mathcal{H}\Phi)}{4\pi G a^2(\bar\rho + \bar P)} = -\Phi - \frac{2}{3}\frac{\mathcal{H}^{-1}\Phi' + \Psi}{1+w}\,, \tag{13.46}$$

such that

$$\Phi = -\frac{3+3w}{5+3w}\mathcal{R}. \tag{13.47}$$

During the matter-radiation transition the curvature perturbation $\mathcal{R}$ remains constant outside the horizon. The change in the equation of state thus implies a change in the amplitude of the Newtonian potential Φ

$$\mathcal{R} = -\frac{3}{2}\Phi_{\mathrm{rd}} = -\frac{5}{3}\Phi_{\mathrm{md}} \quad \Rightarrow \quad \Phi_{\mathrm{md}} = \frac{9}{10}\Phi_{\mathrm{rd}}\,. \tag{13.48}$$

On very large scales we can ignore the terms multiplied by k in Eqs. (13.24) and (13.29)

$$\Theta_0' = \Phi'\,, \qquad\qquad \delta_{\mathrm{m}}' = 3\Phi'\,. \tag{13.49}$$

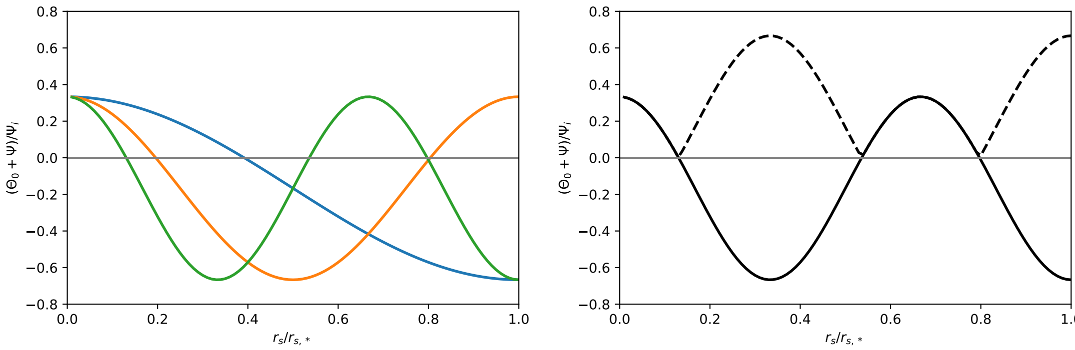


Figure 33: *Left panel:* Observed temperature perturbation $\Theta_0 + \Psi$ in units of the metric perturbation Ψ for several wavenumbers k. Following Eq. (12.65), the modes start at $\Theta_0 + \Psi = \Psi/3$. *Right panel:* As you can see the baryon loading shifts the cuves downwards, leading to the magnitude of the odd peaks being larger than that of the even peaks.

Thus $\delta_\mathrm{m} = 3\Theta_0 + \text{const.}$. For adiabatic initial conditions this constant needs to vanish. With $\Theta_0 = \delta_\gamma/4$ we thus have $\delta_\mathrm{m} = 3/4\delta_\gamma$.
On large scales, for constant potentials $\Phi' = 0$ we have from Eq. (12.3) that

$$-3\mathcal{H}^2\Phi = 4\pi G a^2 \left(\bar{\rho}_\mathrm{m}\delta_\mathrm{m} + \bar{\rho}_\gamma\delta_\gamma\right) . \tag{13.50}$$

Matter Domination:
Eq. (13.50) entails that $\delta_\mathrm{m} = -2\Phi_\mathrm{md}$ and thus $\delta_\gamma = -8/3\Phi_\mathrm{md}$ or $\Theta_0 = -2/3\Phi_\mathrm{md}$. This leads to $\Theta_0 + \Psi = \Psi/3$.

Radiation Domination:
Eq. (13.50) that $\Theta_0 = -\Phi_\mathrm{rd}/2$ and thus $\delta_\gamma = -2\Phi_\mathrm{rd}$ or $\delta_\mathrm{m} = -3/2\Phi_\mathrm{rd}$. Using $\Theta_0 = 2/5\mathcal{R} = -2/3\Psi_\mathrm{md}$ we have for the transfer functions expressing Θ in terms of $\mathcal{R}$ in Eq. (13.19)

$$\tilde{T}_\mathrm{SW} = \frac{\Theta_0 + \Psi}{\mathcal{R}} = -\frac{1+3R}{5}\cos(kr_\mathrm{s}) + \frac{3R}{5} , \tag{13.51}$$

$$\tilde{T}_\mathrm{D} = \frac{v_\mathrm{b}}{\mathcal{R}} = 3c_\mathrm{s}\frac{1+3R}{5}\sin(kr_\mathrm{s}) . \tag{13.52}$$

We show these dependencies in Fig. 33 for $R = 1$.
For modes that enter during radiation domination the gravitational potential is not constant but decaying. Counterintuitively, this decaying potential can enhance the first compression. This happens by pulling the photons into overdensity then decaying and thus not being able to halt the subsequent rarefaction [55].
In Fig. 34 we show the relevant scales for CMB physics.

Hubble radius Fluctuations outside the comoving Hubble radius $r_\mathrm{H} \approx \mathcal{H}^{-1}$ are frozen but they start to evolve once they cross into the horizon.

Sound horizon Photon perturbations remain frozen until they cross the sound horizon $r_\mathrm{s} = c_\mathrm{s}\eta$, which is when they start to oscillate. At early times $R \to 0$ and thus $c_\mathrm{s} \approx 1/\sqrt{3}$, whereas R becomes significant just before recombination, driving the speed of sound to zero. After recombination, the concept of a sound horizon is ill defined.

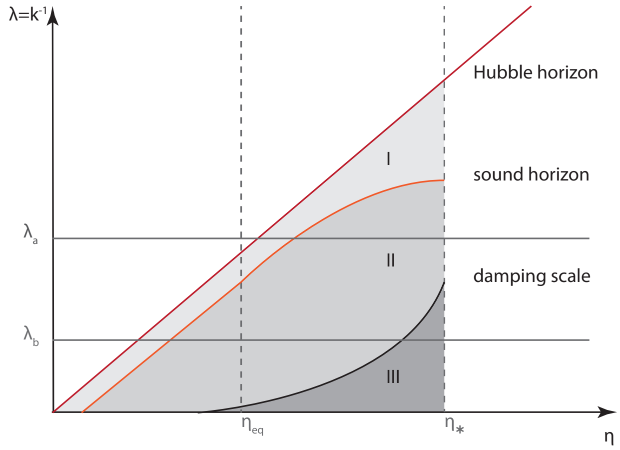


Figure 34: Scales affecting the evolution of perturbations in the photon-baryon fluid. As time proceeds fluctuations are moving through the regions horizontally. Adapted from [?].

Damping scale On scales of order the photon mean free path fluctuations are damped by viscosity and heat conduction arising from diffusion.

13.4 Diffusion Damping

Damping arises from a non-vanishing quadrupole. We will thus need to consider higher orders in $1/\tau' = -1/\Gamma$. The mean free path of the photons in the plasma is given by

$$\lambda_{\mathrm{MFP}} = (a\sigma_{\mathrm{T}} n_{\mathrm{e}})^{-1}\,. \tag{13.53}$$

In a Hubble time the photon scatters $N_{\mathrm{step}} = \Gamma\eta_{\mathcal{H}} = a\sigma_{\mathrm{T}}n_{\mathrm{e}}/\mathcal{H}$ times and thus the root-mean-square displacement of the random walk is

$$\sigma_{\mathrm{D}} = \lambda_{\mathrm{MFP}}\sqrt{N_{\mathrm{step}}} = \frac{1}{\sqrt{a\sigma_{\mathrm{T}}n_{\mathrm{e}}\mathcal{H}}} = \frac{1}{\sqrt{\Gamma\mathcal{H}}}. \tag{13.54}$$

Perturbations on scales below this rms displacement scale will be washed out. We can also see this effect a bit more formally considering the dispersion relation of the photon waves. In the WKB approximation we can write for high frequencies

$$\Theta_l, v_{\mathrm{b}} \propto \exp\left\{\left[i\int d\eta\;\omega\right]\right\} \tag{13.55}$$

leading to $v_{\mathrm{b}}' = i\omega v_{\mathrm{b}}$.
Diffusion damping is relevant on small scales, where the expansion is much slower than the small scale dynamics. It follows that we can drop the gravitational terms and get for the evolution equations

$$\begin{aligned}\Theta_0' &= -\,k\Theta_1\\ 3\Theta_1' &= k\left(\Theta_0 - 2\Theta_2\right) - \Gamma(3\Theta_1 + v_{\mathrm{b}})\\ 5\Theta_2' &= 2k\Theta_1 - \frac{9}{2}\Gamma\Theta_2\end{aligned} \tag{13.56}$$

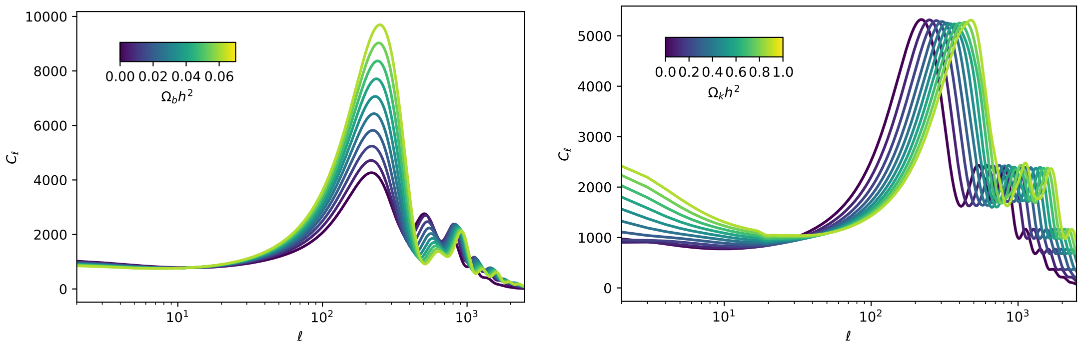


Figure 35: Impact of baryon density and curvature on the angular power spectrum. C_ls calculated using CAMB.

Note that the $l = 2$ equation contains a correction from the quadrupole term in the Thomson cross section, which changes the coefficient of Θ_2 on the right hand side from 5 to 9/2. Since $\Theta_2' \ll \Gamma\Theta_2$ we have

$$\Theta_2 = \frac{4}{9}\frac{k}{\Gamma}\Theta_1 \qquad\qquad \Theta_0 = \frac{ik}{\omega}\Theta_1\,. \tag{13.57}$$

It remains to relate the baryon velocity to the moments of the photon temperature

$$v_{\rm b}' + \left(\mathcal{H} + \frac{\Gamma}{R}\right) v_{\rm b} = -3\frac{\Gamma}{R}\Theta_1 \tag{13.58}$$

As $\mathcal{H} \ll \Gamma$ we have

$$\left(1 + i\frac{\omega R}{\Gamma}\right) v_{\rm b} = -3\Theta_1 \tag{13.59}$$

Thus

$$v_{\rm b} = -3\Theta_1 \left(1 + i\frac{\omega R}{\Gamma}\right)^{-1} \approx -3\Theta_1 \left[1 - i\frac{\omega R}{\Gamma} - \left(\frac{\omega R}{\Gamma}\right)^2\right] \tag{13.60}$$

We can now use this result as well as the solutions for Θ_0 and Θ_2 in the dynamical equation for Θ_1 (the Euler equation)

$$\omega^2(1+R) - \frac{k^2}{3} - i\frac{\omega}{\Gamma}\left(\frac{8}{27}k^2 + R^2\omega^2\right) = 0\,. \tag{13.61}$$

At leading order the dispersion relation is solved by

$$\omega = c_{\rm s} k = \frac{1}{\sqrt{3(1+R)}}k\,. \tag{13.62}$$

Using this solution in the imaginary part of the dispersion relation suppressed by Γ^{-1} and expanding $\omega \to \omega + \delta\omega$ in the leading order real part, we obtain

$$\delta\omega = i\frac{k^2}{\Gamma}\frac{1}{(1+R)}\left(\frac{8}{9} + \frac{R^2}{1+R}\right) \tag{13.63}$$

The imaginary form of the dispersion relation leads to the damping $\exp\{[i\int d\eta\ \delta\omega]\} = \exp\{[-k^2/k_{\mathrm{D}}^2]\}$, where

$$\frac{1}{k_{\mathrm{D}}^2} = \frac{1}{6}\int d\eta\ \frac{1}{\Gamma}\frac{1}{(1+R)}\left(\frac{8}{9}+\frac{R^2}{1+R}\right) \tag{13.64}$$

We see that our naïve estimate of $k_{\mathrm{D}}^2 \sim \sigma_{\mathrm{D}}^{-2} \sim \Gamma$ was correct.
When using the instantaneous recombination approximation discussed above, the finite width of the last scattering surface can be accounted for by an additional Gaussian damping [87] with standard deviation $\sigma_{\mathrm{LSS}} \approx 0.03$.

Acknowledgements

In developing this course and writing these lecture notes, I acknowledge the crucial help of Dr. Tobias Baldauf and Dr. Scott Melville, who taught part of this material in 2019-21 and 2021-22, respectively. The present discussion of large scale structures and the cosmic microwave background is based on a previous version of these notes by Tobias Baldauf.

A Notation, units and conventions

In these notes we use units in which $\hbar = c = k_b = 1$. Therefore energy is temperature and inverse time or inverse length. On the other hand, we try to keep the reduced Planck mass explicit, $M_{\rm Pl} = (8\pi G_N)^{-1/2}$. Beware that some authors use $M_{\rm Pl}$ to indicate the "full" Planck mass $G_N^{-1/2} \simeq 1.2 \times 10^{19}$GeV. The necessary conversion factors can be added using dimensional analysis and

$$\begin{aligned} c &= 3 \times 10^8 \frac{\rm m}{\rm sec}\,, \quad {\rm pc} = 3.2\,{\rm lightyears}\,, \quad {\rm year} = \pi \times 10^7\,{\rm sec}\,, & \text{(A.1)} \\ \hbar c &= 0.2\,{\rm eV}\,\mu{\rm m}\,, \quad M_{\rm Pl} \simeq 2.4 \times 10^{18}{\rm GeV}\,. & \text{(A.2)} \end{aligned}$$

We use the mostly plus signature $(-,+,+,+)$. Latin indices indicate space, $i, j, \dots = \{1, 2, 3\}$, while greek indices run over spacetime, $\mu, \nu, \dots = \{0, 1, 2, 3\}$. Three-dimensional vectors are in boldface, e.g. $\mathbf{k}$ and $\mathbf{x}$. Unless otherwise specified, all tensors are expressed in terms of the FLRW coordinates

$$ds^2 = -dt^2 + a^2 dx^2\,. \tag{A.3}$$

. Standard derivatives are represented with a comma, and covariant derivatives with a semi-column

$$T^{\dots}_{\dots,\mu} \equiv \partial_\mu T^{\dots}_{\dots}\,, \quad T^{\dots}_{\dots;\mu} \equiv \nabla_\mu T^{\dots}_{\dots}\,. \tag{A.4}$$

Symmetrization and anti-symmetrization of a pair of indices is indicated with $(\dots)$ and $[\dots]$ respectively and is defined to have weight one

$$A_{(\mu\nu)} \equiv \frac{1}{2}\left(A_{\mu\nu} + A_{\nu\mu}\right)\,, \quad A_{[\mu\nu]} \equiv \frac{1}{2}\left(A_{\mu\nu} - A_{\nu\mu}\right)\,. \tag{A.5}$$

My convention for the Fourier transform are

$$F(\mathbf{x}) = \int_{\mathbf{k}} F(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}}, \quad F(\mathbf{k}) = \int_{\mathbf{x}} F(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}}\,, \tag{A.6}$$

with

$$\int_{\mathbf{k}} \equiv \int \frac{d^3\mathbf{k}}{(2\pi)^3}\,, \quad \int_{\mathbf{x}} \equiv \int d^3\mathbf{x}\,. \tag{A.7}$$

We sometimes use the shorthand notations: D for Dodelson's Modern Cosmology book [37], W for Weinberg's Cosmology book [104]. For example D 3 is Chapter 3 of Donelson's book, while W AppB is App. B of Weinberg's book.

There are surprisingly many conventions for the name of variables in perturbation theory. In particular, Newtonian gauge is written as

$$\begin{aligned} ds^2 &\equiv -\left(1 + 2\Psi_D\right) dt^2 + a^2 \left(1 + 2\Phi_D\right) dx^i dx^j \delta_{ij} & \text{(A.8)} \\ &\equiv -\left(1 + 2\Phi_W\right) dt^2 + a^2 \left(1 - 2\Psi_W\right) dx^i dx^j \delta_{ij} & \text{(A.9)} \end{aligned}$$

in Dodelson's (D) or Weinberg's (W) notations. The conversion is $\Psi_D = \Phi_W$ and $\Phi_D = -\Psi_W$. In these notes, we use Dodelson's notation everywhere the label W is written explicitly.

B Explicit examples: Contact correlators and Wick's theorem*

In the main text we introduce diagrammatic rules that automatise the calculation of correlators. These are easy to learn and apply, but hide a few details. In this appendix we show how to perform the calculation very explicitly from first principles, without invoking any diagrammatic rules. In practice this approach becomes prohibitively slow and cumbersome when looking beyond the simplest contact correlators.
In the main text, we introduced a very general formalism to compute cosmological correlators. Let's see it in action for the simple example of de Sitter spacetime with a scalar field. This will turn out to be a good approximation of realistic inflationary models. For the moment the metric will be fixed, non-dynamical, so we are neglecting the effect of the scalar field perturbations on the geometry. We will amend this in Sec. ??. The simplest calculation to perform are *contact interactions*, which contribute to correlators already at linear order in H_{int}. An example is depicted on the left-hand side of Figure 36.

Example: cubic interaction The simplest interaction one can think of in particle physics is a cubic potential term $V = \mu\varphi(x)^3$. For this term, we can write the Hamiltonian as

$$H_{\rm int} = -L_{\rm int} = \int d^3x\,\sqrt{-g}\mu\varphi(\mathbf{x},\tau)^3 \tag{B.1}$$

$$= a^4\mu\int_{\mathbf{q_1 q_2 q_3}} \varphi(\mathbf{q_1},\tau)\varphi(\mathbf{q_2},\tau)\varphi(\mathbf{q_3},\tau)\,(2\pi)^3\delta_D^3\left(\mathbf{q_1}+\mathbf{q_2}+\mathbf{q_3}\right)\,. \tag{B.2}$$

where

$$\varphi(\mathbf{q},\tau) = f_q(\tau)a_{\mathbf{q}} + f_q^*(\tau)a^\dagger_{-\mathbf{q}}\,, \tag{B.3}$$

$$f_q(\tau) = \frac{H}{\sqrt{2q^3}}(1+iq\tau)e^{-iq\tau}\,. \tag{B.4}$$

This interaction induces a non-vanishing three-point correlator or *bispectrum*, as it is often called. We use the commutator-form of the in-in formula (3.37) to leading non-trivial order

$$\langle\varphi(\mathbf{k}_1,\tau)\varphi(\mathbf{k}_2,\tau)\varphi(\mathbf{k}_3,\tau)\rangle = i\int_{-\infty}^{\tau} d\tau'\langle[H_{int}(\tau'),\varphi(\mathbf{k}_1,\tau)\varphi(\mathbf{k}_2,\tau)\varphi(\mathbf{k}_3,\tau)]\rangle\,. \tag{B.5}$$

We will eventually be interested in the correlators at late time, $\tau\to 0$. For the time being, we will keep τ general. For any Hermitian operator $\mathcal{O}^\dagger = \mathcal{O}$, we can re-write the commutator as

$$\begin{aligned}\langle[H_{int},\mathcal{O}]\rangle &= \langle H_{int}\mathcal{O}\rangle - \langle\mathcal{O}H_{int}\rangle = \langle H_{int}\mathcal{O}\rangle - \langle\mathcal{O}^\dagger H_{int}^\dagger\rangle \\ &= \langle H_{int}\mathcal{O}\rangle - \langle(H_{int}\mathcal{O})^\dagger\rangle = \langle H_{int}\mathcal{O}\rangle - \langle H_{int}\mathcal{O}\rangle^* = 2i\,\mathrm{Im}\langle H_{int}\mathcal{O}\rangle\,,\end{aligned}$$

where we used that also H_{int} is Hermitian. All equal-time products of fields in real space are Hermitian

$$(\varphi(\mathbf{x}_1)\ldots\varphi(\mathbf{x}_n))^\dagger = \varphi(\mathbf{x}_1)\ldots\varphi(\mathbf{x}_n)\,, \tag{B.6}$$

because $\varphi(\mathbf{x})$ is Hermitian and the equal-time fields commute. By taking the Fourier transform we find

$$(\varphi(\mathbf{k}_1)\ldots\varphi(\mathbf{k}_n))^\dagger = \left(\int_{\mathbf{x}_1\ldots\mathbf{x}_n} e^{-ik_a x_a}\varphi(\mathbf{x}_1)\ldots\varphi(\mathbf{x}_n)\right)^\dagger \tag{B.7}$$

$$= \varphi(-\mathbf{k}_1)\ldots\varphi(-\mathbf{k}_n)\,. \tag{B.8}$$

If the theory is symmetric under spatial parity[82], which we will assume in the following, then we can flip the sign of momenta again and find that also the product of field in Fourier space is an Hermitian operator.
Our correlator becomes

$$-2\,\mathrm{Im}\int_{-\infty}^{\tau} d\tau' a^4(\tau')\int_{\mathbf{q}_1\mathbf{q}_2\mathbf{q}_3}(2\pi)^3\delta_D\left(\mathbf{q}_1+\mathbf{q}_2+\mathbf{q}_3\right)\times \tag{B.9}$$
$$\langle\varphi(\mathbf{q}_1,\tau')\varphi(\mathbf{q}_2,\tau')\varphi(\mathbf{q}_3,\tau')\varphi(\mathbf{k}_1,\tau)\varphi(\mathbf{k}_2,\tau)\varphi(\mathbf{k}_3,\tau)\rangle\,,$$

where we dropped the prime in the integration variable.

Wick's theorem The above kind of expressions are most easily computed using (a variant of) *Wick's theorem.* Let us define the *contraction* of two fields as

$$\varphi^\bullet(\mathbf{q},\tau')\varphi^\bullet(\mathbf{k},\tau) = \varphi(\mathbf{q},\tau')\varphi(\mathbf{k},\tau) - :\varphi(\mathbf{q},\tau')\varphi(\mathbf{k},\tau):\,, \tag{B.10}$$

where $:\cdots:$ denotes *normal ordering* (all creation operators to the left of all annihilation operators) and the bullets $^\bullet$ mark the fields to be contracted. Since all normal ordered products vanish inside an expectation value, we find

$$\langle\varphi^\bullet(\mathbf{q},\tau')\varphi^\bullet(\mathbf{k},\tau)\rangle = \langle\varphi(\mathbf{q},\tau')\varphi(\mathbf{k},\tau)\rangle = f_q(\tau')f_k^*(\tau)(2\pi)^3\delta_D^3(\mathbf{q}+\mathbf{k})\,. \tag{B.11}$$

This is the Fourier space *propagator.* Notice that the order matters, namely

$$\langle\varphi(\mathbf{q},\tau')\varphi(\mathbf{k},\tau)\rangle = \langle\varphi(\mathbf{k},\tau)\varphi(\mathbf{q},\tau')\rangle^*\,. \tag{B.12}$$

This propagator is related to but distinct from the Feynman, advanced and retarded propagators. Wick's theorem then states that

$$\prod_{a=1}^{n}\varphi(\mathbf{k}_a,\tau_a) = \sum_{\substack{\text{pairwise}\\ \text{contr's}}} :\prod_a\varphi(\mathbf{k}_a,\tau_a):\,, \tag{B.13}$$

where the sum runs over all possible ways to *pairwise* contract any subset of the fields in the product. Since $\langle:\mathcal{O}:\rangle = 0$, inside an expectation value the only surviving term is that in which all fields have been contracted,

$$\langle\prod_{a=1}^{2n}\varphi_a\rangle = \sum_{\text{perm's}}\left[\langle\varphi_1^\bullet\varphi_2^\bullet\rangle\ldots\langle\varphi_{2n-1}^\bullet\varphi_{2n}^\bullet\rangle\right]\,, \tag{B.14}$$

$$= \sum_{\text{perm's}}\left[\langle\varphi_1\varphi_2\rangle\ldots\langle\varphi_{2n-1}\varphi_{2n}\rangle\right]\,, \tag{B.15}$$

[82] Notice that specifically for the three-point correlator, invariance under rotations implies parity. This is because all three vectors must lie on a plane, which can be rotated by 180° to invert all vectors.

where we used the shorthand notation $\varphi(\mathbf{k}_a, \tau_a) = \varphi_a$.
Using Wick's theorem in (B.9) we have in principle many possible pairs to contract. But here we are only interested in those contractions between one φ in $H_{int}(\tau')$ and one in $\varphi^3(\tau)$. The sum of all and only such contractions is called a *connected* correlator. We will discuss connected correlators in full generality later on. There are only 3! terms contributing to the connected correlator and they all give the same result, so we pick up a factor of 6. From (B.9), our correlator becomes

$$-2 \times 3! \times \mu \,\mathrm{Im}\left[\prod_{a=1}^{3} f^*_{k_a}(\tau)\right] \int_{-\infty}^{\tau} d\tau' \, \frac{1}{H^4 \tau'^4} \, f_{k_1}(\tau') f_{k_2}(\tau') f_{k_3}(\tau') \tag{B.16}$$

$$= -\frac{3}{2} \frac{\mu H^2}{(k_1 k_2 k_3)^3} \,\mathrm{Im}\left[\prod_{a=1}^{3}(1 - i k_a \tau)\right] \int_{-\infty}^{\tau} \frac{d\tau'}{\tau'^4} \, e^{-i k_T (\tau' - \tau)} \left[\prod_{a=1}^{3}(1 + i k_a \tau')\right], \tag{B.17}$$

where we introduce the "total energy" $k_T = k_1 + k_2 + k_3$. The integral is a bit complicated. First we notice that, thank to the rotation into the lower complex plain for the anti-time ordered factors, the integral converges at $\tau \to -\infty(1 + i\epsilon)$ because of the exponential suppression. The interaction is shutting off in the infinite past, just as we wanted. We can then focus on the upper limit of integration. Upon expanding the product in the integrand one finds integrals of the form

$$\int_{-\infty}^{\tau} \frac{d\tau'}{\tau'^n} \, e^{-i k_T \tau'}, \tag{B.18}$$

for $n = 1, 2, 3, 4$. The strategy is then to use integration by parts to reduce each term to the exponential integral Ei defined by

$$\mathrm{Ei}(x) = -\int_{-x}^{\infty} \frac{e^{-t}}{t} \, dt \,. \tag{B.19}$$

The result of the integral is

$$\int_{-\infty}^{\tau} \frac{d\tau'}{\tau'^4} \, e^{-i k_T \tau'} \left[\prod_{a=1}^{3}(1 + i k_a \tau')\right] = -\frac{i}{3} \sum_{a=1}^{3} \left(k_a^3\right) \mathrm{Ei}(-i k_T \tau) + \tag{B.20}$$

$$- \frac{e^{-i k_T \tau}}{3 \tau^3} \left[1 + i k_T \tau + \left(\sum_{a=1}^{3} k_a^2 - \sum_{a \neq b} k_a k_b\right) \tau^2\right].$$

Using the asymptotic

$$\mathrm{Ei}(-i k_T \tau) \simeq \gamma_E + \log(k_T \tau) - i \frac{\pi}{2}, \tag{B.21}$$

where $\gamma_E \simeq 0.577$ is the Euler-Mascheroni constant, we can take the $\tau \to 0$ limit of (B.17) and find

$$\langle \varphi(k_1) \varphi(k_2) \varphi(k_3) \rangle = (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) \frac{\mu H^2}{2 k_1^3 k_2^3 k_3^3}$$

$$\times \left[\sum_a k_a^3 \left(\gamma_E - 1 + \ln\left(-k_T \tau\right)\right) + k_1 k_2 k_3 - \sum_{a \neq b} k_a^2 k_b\right]. \tag{B.22}$$

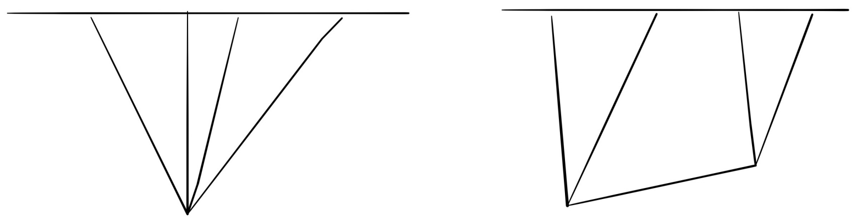


Figure 36: The two diagrams that contribute at tree level to the four-point correlator: the contact diagram on the left-hand side and the exchange diagram on the right-hand side.

Some comments on this result are in order. First, we immediately recognize the ubiquitous momentum-conserving delta functions. It is common to suppress this factor by appending a prime to the correlator or to define B_n as

$$\langle\varphi(\mathbf{k}_1)\dots\varphi(\mathbf{k}_n)\rangle = (2\pi)^3\delta_D^3\left(\sum_{a=1}^{n}\mathbf{k}_a\right)B_n(\mathbf{k}_1,\dots,\mathbf{k}_n) \tag{B.23}$$

$$\langle\varphi(\mathbf{k}_1)\dots\varphi(\mathbf{k}_n)\rangle' = B_n(\mathbf{k}_1,\dots,\mathbf{k}_n)\,. \tag{B.24}$$

Second, the coupling constant μ unsurprisingly appears linearly. Third we see that B_3 only depends on the norm of the momenta but not on their orientation. This will turn out to be a consequence of rotation and translation invariance. The overall scaling with $k \sim k_a$ is $B_n \sim k^{-6}$. We will soon see that this is a consequence of scale invariance. Fourth, the correlator is fully symmetric under any permutations of $\{k_1, k_2, k_3\}$. Finally, the limit $\tau \to 0$ turned out to be log-divergent! This is one of many divergences that show up in dS spacetime. We will see later on that the gauge invariant observables in the problems are actually finite.

Example: quartic derivative interaction Let's compute another correlator. This time we will choose to compute a four-point function B_4. At tree level, this can be generated by a quartic contact interaction as on the left-hand side of Figure 36 or from two cubic interaction as in the *exchange* diagram on the right-hand side. Let us compute the contribution from the contact interaction, which we assume to come from

$$H_{int} = \int_{\mathbf{x}} a^4\,\frac{1}{4!\Lambda^4}\left(\partial_\tau\varphi g^{\tau\tau}\partial_\tau\varphi\right)^2 \tag{B.25}$$

$$= \int_{\mathbf{q}_1\dots\mathbf{q}_4}\frac{1}{4!\Lambda^4}\prod_{a=1}^{4}\varphi'(\mathbf{q}_a)\delta_D^3\left(\sum_{a=1}^{4}\mathbf{q}_\mathbf{a}\right)\,. \tag{B.26}$$

where $\varphi' = \partial_\tau\varphi$, Λ is a coupling constant with dimension of mass and the 4! is for later convenience. Using the same trick as in (B.6), The four-point correlator is

$$B_4 = -\frac{2}{4!\Lambda^4}\,\mathrm{Im}\int_{-\infty}^{\tau}d\tau'\int_{\mathbf{q}_1\dots\mathbf{q}_4}\langle\left[\prod_{a=1}^{4}\varphi'(\mathbf{q}_a,\tau')\right]\left[\prod_{a=1}^{4}\varphi(\mathbf{k}_a,\tau)\right]\rangle\,. \tag{B.27}$$

From the dS mode functions, (2.23), we find

$$f'_k(\tau) = \frac{H}{\sqrt{2k^3}}\partial_\tau\left[(1+ik\tau)e^{-ik\tau}\right] \tag{B.28}$$

$$= \frac{H}{\sqrt{2k^3}}k^2\tau e^{-ik\tau}\,. \tag{B.29}$$

The relevant propagators now are

$$\langle\varphi'(\mathbf{q},\tau')\varphi(\mathbf{k},\tau)\rangle = f'_q(\tau')f^*_k(\tau)(2\pi)^3\delta^3_D(\mathbf{q}+\mathbf{k})\,. \tag{B.30}$$

The correlator becomes

$$B_4 = -\frac{2\times 4!}{4!\Lambda^4}\times\mathrm{Im}\left[\prod_{a=1}^{4}f_{k_a}(\tau)\right]\int_{-\infty}^{\tau}d\tau'\left[\prod_{a=1}^{4}f'_{q_a}(\tau')\right]\,. \tag{B.31}$$

The integral can again be reduced to an exponential integral, but it is now completely finite (again thanks to the $i\epsilon$ rotation of the past infinite boundary). The master integral is

$$\lim_{\tau\to 0}\int_{-\infty(1+i\epsilon)}^{\tau}d\tau' e^{-ik_T\tau'}\left(\tau'\right)^p = -\lim_{\tau\to 0}\eta^{1+p}\,\mathrm{Ei}[-p,ik_T\eta] \tag{B.32}$$

$$= -\frac{(-i)^{p+1}p!}{k_T^{p+1}}\quad\text{for } p\geq 0\,, \tag{B.33}$$

which we will use for $p=4$. Because this has no divergent terms, we can simply take the leading term from $f_k(\tau)$

$$\lim_{\tau\to 0}f_k(\tau) = \frac{H}{\sqrt{2k^3}}\,. \tag{B.34}$$

Finally we find

$$B_4(\tau\to 0) = -\frac{2}{\Lambda^4}\frac{H^8}{2^4(k_1k_2k_3k_4)^{3-2}}\,\mathrm{Im}\left[-\frac{24\,(-i)^5}{k_T^5}\right] \tag{B.35}$$

$$= -\frac{3H^8}{\Lambda^4}\frac{1}{k_T^5k_1k_2k_3k_4}\,. \tag{B.36}$$

Notice that the overall scaling is now $B_4\sim k^{-9}$, which again leads to a scale-invariant trispectrum in position space.

C Graviton polarization tensors

In this appendix, I discuss the graviton polarization tensors. Since all these conditions are invariant under rotations, to find ϵ^s_{ij} explicitly, we can simply choose some convenient $\mathbf{k}$, e.g. $\hat{\mathbf{k}} = \mathbf{k}/k = \hat{\mathbf{z}}$, and then rotate the result. A simple solution to all the conditions in (5.28)-(5.32) is

$$\epsilon^{+2}_{ij}(\hat{\mathbf{z}}) = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & i & 0\\ i & -1 & 0\\ 0 & 0 & 0\end{pmatrix}\quad\text{and}\quad \epsilon^{-2}_{ij}(\hat{\mathbf{z}}) = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & -i & 0\\ -i & -1 & 0\\ 0 & 0 & 0\end{pmatrix}\,. \tag{C.1}$$

This choice has the advantage of describing eigenvectors of rotations around the $\hat{\mathbf{k}}$ axis, namely

$$\epsilon_{ij}^{+2} \to e^{i2\theta}\epsilon_{ij}^{+2}\,, \qquad\qquad \epsilon_{ij}^{-2} \to e^{-i2\theta}\epsilon_{ij}^{-2}\,. \tag{C.2}$$

Indeed they are mapped into each other by parity,

$$\epsilon_{ij}^{+2}(-\mathbf{k}) = \epsilon_{ij}^{+2}(\mathbf{k})^* = \epsilon_{ij}^{-2}(\mathbf{k})\,, \tag{C.3}$$

and viceversa. This is not the only choice since any rotation around $\hat{\mathbf{z}}$ gives a different choice of polarization. More generally, we can use real polarization vectors, which are not eigenstates of rotations. Given wavevector $\mathbf{k}$, we define *real* vectors $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$ to form an orthonormal basis with $\hat{\mathbf{k}} = \mathbf{k}/k$. Then

$$\epsilon_{ij}^{+}(\mathbf{k}) = \hat{u}_i\hat{u}_j - \hat{v}_i\hat{v}_j \quad \text{and} \quad \epsilon_{ij}^{\times}(\mathbf{k}) = \hat{v}_i\hat{u}_j + \hat{v}_j\hat{u}_i\,. \tag{C.4}$$

D Useful formulae

Here I collect some useful formulae and their sign conventions. First GR:

$$S_2 = \int d^4x\sqrt{-g}\left[\frac{M_{\rm Pl}^2}{2}R - \Lambda\right]\,, \qquad R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R + g_{\mu\nu}\Lambda = \frac{1}{M_{\rm Pl}^2}T_{\mu\nu}\,, \tag{D.1}$$

$$R_{\mu\nu} \equiv 2\Gamma^{\rho}_{\mu[\nu,\rho]} + 2\Gamma^{\rho}_{\lambda[\rho}\Gamma^{\Lambda}_{\beta]\alpha}\,, \qquad \Gamma^{\mu}_{\alpha\beta} \equiv \frac{1}{2}g^{\mu\gamma}\left(g_{\alpha\gamma,\beta} + g_{\beta\gamma,\alpha} - g_{\alpha\beta,\gamma}\right)\,, \tag{D.2}$$

$$A^i_{;l} = \frac{\partial A^i}{\partial x^l} + \Gamma^i_{kl}A^k\,, \qquad A_{i;l} = \frac{\partial A_i}{\partial x^l} - \Gamma^k_{il}A_k\,, \tag{D.3}$$

$$A^i_{k;l} = \frac{\partial A^i_k}{\partial x^l} - \Gamma^m_{kl}A^i_m + \Gamma^i_{ml}A^m_k\,, \qquad A_{ik;l} = \frac{\partial A_{ik}}{\partial x^l} - \Gamma^m_{il}A_{mk} - \Gamma^m_{kl}A_{im}\,, \tag{D.4}$$

Then cosmology

$$ds^2 = -dt^2 + a(t)^2\frac{dx^i dx^j\delta_{ij}}{(1+K\mathbf{x}^2/4)^2}\,, \qquad T^{\mu\nu} = \frac{2}{\sqrt{-g}}\frac{\delta S}{\delta g_{\mu\nu}}\,, \tag{D.5}$$

$$T^{\mu}{}_{\nu} = \text{Diag}\left\{-\rho, p, p, p\right\}\,, \qquad \rho = 3M_{\rm Pl}^2\left(H^2 + \frac{K}{a^2}\right)\,, \tag{D.6}$$

$$-\dot{H}M_{\rm Pl}^2 = \frac{1}{2}\left(\rho + p\right)\,, \qquad M_{\rm Pl}^2\frac{\ddot{a}}{a} = -\frac{1}{6}\left(\rho + 3p\right)\,, \tag{D.7}$$

$$\epsilon \equiv -\frac{\dot{H}}{H^2}\,, \qquad \eta_H = \frac{\dot{\epsilon}}{\epsilon H}\,. \tag{D.8}$$

And moving on to field theory

$$X \equiv -\frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi \tag{D.9}$$

$$\varphi(\mathbf{x}) = \int_{\mathbf{k}}\varphi(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{x}}, \qquad \varphi(\mathbf{k}) = \int_{\mathbf{x}}\varphi(\mathbf{x})e^{-i\mathbf{k}\cdot\mathbf{x}}\,, \tag{D.10}$$

$$\varphi(\mathbf{k}, t) = f_k(t)a_{\mathbf{k}} + f_k^*(t)a^{\dagger}_{-\mathbf{k}}\,, \qquad f_k = \frac{H}{\sqrt{2k^3}}(1 + ik\tau)e^{-ik\tau}\,, \tag{D.11}$$

$$\langle\varphi(\mathbf{k})\varphi(\mathbf{k}')\rangle = (2\pi)^3\delta_D^3\left(\mathbf{k} + \mathbf{k}'\right)P(k)\,, \qquad P(k) = \frac{H^2}{2k^3}\,, \tag{D.12}$$

Useful formulae for orthogonal polynomials are Rayleigh expansion

$$e^{i\vec{k}\cdot\vec{r}} = 4\pi\sum_{lm} i^l j_l(kr) Y^*_{lm}(\hat{\vec{k}}) Y_{lm}(\hat{\vec{n}}) = \sum_l (2l+1) i^l j_l(kr) \mathcal{P}_l(\hat{\vec{n}}\cdot\hat{\vec{k}}) \tag{D.13}$$

$$Y_{lm}(\hat{\vec{z}}) = \sqrt{\frac{2l+1}{4\pi}}\delta_{m0} \tag{D.14}$$

$$\delta_{ll'}\delta_{mm'} = \int d^2\Omega\; Y_{lm}(\hat{\vec{n}}) Y^*_{l'm'}(\hat{\vec{n}})\,, \tag{D.15}$$

$$\mathcal{P}_l(\hat{\vec{n}}\cdot\hat{\vec{n}}') = \frac{4\pi}{2l+1}\sum_m Y_{lm}(\hat{\vec{n}}) Y^*_{lm}(\hat{\vec{n}}') \tag{D.16}$$

$$\int d\mu\; \mathcal{P}_l(\mu)\mathcal{P}_{l'}(\mu) = \frac{2}{2l+1}\delta_{l,l'} \tag{D.17}$$

$$(2l+1)\mu P_l(\mu) = (l+1)P_{l+1}(\mu) + lP_{l-1}(\mu)\,. \tag{D.18}$$

E Lesson references and further reading

Sec. 1: A quick review of background cosmology This background material can be found in many excellent textbooks, such as the ones by Weinberg [104], Dodelson [37] (updated by Dodelson and Schmidt in [38]) and Mukhanov [69]. My presentation is based on my own lecture notes for cosmology [71].

Sec. 2: Free fields on curved backgrounds This discussion is based on the nice review by Yi Wang [99]. Weinberg's textbook [104] has also a good discussion that includes dynamical gravity and slow-roll corrections to the mode functions of $\mathcal{R}$. Old textbooks that focus on QFT in curved spacetime are Birrell and Davies [19], Wald [97] and B.de Witt [35].

Sec. 3: Interacting fields and the in-in formalism The standard modern and pedagogical reference for the in-in formalism in cosmology is the appendix of Weinberg's paper on loops in inflation [102]. This focusses on the canonical formalism. Again Wang's lecture notes are also very clear [99]. A reference for the in-in formalism from a path integral perspective (a.k.a. the Schwinger-Keldysh formalism) is [26].

Sec. 4: Correlators from $P(X, \phi)$ theories $P(X, \phi)$ theories were studied in many papers, starting with [7] and [47] in the context of inflation. Chen's review [25] provide some details of the associated primordial non-Gaussianities.

Sec. 5: Pure gravity A very nice discussion of EFT's, on which Sec. 5.1 is based was given by Polchinski in the beautiful lecture notes [78]. A nice and pedagogical introduction to general relativity as an EFT can be found in Donoghue lecture notes [39]. The explicit expansion of the Einstein-Hilbert action to quadratic order in perturbations can be found in Sec. 5.1 of [104] and to cubic and quartic order in [20].

Sec. 6: Gravity and matter The notation for the SVT decomposition in Sec. 6.3 is that of Weinberg [104].

Sec. 7: Symmetries and soft theorems The presentation in Sec. 7.4 parallels in spirit that of [53], but instead of the Schrödinger picture of "wave functionals" I use the more standard interaction picture.

Sec. 8: Phenomenology of primordial non-Gaussianity

Sec. 9: Dynamics in the Newtonian regime The basics of clustering statistics and perturbation theory for large scale structures are nicely summarized in the classic review [18]. Particularly insightful is the discussion in Peebles' book [76], although the observational part is outdated.

Sec. 10: Standard Perturbation Theory Again the classic review [18] is a good reference for the Standard Perturbation Theory (SPT) approach to include non-linear evolution in structure formation.

Sec. 11: The effective field theory approach The original references that developed this approach are [17] and [23]. The relative importance of terms in a scaling universe and renormalizability were discussed in [75].

References

[1] P. A. R. Ade et al. Planck 2015 results. XIII. Cosmological parameters. Astron. Astrophys., 594:A13, 2016.

[2] P. A. R. Ade et al. Improved Constraints on Primordial Gravitational Waves using Planck, WMAP, and BICEP/Keck Observations through the 2018 Observing Season. Phys. Rev. Lett., 127(15):151301, 2021.

[3] Y. Akrami et al. Planck 2018 results. X. Constraints on inflation. Astron. Astrophys., 641:A10, 2020.

[4] Marcelo Alvarez, Tobias Baldauf, J Richard Bond, Neal Dalal, Roland de Putter, Olivier Doré, Daniel Green, Chris Hirata, Zhiqi Huang, Dragan Huterer, et al. Testing inflation with large scale structure: connecting hopes with reality. arXiv preprint arXiv:1412.4671, 2014.

[5] Nima Arkani-Hamed, Paolo Benincasa, and Alexander Postnikov. Cosmological Polytopes and the Wavefunction of the Universe. 2017.

[6] Nima Arkani-Hamed and Juan Maldacena. Cosmological Collider Physics 2. 2015.

[7] C. Armendariz-Picon, T. Damour, and Viatcheslav F. Mukhanov. k - inflation. Phys. Lett., B458:209–218, 1999.

[8] C. Armendariz-Picon, Viatcheslav F. Mukhanov, and Paul J. Steinhardt. A Dynamical solution to the problem of a small cosmological constant and late time cosmic acceleration. Phys. Rev. Lett., 85:4438–4441, 2000.

[9] Richard L. Arnowitt, Stanley Deser, and Charles W. Misner. Dynamical Structure and Definition of Energy in General Relativity. Phys. Rev., 116:1322–1330, 1959.

[10] Valentin Assassi, Daniel Baumann, and Daniel Green. Symmetries and Loops in Inflation. JHEP, 02:151, 2013.

[11] Daniel Babich, Paolo Creminelli, and Matias Zaldarriaga. The Shape of non-Gaussianities. JCAP, 0408:009, 2004.

[12] Henri Bacry and Jean-Marc Lévy-Leblond. Possible kinematics. Journal of Mathematical Physics, 9(10):1605–1614, 1968.

[13] Tobias Baldauf, Lorenzo Mercolli, Mehrdad Mirbabayi, and Enrico Pajer. The Bispectrum in the Effective Field Theory of Large Scale Structure. JCAP, 05:007, 2015.

[14] Tobias Baldauf, Uros Seljak, Leonardo Senatore, and Matias Zaldarriaga. Galaxy Bias and non-Linear Structure Formation in General Relativity. JCAP, 1110:031, 2011.

[15] James M Bardeen. Gauge-invariant cosmological perturbations. Physical Review D, 22(8):1882, 1980.

[16] James M. Bardeen, J. R. Bond, Nick Kaiser, and A. S. Szalay. The Statistics of Peaks of Gaussian Random Fields. Astrophys. J., 304:15–61, 1986.

[17] Daniel Baumann, Alberto Nicolis, Leonardo Senatore, and Matias Zaldarriaga. Cosmological Non-Linearities as an Effective Fluid. JCAP, 1207:051, 2012.

[18] F. Bernardeau, S. Colombi, E. Gaztanaga, and R. Scoccimarro. Large scale structure of the universe and cosmological perturbation theory. Phys. Rept., 367:1–248, 2002.

[19] Nicholas David Birrell and PCW Davies. Quantum fields in curved space. 1984.

[20] James Bonifacio, Harry Goodhew, Austin Joyce, Enrico Pajer, and David Stefanyszyn. The graviton four-point function in de Sitter space. 12 2022.

[21] Lorenzo Bordin, Paolo Creminelli, Andrei Khmelnitsky, and Leonardo Senatore. Light Particles with Spin in Inflation. JCAP, 10:013, 2018.

[22] Giovanni Cabass, Enrico Pajer, and Fabian Schmidt. How Gaussian can our Universe be? JCAP, 1701(01):003, 2017.

[23] John Joseph M. Carrasco, Mark P. Hertzberg, and Leonardo Senatore. The Effective Field Theory of Cosmological Large Scale Structures. JHEP, 09:082, 2012.

[24] Anthony Challinor and Hiranya Peiris. Lecture notes on the physics of cosmic microwave background anisotropies. In Mario Novello and Santiago Perez, editors, American Institute of Physics Conference Series, volume 1132 of American Institute of Physics Conference Series, pages 86–140, May 2009.

[25] Xingang Chen. Primordial Non-Gaussianities from Inflation Models. Adv. Astron., 2010:638979, 2010.

[26] Xingang Chen, Yi Wang, and Zhong-Zhi Xianyu. Schwinger-Keldysh Diagrammatics for Primordial Perturbations. JCAP, 1712(12):006, 2017.

[27] Paolo Creminelli. Conformal invariance of scalar perturbations in inflation. Phys. Rev., D85:041302, 2012.

[28] Paolo Creminelli, Austin Joyce, Justin Khoury, and Marko Simonovic. Consistency Relations for the Conformal Mechanism. JCAP, 04:020, 2013.

[29] Paolo Creminelli, Markus A. Luty, Alberto Nicolis, and Leonardo Senatore. Starting the Universe: Stable Violation of the Null Energy Condition and Non-standard Cosmologies. JHEP, 12:080, 2006.

[30] Paolo Creminelli, Jorge Noreña, Marko Simonović, and Filippo Vernizzi. Single-Field Consistency Relations of Large Scale Structure. JCAP, 12:025, 2013.

[31] Paolo Creminelli, Jorge Norena, and Marko Simonovic. Conformal consistency relations for single-field inflation. JCAP, 1207:052, 2012.

[32] Paolo Creminelli, Ashley Perko, Leonardo Senatore, Marko Simonovi?, and Gabriele Trevisan. The Physical Squeezed Limit: Consistency Relations at Order q^2. JCAP, 1311:015, 2013.

[33] Liang Dai, Enrico Pajer, and Fabian Schmidt. Conformal Fermi Coordinates. JCAP, 1511(11):043, 2015.

[34] Liang Dai, Enrico Pajer, and Fabian Schmidt. On Separate Universes. JCAP, 1510(10):059, 2015.

[35] Bryce S DeWitt. Quantum field theory in curved spacetime. Physics Reports, 19(6):295–357, 1975.

[36] Scott Dodelson. Modern cosmology. Academic Press, San Diego, CA, 2003.

[37] Scott Dodelson. Modern Cosmology. Academic Press, Amsterdam, 2003.

[38] Scott Dodelson and Fabian Schmidt. Modern cosmology. Academic Press, 2020.

[39] John F. Donoghue. General relativity as an effective field theory: The leading quantum corrections. Phys. Rev., D50:3874–3888, 1994.

[40] S. Dubovsky, T. Gregoire, A. Nicolis, and R. Rattazzi. Null energy condition and superluminal propagation. JHEP, 03:025, 2006.

[41] Ruth Durrer. The Cosmic Microwave Background. Cambridge Univ. Press, Cambridge, 2008.

[42] Bernardo Finelli, Garrett Goon, Enrico Pajer, and Luca Santoni. Soft Theorems For Shift-Symmetric Cosmologies. Phys. Rev., D97(6):063531, 2018.

[43] Bernardo Finelli, Garrett Goon, Enrico Pajer, and Luca Santoni. The Effective Theory of Shift-Symmetric Cosmologies. JCAP, 1805(05):060, 2018.

[44] D. J. Fixsen, E. S. Cheng, J. M. Gales, John C. Mather, R. A. Shafer, and E. L. Wright. The Cosmic Microwave Background spectrum from the full COBE FIRAS data set. Astrophys. J., 473:576, 1996.

[45] Tian-Fu Fu and Qing-Guo Huang. The four-point correlation function of graviton during inflation. JHEP, 07:132, 2015.

[46] T. Garidi. What is mass in de Sitterian physics? 9 2003.

[47] Jaume Garriga and Viatcheslav F. Mukhanov. Perturbations in k-inflation. Phys. Lett., B458:219–225, 1999.

[48] Harry Goodhew. The Cosmological Implications of Unitarity. PhD thesis, Cambridge U., DAMTP, 2023.

[49] M. H. Goroff, Benjamin Grinstein, S. J. Rey, and Mark B. Wise. Coupling of Modes of Cosmological Mass Density Fluctuations. Astrophys. J., 311:6–14, 1986.

[50] Alan H. Guth. The Inflationary Universe: A Possible Solution to the Horizon and Flatness Problems. Phys. Rev., D23:347–356, 1981.

[51] Jemal Guven, Ben Lieberman, and Christopher T. Hill. Schrodinger Picture Field Theory in Robertson-walker Flat Space-times. Phys. Rev., D39:438, 1989.

[52] Kurt Hinterbichler, Lam Hui, and Justin Khoury. Conformal Symmetries of Adiabatic Modes in Cosmology. JCAP, 1208:017, 2012.

[53] Kurt Hinterbichler, Lam Hui, and Justin Khoury. An Infinite Set of Ward Identities for Adiabatic Modes in Cosmology. JCAP, 1401:039, 2014.

[54] Wayne Hu. Concepts in CMB anisotropy formation. Lect. Notes Phys., 470:207, 1996.

[55] Wayne Hu and Naoshi Sugiyama. Anisotropies in the cosmic microwave background: An Analytic approach. Astrophys. J., 444:489–506, 1995.

[56] Lam Hui, Austin Joyce, and Sam S. C. Wong. Inflationary soft theorems revisited: A generalized consistency relation. JCAP, 1902:060, 2019.

[57] Erdal Inonu and Eugene Paul Wigner. On the contraction of groups and their representations. Proceedings of the National Academy of Sciences, 39(6):510–524, 1953.

[58] Bhuvnesh Jain and Edmund Bertschinger. Selfsimilar evolution of cosmological density fluctuations. Astrophys. J., 456:43, 1996.

[59] A. Kehagias and A. Riotto. Symmetries and Consistency Relations in the Large Scale Structure of the Universe. Nucl. Phys. B, 873:514–529, 2013.

[60] Eiichiro Komatsu and David N. Spergel. Acoustic signatures in the primary microwave background bispectrum. Phys. Rev., D63:063002, 2001.

[61] David Langlois and Karim Noui. Degenerate higher derivative theories beyond Horndeski: evading the Ostrogradski instability. JCAP, 02:034, 2016.

[62] Hayden Lee, Daniel Baumann, and Guilherme L. Pimentel. Non-Gaussianity as a Particle Detector. JHEP, 12:040, 2016.

[63] Mang Hei Gordon Lee, Ciaran McCulloch, and Enrico Pajer. Leading Loops in Cosmological Correlators. 5 2023.

[64] Shing Yan Li, Yi Wang, and Siyi Zhou. KLT-Like Behaviour of Inflationary Graviton Correlators. JCAP, 12:023, 2018.

[65] Andrei D. Linde. A New Inflationary Universe Scenario: A Possible Solution of the Horizon, Flatness, Homogeneity, Isotropy and Primordial Monopole Problems. Adv. Ser. Astrophys. Cosmol., 3:149–153, 1987.

[66] David H. Lyth, Karim A. Malik, and Misao Sasaki. A General proof of the conservation of the curvature perturbation. JCAP, 05:004, 2005.

[67] Juan Martin Maldacena. Non-Gaussian features of primordial fluctuations in single field inflationary models. JHEP, 05:013, 2003.

[68] Lorenzo Mercolli and Enrico Pajer. On the velocity in the Effective Field Theory of Large Scale Structures. JCAP, 1403:006, 2014.

[69] V. Mukhanov. Physical Foundations of Cosmology. Cambridge University Press, Oxford, 2005.

[70] Hugh Osborn. Lectures on Conformal Field Theories in more than two dimensions. 2010.

[71] Enrico Pajer. Cosmology lecture notes. 2018.

[72] Enrico Pajer and Sadra Jazayeri. Systematics of Adiabatic Modes: Flat Universes. JCAP, 1803(03):013, 2018.

[73] Enrico Pajer, Guilherme L. Pimentel, and Jaap V. S. Van Wijck. The Conformal Limit of Inflation in the Era of CMB Polarimetry. JCAP, 1706(06):009, 2017.

[74] Enrico Pajer, Fabian Schmidt, and Matias Zaldarriaga. The Observed Squeezed Limit of Cosmological Three-Point Functions. Phys. Rev., D88(8):083502, 2013.

[75] Enrico Pajer and Matias Zaldarriaga. On the Renormalization of the Effective Field Theory of Large Scale Structures. JCAP, 1308:037, 2013.

[76] Phillip James Edwin Peebles. The large-scale structure of the universe. Princeton university press, 2020.

[77] Joseph Polchinski. Renormalization and Effective Lagrangians. Nucl. Phys., B231:269–295, 1984.

[78] Joseph Polchinski. Effective field theory and the Fermi surface. In Theoretical Advanced Study Institute (TASI 92): From Black Holes and Strings to Particles Boulde pages 0235–276, 1992.

[79] Suvrat Raju. Four Point Functions of the Stress Tensor and Conserved Currents in AdS_4/CFT_3. Phys. Rev. D, 85:126008, 2012.

[80] V. A. Rubakov. Phantom without UV pathology. Theor. Math. Phys., 149:1651–1664, 2006. [Teor. Mat. Fiz.149,409(2006)].

[81] Slava Rychkov. EPFL Lectures on Conformal Field Theory in $D \geq 3$ Dimensions. SpringerBriefs in Physics. 2016.

[82] D. S. Salopek and J. R. Bond. Nonlinear evolution of long wavelength metric fluctuations in inflationary models. Phys. Rev. D, 42:3936–3962, 1990.

[83] Misao Sasaki and Ewan D. Stewart. A General analytic formula for the spectral index of the density perturbations produced during inflation. Prog. Theor. Phys., 95:71–78, 1996.

[84] Koenraad Schalm, Gary Shiu, and Ted van der Aalst. Consistency condition for inflation from (broken) conformal symmetry. JCAP, 03:005, 2013.

[85] Roman Scoccimarro and Joshua Frieman. Loop corrections in nonlinear cosmological perturbation theory. Astrophys. J. Suppl., 105:37, 1996.

[86] David Seery, James E. Lidsey, and Martin S. Sloth. The inflationary trispectrum. JCAP, 01:027, 2007.

[87] Uros Seljak. A Two fluid approximation for calculating the cosmic microwave background anisotropies. Astrophys. J., 435:L87–L90, 1994.

[88] Uros Seljak and Matias Zaldarriaga. A Line of sight integration approach to cosmic microwave background anisotropies. Astrophys. J., 469:437–444, 1996.

[89] David Simmons-Duffin. The Conformal Bootstrap. In Proceedings, New Frontiers in Fields and Strings (TASI 2015): Boulder, CO, USA, June 1-26, 2015, pages 1–74, 2017.

[90] D. T. Son. Low-energy quantum effective action for relativistic superfluids. 2002.

[91] Alexei A. Starobinsky. A New Type of Isotropic Cosmological Models Without Singularity. Adv. Ser. Astrophys. Cosmol., 3:130–133, 1987.

[92] G 't Hooft. Naturalness, chiral symmetry, and spontaneous chiral symmetry breaking. Recent developments in gauge theories, pages 135–157, 1980.

[93] Gerard 't Hooft and M. J. G. Veltman. One loop divergencies in the theory of gravitation. Ann. Inst. H. Poincare Phys. Theor., A20:69–94, 1974.

[94] Ryuichi Takahashi. Third Order Density Perturbation and One-loop Power Spectrum in a Dark Energy Dominated Universe. Prog. Theor. Phys., 120:549–559, 2008.

[95] Atsushi Taruya and Stéphane Colombi. Post-collapse perturbation theory in 1D cosmology – beyond shell-crossing. Mon. Not. Roy. Astron. Soc., 470(4):4858–4884, 2017.

[96] David Tong. Quantum Field Theory. 2006.

[97] Robert M Wald. Quantum field theory in curved spacetime and black hole thermodynamics. University of Chicago press, 1994.

[98] David Wands. Local non-Gaussianity from inflation. Class. Quant. Grav., 27:124002, 2010.

[99] Yi Wang. Inflation, Cosmic Perturbations and Non-Gaussianities. Commun. Theor. Phys., 62:109–166, 2014.

[100] Steven Weinberg. Adiabatic modes in cosmology. Phys. Rev., D67:123504, 2003.

[101] Steven Weinberg. Damping of tensor modes in cosmology. Phys. Rev., D69:023503, 2004.

[102] Steven Weinberg. Quantum contributions to cosmological correlations. Phys. Rev., D72:043514, 2005.

[103] Steven Weinberg. The Quantum theory of fields. Vol. 1: Foundations. Cambridge University Press, 2005.

[104] Steven Weinberg. Cosmology. 2008.

[105] Edward Witten. Anti-de Sitter space and holography. Adv. Theor. Math. Phys., 2:253–291, 1998.

Field Theory in Cosmology: Example Sheet 1

1. For a $P(X,\phi)$ theory

$$S = \int \sqrt{-g} P(X,\phi)\,, \tag{1}$$

compute the equations of motion. Compute the energy-momentum tensor and find the identification upon which it reduces to that of a perfect fluid

$$T_{\mu\nu} = (\rho + p)\, u_\mu u_\nu + g_{\mu\nu} p\,. \tag{2}$$

Re-derive the equations of motion by combining the two Friedmann equations, which for a perfect fluid take the general form

$$3M_{\mathrm{Pl}}^2 H^2 = \rho\,, \qquad -\dot{H} M_{\mathrm{Pl}}^2 = \frac{1}{2}\left(\rho + p\right)\,. \tag{3}$$

2. Compute the power spectrum of a massive scalar field in de Sitter. Consider the action

$$S = -\int \sqrt{-g} \frac{1}{2}\left[\partial_\mu \phi \partial^\mu \phi - m^2 \phi^2\right]\,, \tag{4}$$

for some mass m. Write $\phi(\mathbf{k})$ in terms of creation and annihilation operators $\{a_{\mathbf{k}}, a^\dagger_{\mathbf{k}}\}$ and mode functions f_k. Derive the equation that $f_k(\tau)$ has to satisfy from the action (4), using conformal time. To solve this equation, re-write it as an equation for $g_k = (-\tau)^{-3/2} f_k$, and then use the fact that the two linear independent solution of Bessel's differential equation,

$$x^2 \partial_x^2 y + x\partial_x y + (x^2 - \alpha^2) y = 0\,, \tag{5}$$

can be taken to be the two Hankel functions $H_\alpha^{(1,2)}$. Now that you have the most general solution for f_k, with two integration constant, match this solution in the $-k\tau \to \infty$ limit to the flat space solution. You may use the following expansions of the Hankel functions for $x \to \infty$

$$H_\alpha^{(1)}(x) \simeq \sqrt{\frac{2}{\pi}} \frac{e^{ix}}{\sqrt{x}}\,, \qquad H_\alpha^{(2)}(x) \simeq \sqrt{\frac{2}{\pi}} \frac{e^{-ix}}{\sqrt{x}}\,, \tag{6}$$

which are valid up to an irrelevant (α-dependent) phase. You should find

$$f_k(\tau) = \frac{\sqrt{\pi} H}{2} (-\tau)^{3/2} H_\nu^{(1)}(-k\tau)\,, \quad \nu = \sqrt{\frac{9}{4} - \frac{m^2}{H^2}}\,. \tag{7}$$

3. Compute the two-point correlators

$$\lim_{\tau\to 0} \langle \phi(\mathbf{k}) \pi(\mathbf{k}')\rangle = (2\pi)^3 \delta_D^3\left(\mathbf{k} + \mathbf{k}'\right) \frac{1}{2k\tau}\,, \tag{8}$$

$$\lim_{\tau\to 0} \langle \phi(\mathbf{k}) \pi(\mathbf{k}')\rangle = \lim_{\tau\to 0} \langle \pi(\mathbf{k}) \phi(\mathbf{k}')\rangle\,, \tag{9}$$

$$\lim_{\tau\to 0} \langle \pi(\mathbf{k}) \pi(\mathbf{k}')\rangle = (2\pi)^3 \delta_D^3\left(\mathbf{k} + \mathbf{k}'\right) \frac{k}{2H^2\tau^2}\,. \tag{10}$$

4. Using the power spectrum derived in the lecture, compute the (real space) correlation function at separate points for a massless scalar field in dS and show that it is IR divergent:

$$\langle \phi(\mathbf{x}) \phi(\mathbf{0})\rangle = \frac{H^2}{(2\pi)^2} \int_0^\infty d\tilde{k}\, \frac{\sin \tilde{k}}{\tilde{k}^2}\,. \tag{11}$$

5. Compute the amount of particle production in dS. In the lectures, we fixed the mode functions by demanding that φ creates positive-energy particles at $k\tau \to -\infty$. Let's instead require that φ creates positive-energy particles at some finite $|\tau_*| > \infty$, still satisfying $|k\tau_*| \gg 1$. The quantized field then takes the form

$$\varphi(\mathbf{k}) = g_k b_{\mathbf{k}} + g_k^* b^\dagger_{-\mathbf{k}} \,, \tag{12}$$

where $\{b_{\mathbf{k}}, b^\dagger_{\mathbf{k}}\}$ are a new set of creation and annihilation operators. Define the new vacuum state $|\tilde{0}\rangle$. Find g_k by matching to the Minkowski vacuum at τ_* (you may multiply g_k by a convenient phase)

$$g_k = \frac{H}{\sqrt{2k^3}}\left[1 + \frac{i}{k\tau_*} - \frac{1}{2(k\tau_*)^2}\right] f_k(\tau) + e^{-2ik\tau_*}\frac{H}{\sqrt{2k^3}}\frac{1}{2(k\tau_*)^2} f_k^*(\tau) \,,$$

By matching (12) to the expressions for $\varphi(\mathbf{k})$ we found in the lectures (i.e. matching to Minkowski at $|\tau_*| \to \infty$), show that the two sets of ladder operators are related,

$$a_{\mathbf{k}} = \frac{\sqrt{2k^3}}{H}\left(\alpha b_{\mathbf{k}} + \beta^* b^\dagger_{-\mathbf{k}}\right) \,, \qquad a^\dagger_{\mathbf{k}} = \frac{\sqrt{2k^3}}{H}\left(\beta b_{-\mathbf{k}} + \alpha^* b^\dagger_{\mathbf{k}}\right) \,, \tag{13}$$

This relation is called a *Bogoliubov transformation.* Invert it to give

$$b_{\mathbf{k}} = \frac{\sqrt{2k^3}}{H}\left(\alpha^* a_{\mathbf{k}} + \beta^* a^\dagger_{-\mathbf{k}}\right) \,, \qquad b^\dagger_{\mathbf{k}} = \frac{\sqrt{2k^3}}{H}\left(\beta a_{-\mathbf{k}} + \alpha a^\dagger_{\mathbf{k}}\right) \,, \tag{14}$$

Now we want to ask what a detector that measures $b_k^\dagger$ excitations would measure in the Bunch Davies vacuum $|0\rangle$, which we defined in the lecture as $a_{\mathbf{k}}|0\rangle = 0$. To this end, let's define the "b-particle" number operator

$$N_b(\mathbf{k}) = b^\dagger_{\mathbf{k}} b_{\mathbf{k}} \,. \tag{15}$$

Compute the expectation value of $N_b(\mathbf{k})$ on the state $|\tilde{0}\rangle$ and on the Bunch-Davies vacuum $|0\rangle$. To understand the singular factor $\delta_D^3(\mathbf{0})$, work at finite volume

$$(2\pi)^3\delta_D^3(\mathbf{0}) = \lim_{V\to\infty}\int_V d^3x e^{-i\mathbf{0}\cdot\mathbf{x}} = \lim_{V\to\infty} V \,, \tag{16}$$

and define the number density of particles, $n_b(\mathbf{k}) \equiv N_b(\mathbf{k})/V$, instead of the total number $N_b(\mathbf{k})$. You should find that the Bunch-Davies state has a non-vanishing density of b-type particles given by

$$\langle 0| n_b(\mathbf{k}) |0\rangle = \frac{1}{4(k\tau)^4} \neq 0 \,. \tag{17}$$

6. The fact that an FLRW background is invariant under translations, $\mathbf{x} \to \mathbf{x} + \mathbf{b}$, implies that also correlators must be invariant

$$\langle \phi(\mathbf{x}_1)\dots\phi(\mathbf{x}_n)\rangle = \langle \phi(\mathbf{x}_1 + \mathbf{b})\dots\phi(\mathbf{x}_n + \mathbf{b})\rangle \,. \tag{18}$$

Using this, prove that momentum space correlators must always be proportional to a delta function of the total momentum

$$\langle \phi(\mathbf{k}_1)\dots\phi(\mathbf{k}_n)\rangle \propto \delta_D^3\left(\sum_{a=1}^n \mathbf{k}_a\right) \,. \tag{19}$$

7. For the metric

$$\begin{aligned} ds^2 &= -dt^2 + a^2\left(\delta_{ij} + \gamma_{ij}\right) dx^i dx^j && (20)\\ &= \frac{1}{H^2\tau^2}\left[-d\tau^2 + \left(\delta_{ij} + \gamma_{ij}\right) dx^i dx^j\right] \,, && (21) \end{aligned}$$

where $\gamma_{ii} = \partial_i \gamma_{ij} = 0$, we want to expand the Einstein-Hilbert action in de Sitter

$$S_2 = \int d^4x \sqrt{-g} \left[\frac{M_{\rm Pl}^2}{2} R - \Lambda \right] \tag{22}$$

to second order in γ to find the action of a free graviton. You already performed a similar expansion around Minkowski in the General Relativity course and it was a painful calculation. Instead of doing it again, let's use a trick. Start by noticing that the dS metric is proportional to the Minkowski one

$$g_{\mu\nu}^{\rm dS} = a^2 g_{\mu\nu}^{\rm Mink} = \frac{1}{H^2 \tau^2} g_{\mu\nu}^{\rm Mink} \,, \tag{23}$$

with the identification $\tau^{(\rm dS)} = t^{(\rm Mink)}$. Notice that by the Friedmann equation

$$3 M_{\rm Pl}^2 H^2 = \Lambda \tag{24}$$

A metric with this property is called *conformally flat*. Given an arbitrary function Ω of the coordinates, the rescaling

$$g_{\mu\nu} \to \tilde{g}_{\mu\nu} = \Omega^2 g_{\mu\nu} \,, \tag{25}$$

is called a Weyl transformation. Various GR tensors transform quite easily under a Weyl rescaling. For example, the Ricci scalars $\tilde{R} \equiv \tilde{g}^{\mu\nu} \tilde{R}_{\mu\nu}$ and $R \equiv g^{\mu\nu} R_{\mu\nu}$ for the two metrics are related by [this can be proven by direct calculation, if you wish]

$$\tilde{R} = \Omega^{-2} \left[R - 6 \nabla_\mu \nabla^\mu \ln \Omega - 6 (\nabla_\mu \ln \Omega)(\nabla^\mu \ln \Omega) \right] . \tag{26}$$

Now recall that in Minkowski, you found

$$\begin{aligned} S_2^{\rm Mink} &= \frac{M_{\rm Pl}^2}{2} \int d^4x \sqrt{-g} R & (27) \\ &= \frac{M_{\rm Pl}^2}{8} \int d^3x dt \left[\dot{\gamma}_{ij} \dot{\gamma}_{ij} - \partial_k \gamma_{ij} \partial_k \gamma_{ij} \right] + \mathcal{O}(\gamma^3) \quad \text{(Minkowski)} & (28) \end{aligned}$$

Use (26) to rewrite the Einstein-Hilbert action around dS in terms of that around Minkowski, for which you can use the expansion above. You should find that around dS the graviton free action is

$$S_2 = \frac{M_{\rm Pl}^2}{8} \int d^3x d\tau \, a^2 \left[\gamma'_{ij} \gamma'_{ij} - \partial_k \gamma_{ij} \partial_k \gamma_{ij} \right] . \tag{29}$$

8. Prove that the two in-in expressions for a generic in-in correlator

$$\begin{aligned} \langle \mathcal{O}(t) \rangle &= \sum_{N=0}^{\infty} i^N \int_{-\infty}^{t} dt_N \int_{-\infty}^{t_N} dt_{N-1} \ldots \int_{-\infty}^{t_2} dt_1 & (30) \\ &\quad \times \langle 0 | [\hat{H}_{int}(t_1), [\hat{H}_{int}(t_2), \ldots [\hat{H}_{int}(t_N), \mathcal{O}(t)] \ldots]] | 0 \rangle \,, \\ \langle \mathcal{O}(t) \rangle &= \langle 0 | \left[\bar{T} e^{\left(i \int_{-\infty(1+i\epsilon)}^{t} dt' \hat{H}_{int}(t') \right)} \right] \mathcal{O}(t) \left[T e^{\left(-i \int_{-\infty(1-i\epsilon)}^{t} dt' \hat{H}_{int}(t') \right)} \right] | 0 \rangle \,, & (31) \end{aligned}$$

are indeed equivalent. Proceed by induction. First prove that they are equivalent at order $N = 0$ and $N = 1$. Then, assuming that they agree at order $N - 1$, take the time derive of each Nth-order expression and rewrite it as the correlators of some other field to order $N - 1$. This proves that the expression agree to order N up to a constant. By taking the limit $t \to -\infty$ show that the constant has to vanish.

9. Using the in-in formalism, compute the bispectrum in a $P(X)$ theory induced by the interactions $\dot{\varphi}^3$ and $\dot{\varphi}(\partial_i \varphi)^2$.

10. The fact that the de Sitter metric,

$$ds^2 = \frac{-d\tau^2 + dx^i dx^j \delta_{ij}}{\tau^2 H^2} \,, \tag{32}$$

is invariant under dilations, $\{\tau, \mathbf{x}\} \to \lambda\{\tau, \mathbf{x}\}$, implies that equal time correlators that do not depend on time, such as for example the power spectrum of a massless scalar field or of the graviton at $\tau \to 0$, must obey

$$\langle \phi(\mathbf{x}_1)\phi(\mathbf{x}_2)\ldots\phi(\mathbf{x}_n)\rangle = \langle \phi(\lambda\mathbf{x}_1)\phi(\lambda\mathbf{x}_2)\ldots\phi(\lambda\mathbf{x}_3)\rangle\,. \tag{33}$$

Using this, prove that momentum space correlators B_n, defined as

$$\langle \phi(\mathbf{k}_1)\ldots\phi(\mathbf{k}_n)\rangle = (2\pi)^3\delta^3_D\left(\sum_{a=1}^{n}\mathbf{k}_a\right)B_n(\mathbf{k}_1,\ldots,\mathbf{k}_n)\,. \tag{34}$$

must scale as

$$B_n(\lambda\mathbf{k}_1,\ldots,\lambda\mathbf{k}_n) = \frac{1}{\lambda^{3(n-1)}}B_n(\mathbf{k}_1,\ldots,\mathbf{k}_n)\,. \tag{35}$$

Field Theory in Cosmology: Example Sheet 2

1. Reproduce the constraint equations by varying the action

$$S = \int d^4x\sqrt{h}N\left\{\frac{M_{\rm Pl}^2}{2}\left[{}^{(3)}R + K_{ij}K^{ij} - K^2\right] + P(X,\phi)\right\} . \tag{1}$$

   with respect to N and N^i.

2. Derive the linear-order gauge transformations of A, B, ψ and h_{00}, for a generic change of coordinates $x^\mu \to x^\mu + \epsilon^\mu$.

3. Derive the gauge transformation from Newtonian gauge to flat gauge and viceversa. In particular, given some generic perturbations $\{A^N, h_{00}^N, \varphi^N\}$ in Newtonian gauge, determine teh corresponding perturbations $\{h_{00}^f, \psi^f, \varphi^f\}$ in flat gauge.

4. Solve the $\delta S/\delta N^i$ constraint to find δN, working in flat gauge to linear order.

5. In the lecture, we prove the conservation of $\mathcal{R}$ and γ_{ij} on superHubble scales in the presence of a generic energy-momentum tensor $T_{\mu\nu}$ by working in comoving gauge. Prove again the conservation of $\mathcal{R}$ by working in Newtonian gauge. In particular, you might want to start with the change of coordinates

$$\epsilon^\mu = \left\{\epsilon(t), \lambda x^i\right\} . \tag{2}$$

   and show that the gauge transformations are

$$\begin{aligned}
&\Phi = -\dot\epsilon\,, && \Psi = H\epsilon - \frac{\lambda}{3}\,. && \\
&\delta\rho = -\dot{\bar\rho}\epsilon\,, && \delta u = \epsilon\,, && \pi^S = 0\,, \qquad (3)\\
&\delta p = -\dot{\bar p}\epsilon\,, && \varphi = -\epsilon\dot{\bar\phi}\,. && \qquad (4)
\end{aligned}$$

   Then use the scalar part of the ij components of the Einstein's equation,

$$k_ik_j\left(\Phi - \Psi\right) = 0, \tag{5}$$

   to impose the physicality condition on $\epsilon(t)$. Your final result should be

$$\mathcal{R} = \frac{\lambda}{3}\,, \qquad \varphi = -\dot{\bar\phi}\frac{\mathcal{R}}{a}\int_T^t a(t')dt'\,, \qquad \Phi = \Psi = \mathcal{R}\left[-1 + \frac{H}{a}\int_T^t a(t')dt'\right] . \tag{6}$$

6.