A search for transient reductions in the inflaton speed of sound in cosmological data, and other topics

Torrado Cacho, J.

Chapter 1

Introduction

In this chapter, we review how to characterise inflation through cosmological observations. After a brief description of FLRW cosmology, we describe slow-roll inflation and the effect of possible transient reductions in the speed of sound of the inflaton. Later we describe the Cosmic Microwave Background and how it can be used to constrain features in the inflationary potential. Finally, we review the statistical tools commonly used to achieve the characterisation of inflation through general cosmological data sets.

For the sake of brevity, most of the detailed calculations are not shown. Everywhere in this thesis, we use natural units, $c = \hbar = 1$, but keep the Newton's gravitational constant (and thus Planck mass) explicit .

1.1 Background Cosmology – Large scale geometry of the universe

1.1.1 Description and evolution of homogeneous cosmologies

We observe the CMB radiation and the LSS to be on large scales mostly isotropic. Isotropy from here, combined with the Copernican principle – *our place in the Universe is not special* – implies homogeneity of the space.

In this thesis we will always assume the space-time to be described on cosmological scales as a perfect fluid, determined by only two quantities, its density ρ and pressure p, which is by definition homogeneous, isotropic, having no viscosity, shear tension or energy conduction – it can just uniformly compress or expand.

To describe this system in the context of General Relativity, one normally chooses a *comoving frame* in which the 3-surfaces of homogeneity are synchronous and the same 3-d coordinates are assigned to the same particle of the fluid at all times. In this frame, the metric describing the very large scales reads (we take

$c = 1$)

$$\mathrm{d}s^2 = g_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu \coloneqq -\mathrm{d}t^2 + a^2(t)\,\mathrm{d}l^2\ , \tag{1.1}$$

where $a(t)$ is a *scale factor* and $\mathrm{d}l^2$ is the metric of the 3-space:

$$\mathrm{d}l^2 \coloneqq \mathrm{d}\chi^2 + \Sigma^2(\chi)\left(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2\right)\ , \quad \Sigma(\chi) = \begin{cases} \sinh\chi & \text{if} \quad \kappa = -1 \\ \chi & \text{if} \quad \kappa = 0 \\ \sin\chi & \text{if} \quad \kappa = +1 \end{cases}\ , \tag{1.2}$$

where κ is a parameter defining the 3-curvature, proportional to the Ricci scalar of the 3-space.

We often define the *conformal time*, τ, by factoring out the scale factor:

$$\mathrm{d}s^2 = a^2(\tau)\left(-\mathrm{d}\tau^2 + \mathrm{d}l^2\right), \qquad \mathrm{d}\tau = \frac{\mathrm{d}t}{a}\ . \tag{1.3}$$

On the other side of the Einstein equation, the *matter* side, we have the stress-energy tensor of a perfect fluid in comoving coordinates, which is simply

$$T_{\mu\nu} = \mathrm{diag}\left(\rho, p\,g_{ii}\right)\ . \tag{1.4}$$

Finally, from the Einstein equation and the continuity equation $\nabla_\alpha T^{\alpha\mu} = 0$ we derive the commonly used pair of Friedmann equations

$$3H^2 + 3\frac{\kappa}{a^2} - \Lambda = 8\pi G\,\rho\ , \tag{1.5a}$$

$$\dot{\rho} + 3H\left(\rho + p\right) = 0\ , \tag{1.5b}$$

where we have introduced the Hubble parameter, $H \coloneqq {}^{\mathrm{d}}\!/_{\mathrm{d}t}\log a = \frac{\dot{a}}{a}$, and where Λ is the Einstein's Cosmological constant, and G is Newton's gravitational constant.

The perfect fluid that fills the Universe has, from the very-large-scale gravitational point of view, two different components: matter (M), and radiation (R), with equations of state

$$p = \omega\rho \qquad \text{with} \qquad \omega = \begin{cases} \frac{1}{3} & \text{Radiation} \\ 0 & \text{Matter} \\ -1 & \text{Dark Energy} \end{cases}\ . \tag{1.6}$$

Note that we have included another component: Dark Energy, which is the effect of the cosmological constant (Λ) in the dynamics of the Universe when treated as a fluid, i.e. when taken to the RHS in eq. (1.5a); we can define $\rho_\Lambda \coloneqq {}^{\Lambda}\!/_{8\pi G}$, and, since ρ_Λ is constant, the continuity equation imposes $\omega_\Lambda = -1$. From the continuity equation we can derive the evolution of the density of each species i with $a(t)$:

$$\rho_i = \rho_{i,0}\left(\frac{a}{a_0}\right)^{-3(1+\omega_i)}\ , \tag{1.7}$$

where the sub-index 0 refers to the values in a particular instant, usually *today*.

The solution for a flat FLRW Universe dominated by a fluid satisfying the *strong energy condition* $\omega > -1/3$ (i.e. matter or radiation) is

$$a(t) = a_0 t^{\frac{2}{3(1+\omega)}} . \tag{1.8}$$

For Dark Energy, with equation of state given by $\omega = -1$,

$$a(t) = a_0 e^{Ht} . \tag{1.9}$$

It is useful to derive an equation for the acceleration of the scale factor a by differentiating w.r.t. time eq. (1.5a) and inserting (1.5b):[1]

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(1 + 3\omega)\rho . \tag{1.10}$$

This equation makes clear that the expansion of the Universe accelerates when it is dominated by species with $\omega_i < -1/3$, i.e. it decelerates whenever Matter or Radiation dominate, and accelerates whenever Dark Energy dominates. The order in which the different species dominate the expansion can be guessed from eq. (1.7): radiation energy density decays the fastest with time, so at some point matter takes over; finally, since the energy density of Dark Energy is constant, it becomes eventually the dominating species. Therefore, unless something happens which is extraneous to the model as defined so far, the dynamics are dominated in order by

$$\text{Radiation} \quad \rightarrow \quad \text{Matter} \quad \rightarrow \quad \text{Dark Energy} .$$

According to the Supernovæ data [35, 34], the Universe reached a Dark-Energy-dominated stage at $t \simeq 9.8\,\text{Gyr}$, approximately $4\,\text{Gyr}$ ago.

Data coming from different surveys, including CMB, Gravitational Lensing and type Ia Supernovæ redshifts, point clearly towards a very close to flat Universe, $\kappa/(aH)^2 \simeq 0$, populated today by an insignificant amount of radiation, a significant amount matter and an even more significant proportion of Dark Energy [36]. This, together with Standard Model particle physics, is the model of the Universe we will assume for now.

1.1.2 Causal structure of the Universe – Horizons and Hubble radius

Let us now talk about the *causal structure* of the universe: which events could ever be influenced by which other events. We define the *(comoving) particle horizon* of an observer at a time t as the sphere enclosing all points of the fluid that have

[1] Notice that this equation can also be obtained from the spatial component of the Einstein equation.

ever been in causal contact with the observer, i.e. from which a photon emitted at any time may have ever reached the observer. Unsurprisingly, its radius is equal to the (conformal) length that a photon has been able to travel since the origin of the Universe $t = \delta t$ (we admit that the FLRW is not valid from the very beginning of the Universe $t = 0$, if it exists):

$$r_{\mathrm{p}}(t) \coloneqq \int_{\delta t}^{t} \frac{\mathrm{d}\tilde{t}}{a(\tilde{t})} \,. \tag{1.11}$$

If this integral is finite, there exists a particle horizon, meaning that there is a distance such that two points which are separated by more than it, have never been in causal contact by the time t. In a FLRW Universe dominated by radiation, as ours is initially, eq. (1.8) tells us that the integral converges and is dominated by its upper limit. A divergence in the integral (1.11) implies that every pair of points in the fluid have already been in causal contact with each other by the time t. It is obvious that the particle horizon can only grow with time: if a point in the fluid was ever in the horizon, it will always be.

The equivalent concept towards the future is the *(comoving) event horizon* of an observer at a time t: it is the sphere enclosing all points of the fluid at a time t that will *ever* be in causal contact with us before the end of the Universe t_{end}:

$$r_{\mathrm{e}}(t) \coloneqq \int_{t}^{t_{\mathrm{end}}} \frac{\mathrm{d}\tilde{t}}{a(\tilde{t})} \,. \tag{1.12}$$

Again, this integral may diverge, meaning that we will be able to observe any point of the Universe if we are patient enough. This is the case of a decelerating Universe without a final singularity. In the opposite case of a converging integral, e.g. for an ever-faster expanding Universe, as time goes by more and more distant objects whose light ever reached us will never again be observable.

Let us now define the *recessional velocity* v_{rec} and the *radial peculiar velocity* v_{pec} of a galaxy in a point χ of the fluid. The total velocity of an object today is equal to the derivative of its physical distance $d = a\chi$ with respect to time

$$v_{\mathrm{total}} = \underbrace{\dot{a}\chi}_{v_{\mathrm{rec}}} + \underbrace{a\dot{\chi}}_{v_{\mathrm{pec}}} \,. \tag{1.13}$$

With this definition, a point of the fluid has only recessional velocity, and photons have, on top of a recessional velocity, a peculiar velocity of $v_{\mathrm{pec}} = c \equiv 1$. Neither the total or the recessional velocity are properly defined 4-speeds, as the peculiar velocity is (notice how the velocities addition rule of Relativity is not applied here, as it would be for local 4-speeds). Notice how this definition implies the Hubble law automatically:

$$v_{\mathrm{rec}} = \frac{\dot{a}}{a}\, a\chi = Hd \,. \tag{1.14}$$

We define the *comoving Hubble radius* $r_{\rm H}$ as the comoving distance to the points of the fluid that recede from us at the speed of light, $v_{\rm rec} = c \equiv 1$:

$$r_{\rm H}(t) = \frac{1}{aH} \,, \tag{1.15}$$

and the *comoving Hubble sphere* as the surface containing all such points. Recessional velocities higher than the speed of light are possible (see e.g. [15]). Photons emitted towards us from a superluminally receding galaxy actually recede from us, while at the same time any photon overtakes any observer at the speed of light, thus making no conflict with GR. The Hubble radius is *always* contained between the particle and event horizons.

Since photons emitted from outside the Hubble sphere recede from us, one would be tempted to think that the *Hubble radius* is the true event horizon: galaxies outside it (*super-Hubble*), will never be observable. That is definitely not true (see e.g. [16]): the size of the comoving Hubble sphere changes with time, growing in decelerated Universes and shrinking in accelerated ones. A growing Hubble sphere may overtake the receding photons emitted from a super-Hubble galaxy, thus making it observable (see fig. 1.1). Notice that the *physical* Hubble radius $ar_{\rm H} = H^{-1}$ grows for Universes dominated by physical fluids, $\omega > 1/3$, and stays constant for Dark Energy dominated Universes, $\omega = -1$.

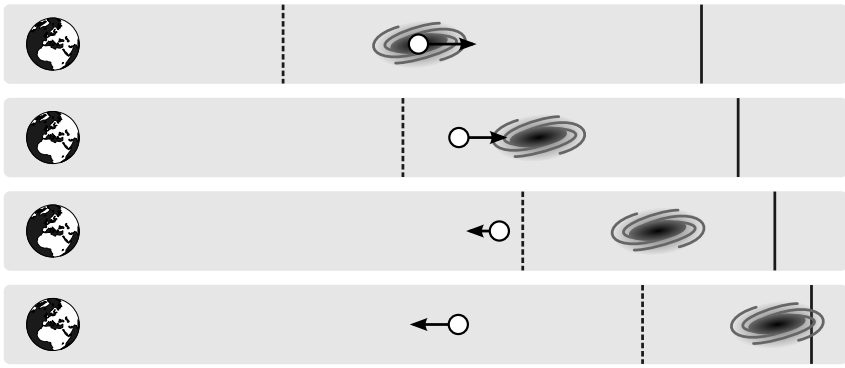


Figure 1.1: Sequence of the path of a photon (white spot) emitted by a super-Hubble galaxy in a decelerating Universe: the supper-Hubble galaxy becomes observable when the initially receding photon is overtaken by the expanding Hubble sphere (dotted line). This only happens when the galaxy is sub-event-horizon (continuous line) at the instant of emission. The arrow represents the total velocity of the photon, as in eq. (1.13). The plot's x direction represents physical distances.

1.2 The *horizon problem* and the need for inflation

1.2.1 The *horizon problem*

The CMB radiation was emitted when the radiation temperature of the Universe was such that the electrons would fall into the nuclei to form atoms, leaving the photons free to travel through the Universe. We call this moment, which has a short but different-from-zero duration, *recombination*. The CMB is astonishingly uniform, reproducing very precisely a black body spectrum with a temperature $T_{\rm CMB} = 2.725\,{\rm K}$. On top of it, once known foreground effects have been accounted for, one finds inhomogeneities of order $10^{-5} T_{\rm CMB}$, which are quite correlated on all scales accessible to the current experiments.

Now let us do a small calculation: under the assumptions stated in section 1.1, using eq. (1.11), we can calculate the size of the particle horizon at *recombination* $t_{\rm rec}$, the moment when the CMB radiation was emitted; then, we can calculate the angle that separates today in the sky two points that were never in causal contact – it turns out to be $\simeq 2^\circ$. This has a very simple consequence: points separated in the sky by bigger angles should show fluctuations of their CMB temperature of order $1 \cdot T_{\rm CMB}$, and those fluctuations should be uncorrelated, i.e. the CMB temperature spectrum should be 0 for $\ell \lesssim 90$ (we discuss the definition of the CMB power spectrum later in section 1.4).

As we just stated, this is certainly not what happens. Either our calculation of the particle horizon at recombination is wrong (i.e. the evolution of $a(t)$ from $t = 0$ to $t_{\rm rec}$ is not as we think it is), or the initial conditions of the Universe were *really* special.

1.2.2 The *flatness conundrum*

Let us recover the curvature in the Friedmann equation (1.5a) (and include the cosmological constant in the density term) and rewrite it as

$$1 - \underbrace{\frac{8\pi G}{3H^2}\,\rho}_{\Omega} = -\frac{\kappa}{(aH)^2}\,. \tag{1.16}$$

Notice how the newly defined, positive quantity Ω controls the curvature of the model:

$$\begin{cases} \Omega < 1 & \text{if} \quad \kappa = -1 \\ \Omega = 1 & \text{if} \quad \kappa = 0 \\ \Omega > 1 & \text{if} \quad \kappa = +1 \end{cases} . \tag{1.17}$$

Let us now show that the value for a flat Universe, $\Omega = 1$, is unstable. For that, let us calculate its derivative w.r.t. a:

$$\frac{\mathrm{d}}{\mathrm{d}a}\,(1-\Omega) = \frac{1+3\omega}{a}\Omega\,(1-\Omega)\,. \tag{1.18}$$

Notice the sign of the derivative:

$$\operatorname{sign}\left(\frac{\mathrm{d}}{\mathrm{d}a}\left(1-\Omega\right)\right) = \operatorname{sign}\left(1+3\omega\right)\operatorname{sign}\left(1-\Omega\right)\ . \tag{1.19}$$

For matter and radiation ($\omega = 0,\ ^1/_3$) the value of $(1-\Omega)$ goes away from 0 as a grows, curving the Universe more and more; instead, for a fluid with negative pressure $\omega < -^1/_3$, the Universe tends towards flatness.

Today [36], our bound for the curvature is $|1-\Omega_0| \simeq 0.05$. Knowing (see sec. 1.1) that the Universe spent most of its history dominated by matter and radiation, we can extrapolate this value towards the early stages of the Universe, e.g. at the epoch corresponding to energy scales of order 10^{16} GeV, where we find $|1-\Omega| \simeq 10^{-55}$, many orders of magnitude smaller. Thus, the Universe should have started its history at stunningly low values of the curvature. This apparent *fine-tuning* of the initial conditions could use a physical explanation.

1.2.3 Digression – Problems and conundrums

We just acknowledged two observations which are surprising when put into the context of the cosmological model as we have established it so far. Are those surprises *physical problems*, i.e. do they *need* to be addressed in order for the theory to be correct or to be complete? In order to answer those questions, let us consider the issue in a more abstract way.

Let's first divide the predictions that models make into two types: deterministic and probabilistic. In turn, probabilistic predictions may come in the form of a well-defined probability distribution function (pdf), or as a lack of predictivity for a certain parameter or outcome, in which case only the main features of a *prior* pdf are known (e.g. equal probability for unpreferred values).

One deterministic prediction would be e.g. how much will a spring elongate when pulled with a certain amount of force. One probabilistic prediction is e.g. the life of a radioactive nucleus, whose decay is a random process. When a deterministic prediction is contradicted by a properly done experiment where observational errors have been accounted for, we have a *physical problem* at hand: our model for the phenomenon is apparently not correct, and we must explain the new observation by modifying it.

When testing a probabilistic prediction the situation is more complicated, since it depends on the amount of samples of the distribution that me have got or may be able to get soon: without enough samples, we cannot characterise the distribution with enough certainty. Regarding sample collection, we distinguish two different situations[2]

[2]A word of caution: the ability to gather samples changes with time: we have continuously achieved previously unimaginable measurements. Thus, questions that were previously philosophical, or simply non-existing, turn suddenly into physical ones. This is often associated with *paradigm shifts* [28].

a) We have collected, or will be able to collect soon, enough samples to test the features of a pdf. In this case, if there is a contradiction with the pdf predicted by the model, or if the model does not predict a particular pdf, we do have a *physical problem*: we have in the first case to modify our model, and in the second one we should extend it to include the new information.

b) We have neither collected, nor will be able to collect soon, enough samples to be able to tell anything about the pdf with minimum reliability. In this case, the model, despite predicting a certain pdf, cannot be tested against the data. Then, no contradiction between the model and the data can be claimed, even if the data looks surprising, a situation which we often call *fine-tuning*. There is no *physical problem* at hand, but a philosophical one; in the context of physics, we can call this situations *conundrums*.

In conclusion, in order to call an apparently surprising result a *physical problem*, we need to be able to assess whether there is actually a contradiction between the prediction and the observation. If we cannot properly perform such assessment, we should call them something else, e.g. *conundrums*, since we cannot resolve the situation using scientific methods. Answering conundrums is satisfactory from the point of view of Physics, and can be argued to be part of the scientific endeavour; but those answers cannot be falsified, so the quest is only worthy when they are addressed by simple principles that at the same time solve real physical problems. Solutions to conundrums are the *icing on the cake*, and sometimes good indications of where our model may be extended, but they may also be *red herrings* that distract us from looking at the real physical problems.

Now back to our apparent problems

Horizon problem *The size of the particle horizon at recombination implies that points in the sky separated by more than* 2° *should have big temperature variations (of order the background temperature* $T_{\rm CMB}$*), which should be uncorrelated.* The prediction is probabilistic – it is not impossible that there are no such variations, just unlikely – and the pdf is not completely characterised, but its central value, $T_{\rm CMB}$, is well determined. In this case, we do have a big number of samples – every pair of regions in the sky separated by $> 2^\circ$, the size of those regions being determined by the resolution of the experiment. They provide a statistically very significant contradiction of the predicted result: the variations on these scales are very small and strongly correlated – here is a *physical problem*: our model is incomplete.

Flatness conundrum *In a Universe dominated most of its history by matter and radiation, as ours has been, the curvature grows with time, so finding a very flat Universe today is quite unexpected, since it would imply an incredibly low amount of curvature in the past.* Here, we have a deterministic prediction: given a value of the curvature in the past, we would know the current one. But we do not know the value in the past, only in the present

day, so we are concerned by a different, probabilistic prediction: if we assume that all values of the curvature were equally likely in the past (or at least approximately so), we expect to find very high curvatures today, flat Universes being highly unlikely. There is only one possible sample here, which is the observable Universe. The fact that our Universe is very improbable may not be appealing, but with only one sample at hand we may as well just have been lucky. There is no physical problem here, but an interesting *conundrum*. We may state that solving this conundrum is possible by modifying the model to increase the probability of a flat Universe today.

Any solution to the horizon problem (which also predicts the correct spectrum for the anisotropies) is equally good, unless further evidence against it is found. But those solutions that in addition address the flatness conundrum without introducing new assumptions deserve special attention.

1.2.4 Inflation as a possible solution

Let us see how both the horizon problem and the flatness conundrum can be addressed by assuming that the Universe during his very early moments underwent a stage of accelerated expansion [1], or *Inflation*.

Consider the horizon problem. Given a long enough period of accelerated expansion, it is intuitively immediate that every point in the Universe is causally connected to every other one, since the points that appear to be today very far away must have been very close in the past. To quantify this, let us solve the particle horizon, eq. (1.11), for a flat Universe dominated by a particular fluid, indicated by i, with equation of state parametrised by ω_i:

$$r_{\mathrm{p}} = \frac{C}{\frac{1}{2}(1+3\omega_i)}\, a^{\frac{1}{2}(1+3\omega_i)}\Big|_{a_{\mathrm{ini}}}^{a_{\mathrm{rec}}} , \tag{1.20}$$

where C is a positive constant, irrelevant here, and a_{ini} and a_{rec} are respectively the scale factor at the beginning of the Universe and at recombination. As eq. (1.10) proves, for fluids with negative pressure such that $\rho + 3p < 0$, or equation of state $\omega_i < -1/3$, the expansion of the Universe is accelerated. In the particle horizon, those values of ω_i make the exponent of the a negative, turning the lower bound into the biggest contribution: for values of the initial scale factor close to zero, the size of the comoving particle horizon tends towards infinity. Our initial intuition gets thus confirmed: an initial phase of accelerated inflation can solve the horizon problem.

Since in our model the Universe starts dominated by radiation, in order to get this phase of accelerated inflation we need to assume that the Universe was initially dominated by a different fluid that later decayed into the Standard Model species that constitute the Universe today, at energies such that they behave as radiation. As we argued, this candidate must behave as a fluid with negative

pressure, $\rho + 3p < 0$ ($\omega < -1/3$). We look for a substance with that feature, the simplest one being a scalar field in a *slow-roll* regime, as we will describe.

Let us assume that at very early times the Universe's dynamics is dominated by a single, homogeneous scalar field $\phi(t, \boldsymbol{x}) = \phi(t)$, which we will call *inflaton*, that is minimally coupled to gravity, with an action

$$S = \int \mathrm{d}^4x \sqrt{-g} \left[\frac{M_{\rm Pl}^2}{2} \mathrm{R} - \frac{1}{2} g^{\mu\nu} \partial_\mu \phi \, \partial_\nu \phi - V(\phi) \right] , \tag{1.21}$$

where $M_{\rm Pl} = \sqrt{\frac{\hbar c}{8\pi G}}$ is the Planck mass, R is the 4-dimensional Ricci scalar and $V(\phi)$ is a potential which is so far arbitrary. Using the Klein-Gordon equation for the scalar field, and the Friedmann equations for its stress-energy tensor, we arrive at the relations (of which only two are independent)

$$\ddot{\phi} + 3H\dot{\phi} = -\frac{\mathrm{d}V}{\mathrm{d}\phi} , \qquad H^2 = \frac{1}{3M_{\rm Pl}} \left(\frac{1}{2}\dot{\phi}^2 + V \right) , \qquad \dot{H} = \frac{1}{M_{\rm Pl}} \left(-\frac{1}{2}\dot{\phi}^2 \right) . \tag{1.22}$$

From this point onwards, for the sake of simplicity, we take $M_{\rm Pl} = 1$. Notice how H^2 is proportional to the sum of the kinetic ($K \sim 1/2\, \dot{\phi}^2$) and the potential energies of the scalar field, and how $-\dot{H}$ equals the kinetic energy. Calculating the stress-energy tensor from the action above, and comparing with that of a perfect fluid, eq. (1.4), we find the density and pressure of the scalar field:

$$\rho_\phi = \frac{1}{2}\dot{\phi}^2 + V , \qquad p_\phi = \frac{1}{2}\dot{\phi}^2 - V . \tag{1.23}$$

Now consider the limit $K \ll V$. In that limit, we immediately get accelerated expansion: $\rho + 3p < 0$, in particular with an equation of state for the scalar field $\omega_\phi \simeq -1$. From the third equation in (1.22), we get an almost constant Hubble parameter, which implies that the scale factor behaves as

$$a \propto e^{Ht} , \qquad H \simeq \text{const} . \tag{1.24}$$

This allows us to define a useful new time scale, e-folds N, as $\mathrm{d}N = H\mathrm{d}t$: during one e-fold, the scale factor a grows by a factor of e.

One can now quantify how long Inflation must last in order to solve the horizon problem, i.e. to get all the scales in the CMB sky into the particle horizon. This turns out to be $50 - 60$ e-folds. With respect to the flatness conundrum, we can see from eqs. (1.18) and (1.19) that flatness becomes an attractor for $\rho + 3p < 0$, so a flat Universe becomes more likely regardless of the original curvature before Inflation. Notice how, unlike the horizon *problem*, the flatness *conundrum* has no quantitative solution; but it is remarkable that it is addressed by the same solution as the horizon problem, which speaks in favour of accelerated expansion against other possible solutions that do not make the Universe flatten with time.

Let us now talk some more about the physical meaning of this framework of Inflation, and incidentally also about its *duration*. Inflation would last as long as the approximation $K \ll V$ holds. That condition describes the dynamics of a field which is *slowly rolling* down a potential. This means that in the first equation in (1.22) the acceleration is very small, and the evolution of the field is dominated by the other terms: the friction term $3H\dot{\phi}$ and the slope of the potential ${}^{\mathrm{d}V}/_{\mathrm{d}\phi}$. In fig. 1.2 we display two possible realisations of a slow-roll potential. We call this whole setting *single-field, slow-roll Inflation*. A slow-roll setting can also be realised in an effectively-single-field theory resulting from integrating out the heavy degrees of freedom (though multiple-field scenarios in which more than one field is light are also possible).

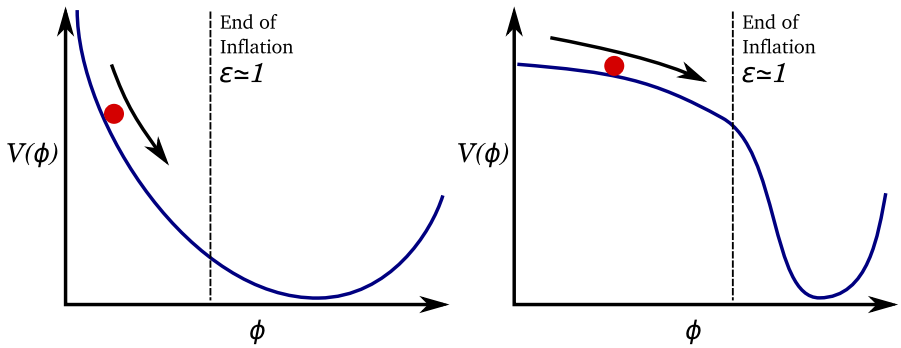


Figure 1.2: Depiction of two simple slow-roll potentials. In the left, the high slope of the potential ${}^{\mathrm{d}V}/_{\mathrm{d}\phi}$ gets compensated by an equally large value of the friction term $3H\dot{\phi}$. In the right, both the slope and the friction are smaller. In both cases, inflation ends whenever one term becomes much bigger than the other, creating a large acceleration.

In order to quantify the slow-roll condition $K \ll V$, we define the positive quantity ϵ, called *first slow-roll parameter*:

$$\epsilon \coloneqq -\frac{\dot{H}}{H^2} = -\frac{\mathrm{d}\log H}{\mathrm{d}N} = \frac{1}{3}\frac{K}{K+V} < 1 \,. \tag{1.25}$$

Its being smaller than 1 guarantees that $\ddot{a} > 0$, hence the Universe's expansion is accelerated. In order for Inflation to last enough time we must also guarantee that the slow-roll condition is satisfied for long enough, i.e. ϵ does not grow too fast, which is given by the *second slow-roll parameter*:

$$|\eta| \coloneqq \left|\frac{\dot{\epsilon}}{H\epsilon}\right| = \left|\frac{\mathrm{d}\log\epsilon}{\mathrm{d}N}\right| < 1 \,. \tag{1.26}$$

We could keep adding higher derivative terms to the description, but this is enough to describe the current data with sufficient precision.

1.3 Primordial perturbations

The former description of the Universe is so far missing a very important ingredient: it is completely homogeneous, lacking any kind of actual structure, such as clusters, galaxies, stars.... We expect small inhomogeneities to be created at some point in the Universe history. In this section, we will see how they get created by Inflation, how to treat them mathematically, and which are their main features.

1.3.1 Primordial perturbations in quantized slow-roll inflation

In order to study perturbations on top of the FLRW metric, we have to write a perturbed version of the metric (1.1), $\tilde{g}_{\mu\nu} \coloneqq g_{\mu\nu} + \delta g_{\mu\nu}$. Its particular shape depends upon some gauge freedom. In the context of Inflation, we normally work in the *comoving gauge*, in which the choice of coordinates is such that the scalar momentum density vanishes, i.e. the 3-surfaces of constant time are everywhere orthogonal to the fluid's flow. In this gauge, the spatial part of the perturbed metric reads

$$\delta g_{ij} = a^2(t)(-2)\zeta(t,\boldsymbol{x})\delta_{ij} + a^2(t)h_{ij}(t,\boldsymbol{x}) \,, \tag{1.27}$$

where $\zeta(t,\boldsymbol{x})$ is a scalar and $h_{ij}(t,\boldsymbol{x})$ is a transverse, traceless tensor. In this gauge, the curvature of the 3-surfaces of constant time is $\mathrm{R} = {}^{4}\!/_{a^2}\nabla^2\zeta$. Hence, we call the scalar ζ *comoving curvature perturbation*.[3] On the other hand, the tensor h_{ij} characterises the *gravitational waves*, which we will not study here.

The Einstein equations of the perturbed metric allow us to calculate the remaining metric perturbations $\delta g_{0\mu}$ also in terms of (ζ, h_{ij}). Substituting them into the action for the scalar field, eq. (1.22), and expanding in powers of ζ, we arrive at the *quadratic* action for the comoving curvature perturbation (terms of higher order are ignored)

$$S_2 = \frac{1}{2}\int \mathrm{d}^4x\, a^3 \left(\frac{\dot{\phi}}{H}\right)^2 \left[\dot{\zeta}^2 - \frac{1}{a^2}\left(\partial_i\zeta\right)^2\right] \,. \tag{1.28}$$

The definition of the comoving gauge, with synchronous 3-surfaces orthogonal to the fluid's flow, implies that the density of the fluid must be constant over those surfaces, in this case $\phi(t,\boldsymbol{x}) = \bar{\phi}(t)$ over those surfaces, since it is the only fluid populating the Universe.

After all remaining gauge freedom is removed using the Einstein equations, we are left with a single scalar degree of freedom, ζ, and two degrees of freedom for each of the two independent tensor perturbations h_{ij} (spin 2, massless). All

[3] Notice that it is more common in the literature (see e.g. [40]) to refer to the comoving curvature perturbation with the symbol $\mathcal{R}$.

five of them are sourced by the only degree of freedom of the theory: the scalar field ϕ.

Now, one can solve the equations of motion for the comoving curvature perturbation arising from the action (1.28). This is often done in terms of the *Mukhanov variable*

$$v \coloneqq z\zeta\,, \qquad z^2 \coloneqq 2a^2\epsilon\,. \tag{1.29}$$

We quantise the action in terms of this variable and derive the equations of motion in the Fourier transformed 3-space, named *Mukhanov-Sasaki equation* [33, 37]:

$$v_{\boldsymbol{k}}'' + \left(k^2 - \frac{z''}{z}\right) v_{\boldsymbol{k}} = 0\,, \tag{1.30}$$

where the primes mean derivatives with respect to conformal time. Conventionally, we impose as initial conditions at the beginning of Inflation, or *choose as a vacuum*, that the mode functions are Minkowski states,

$$\lim_{\tau\to-\infty} v_{\boldsymbol{k}}(\tau) = \frac{1}{\sqrt{2k}} e^{-ik\tau}\,, \tag{1.31}$$

a choice called *Bunch-Davies vacuum* [12]. Solving in the approximation $H \simeq$ const, $\epsilon =$ const (quasi de-Sitter Universe), we arrive at the solution for the modes of the comoving curvature perturbation in terms of the Mukhanov variable:

$$v_{\boldsymbol{k}}(\tau) = \frac{1}{\sqrt{2k}}\, e^{-ik\tau} \left(1 - \frac{i}{k\tau}\right)\,. \tag{1.32}$$

Note that this equation resembles that of an harmonic oscillator with a time dependent mass $m^2(\tau) \coloneqq -z''/z \simeq 2/\tau^2$, where the last approximation is true in the quasi de-Sitter limit. It becomes immediately obvious that the modes of comoving curvature perturbations are oscillatory unless the exponent $k\tau$ vanishes. In the quasi de-Sitter approximation, $H =$ const, $\epsilon =$ const, we have for a conformal time interval $\tau \simeq (aH)^{-1}$. This means that whenever $k \ll aH$, the curvature perturbations are *constant*:

$$\lim_{k/(aH)\to 0} \dot{\zeta}(t, \boldsymbol{k}) = 0\,. \tag{1.33}$$

Remembering the definition of the comoving Hubble radius, eq. (1.15), $r_{\mathrm{H}} = 1/aH$, we learn that this *freeze-out scale* corresponds precisely to the Hubble sphere.[4]

[4]It is common in the literature to refer to this scales as *sub-* or *super-Hubble horizon*, or even worse, simply *sub-* or *super-horizon*. As explained in section 1.1.2, the Hubble radius is not a cosmological horizon, as it is manifested in the context of Inflation by how scales exit and re-enter it; instead, cosmological horizons can only ever be escaped (in the case of the event horizon) or entered (in the case of the particle horizon) a single time, the reverse process not being possible.

Thus, we state that the comoving curvature perturbation is conserved on super-Hubble scales. Since in the de-Sitter Universe the Hubble sphere gets smaller with time, as Inflation progresses modes of decreasing wavelengths get frozen as they exit the Hubble sphere. Once Inflation is finished and we return to a radiation-dominated Universe, the Hubble sphere starts growing, eventually allowing for interactions between regions separated by previously frozen distances. Those interactions give birth to the CMB anisotropies that we can see today, as explained in section 1.4.

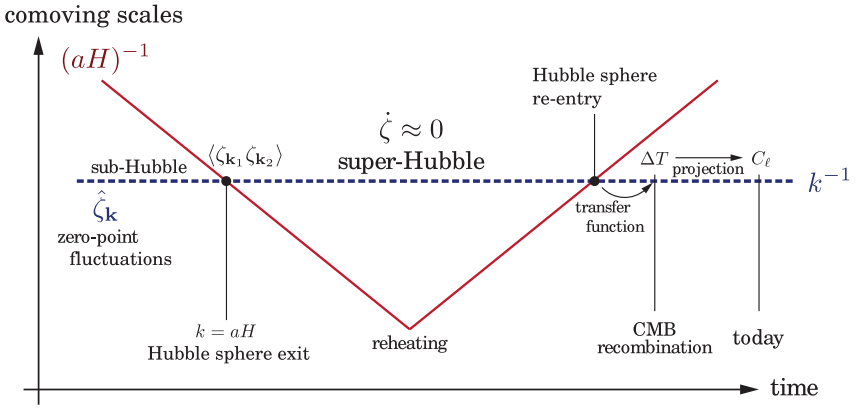


Figure 1.3: Depiction of how a mode of a particular comoving scale (dashed) leaves the Hubble sphere (solid) during Inflation, getting frozen, and how it eventually re-enters during the radiation-dominated regime. Figure from [10].

In the next section, we will review the treatment of the correlation functions of curvature perturbations, which ultimately determine the CMB observables. After it, we will discuss the main features of the perturbations created in this inflationary setting: namely, that they are nearly *scale-invariant*, and nearly *gaussian*.

1.3.2 Correlation functions – spectrum and bispectrum

Here we focus on studying the statistical properties of the perturbations, which we assume to be generated by a random process. In particular, We are interested in correlations between its values in different point of space. In Fourier space, the n-point correlation functions are transformed simply to products of the function itself. We define the 2- and 3-point correlation functions in the Fourier space, called respectively *power spectrum* and *bispectrum*, as

$$(2\pi)^3 \, \delta_{\mathrm{D}}^{(3)} \left(\boldsymbol{k}_1 + \boldsymbol{k}_2\right) \, P_\zeta(\boldsymbol{k}_1) \coloneqq \zeta(\boldsymbol{k}_1)\zeta(\boldsymbol{k}_2) \, , \tag{1.34a}$$

$$(2\pi)^3 \, \delta_{\mathrm{D}}^{(3)} \left(\boldsymbol{k}_1 + \boldsymbol{k}_2 + \boldsymbol{k}_3\right) \, B_\zeta(\boldsymbol{k}_1, \boldsymbol{k}_2) \coloneqq \zeta(\boldsymbol{k}_1)\zeta(\boldsymbol{k}_2)\zeta(\boldsymbol{k}_3) \, . \tag{1.34b}$$

Of course, we can go on and study higher order correlation functions, but those two are enough to describe the current data with enough precision.

We are interested in studying not the particular realisation of the perturbations in our Universe, but the properties of the probability distribution that generated them. These properties show up when averaging over the *ensemble* of all possible *realisations* of the random field ζ which may have been generated by the same physical mechanism. We will denote this ensemble average as $\langle\,\rangle$. Since we expect that mechanism to generate *statistically* isotropic perturbations, the spectrum and bispectrum must depend on norms only, not directions, so the ensemble-averaged correlation functions are

$$(2\pi)^3\,\delta_{\mathrm{D}}^{(3)}\,(\boldsymbol{k}_1+\boldsymbol{k}_2)\;P_\zeta(k_1)=\langle|\zeta(\boldsymbol{k}_1)|^2\rangle\;, \tag{1.35a}$$

$$(2\pi)^3\,\delta_{\mathrm{D}}^{(3)}\,(\boldsymbol{k}_1+\boldsymbol{k}_2+\boldsymbol{k}_3)\;B_\zeta(k_1,\,k_2,\,k_3)\coloneqq\langle\zeta(\boldsymbol{k}_1)\zeta(\boldsymbol{k}_2)\zeta(\boldsymbol{k}_3)\rangle\;. \tag{1.35b}$$

Notice that in the spectrum formula we have made use of the fact that the curvature perturbation is a real function in the space of positions. On the other hand, notice that we have substituted the dependence of the bispectrum from the vectors $(\boldsymbol{k}_1,\,\boldsymbol{k}_2)$, to the norms $(k_1,\,k_2,\,k_3)$. We have done so because, though the first two vectors and the delta completely determine the third vector, once the ensemble average erases the direction information, the norms of the two first vectors do not specify that of the third. There is a range of values of k_3 allowed for each pair $(k_1,\,k_2)$, which are given by the *triangle condition* imposed by the delta:

$$\boldsymbol{k}_1+\boldsymbol{k}_2+\boldsymbol{k}_3=0\;. \tag{1.36}$$

The triangle condition enforces the k_i's to live inside a regular-triangular pyramid in the $(k_1,\,k_2,\,k_3)$ space (see fig. 1.4, and for a review see [31]). Using the *triangular inequality*, for every permutation $(i,\,j,\,k)\in\mathrm{Perm}(1,2,3)$:

$$\boldsymbol{k}_i=-(\boldsymbol{k}_j+\boldsymbol{k}_k)\Rightarrow k_i=|\boldsymbol{k}_j+\boldsymbol{k}_k|\Rightarrow|k_j-k_k|\leq k_i\leq k_j+k_k\;, \tag{1.37}$$

where the equality in the last inequation defines the three sides of the pyramid for $i=1,2,3$.

Since our experiments have a limited precision, we assume that there is a minimum scale accessible to us, i.e, a maximum wave number K such that $k_i\leq K,\;\forall i$, then

$$k_1+k_2+k_3\leq|k_1+k_2|+k_3\leq 2k_3\leq 2K \tag{1.38}$$

and the pyramid turns into a regular tetrahedron with base in the plane

$$k_1+k_2+k_3=2K \tag{1.39}$$

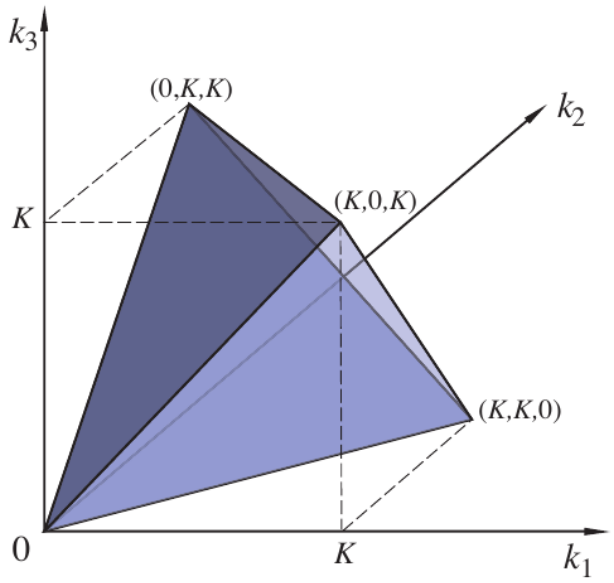


Figure 1.4: Regular Tetrahedron of allowed wave-numbers, limited by the conditions of eq. (1.37) and (1.39). Figure from [31, fig. 2]

Regarding the shape of the ($\boldsymbol{k}_1$, $\boldsymbol{k}_2$, $\boldsymbol{k}_3$) triangles, three different limits are often studied:

squeezed or *local*	$k_3 \ll k_1 \simeq k_2$	*edges* of the tetrahedron
equilateral	$k_3 \simeq k_1 \simeq k_2$	*axis* of the tetrahedron
flattened	$k_3 \simeq k_1 + k_2$	*heights of the faces* of the tetrahedron

1.3.3 Predictions of slow-roll inflation

Power spectrum – scale-invariance

In the QFT framework described in section 1.3.1, one can calculate the predicted power spectrum for the comoving curvature perturbation [12]

$$P_\zeta(k) = \frac{1}{4k^3} \left.\frac{H^2}{\epsilon}\right|_{k=aH} , \tag{1.40}$$

where the spectrum for each comoving wavelength k can be evaluated when said length becomes larger than the (shrinking) Hubble sphere, thus getting *frozen*. We can define a *dimensionless* version of the power spectrum as

$$\mathcal{P}(k) \coloneqq \frac{k^3}{2\pi^2} P_\zeta(k) = \frac{1}{8\pi^2} \left.\frac{H^2}{\epsilon}\right|_{k=aH} . \tag{1.41}$$

What are the particular features of this spectrum? From the slow-roll conditions we know that during Inflation H decreases slowly as ϵ grows slowly (i.e. the kinetic energy increases), so for the smaller wavelengths (higher k) we expect $\mathcal{P}(k)$ to *slowly* decrease, according to (1.41). The spectrum will thus be nearly *scale-invariant*, with a little less power for higher k. Notice that scale-invariance is a natural consequence of the lack of large-scale standard rulers in the early Universe.

When dealing with constraining the power spectrum with observations, we often prefer a more phenomenological parametrisation that reveals more directly its dependence on the scale. We start by assuming a general power law spectrum, $\mathcal{P}(k) \sim k^n$, and then Taylor-expand it in its logarithm around a pivot scale k_0:

$$\log \frac{\mathcal{P}(k)}{\mathcal{P}(k_0)} \simeq 0 + (n_\mathrm{s} - 1) \log \frac{k}{k_0} + \frac{\alpha_\mathrm{s}}{2} \left(\log \frac{k}{k_0} \right)^2 + \dots \,, \tag{1.42}$$

where we have defined the *spectral index*[5] n_s and the *running* α_s, and ignored higher order terms. For a nearly scale invariant spectrum, we should get very small values of $n_\mathrm{s} - 1$ and an even smaller for α_s, which can often be neglected. Calling A_s the value of the spectrum at the pivot scale, we arrive at the very well known expression

$$\mathcal{P}(k) \simeq A_\mathrm{s} \left(\frac{k}{k_0} \right)^{n_\mathrm{s}-1} \,. \tag{1.43}$$

The value of the spectral index can be calculated from eq. (1.40):

$$n_\mathrm{s} - 1 = -2\epsilon - \eta \,, \tag{1.44}$$

where we omit contributions from higher powers of ϵ and η. We don't know the precise values for ϵ and η, but we know that they result in a value for n_s close to 1 but slightly smaller. This small deviation from unity constitutes a strong test for the Inflation framework described here: slow-roll, single-field inflation with canonical kinetic terms and Bunch-Davies vacuum. This small deviation was confirmed by WMAP and later by Planck. The latter finds a 95%c.l. interval [8]

$$n_\mathrm{s} = 0.9616 \pm 0.0094 \,. \tag{1.45}$$

Bispectrum – gaussianity

One of the defining features of a Gaussian distribution is that it is completely characterised by its mean and its variance, strictly its 1st order momentum and 2nd order central momentum. Higher order momenta, are completely determined by those two. In particular, all the odd momenta are zero.

[5] We would normally call *spectral index* the n in $\mathcal{P}(k) \sim k^n$, but in cosmology it is customary to give that name to the difference $n_\mathrm{s} - 1$.

The fluctuations of the inflaton have a mean value of zero, and their higher order momenta are given by its n-point correlation functions (we just described the predicted 2-point correlation). If inflationary primordial fluctuations were gaussian, they would be completely determined by a constant power spectrum, and have an vanishing bispectrum.

The contributions to the bispectrum would arise mainly from the order-3 terms in the action of the comoving curvature perturbation. Those terms account for interactions between the fields present during inflation. Since the only field in the theory is minimally coupled to gravity, we expect to have only a very small contribution to the bispectrum. The full bispectrum for single-field, slow-roll inflation was first calculated in [32, 4] and it is

$$\begin{aligned} B_\zeta(k_1,\, k_2,\, k_3) = \frac{(2\pi)^4 \mathcal{P}(\sum_i k_i)^2}{\left(k_1 k_2 k_3\right)^2} \frac{1}{8 k_1 k_2 k_3} \times \\ \times \left[(3\epsilon - 2\eta) \sum_{i=1}^{3} k_i^3 + \epsilon \sum_{i \neq j} k_i k_j^2 + \epsilon \frac{4}{\sum_{i=1}^{3} k_i} \sum_{i \neq j} k_i^2 k_j^2 \right] . \end{aligned} \tag{1.46}$$

Now, in the same way we did with the power spectrum, we often define a dimensionless version of the bispectrum by factoring out $(k_1 k_2 k_3)^{-2}$, and, in order to compare its value with that of the power spectrum, also a factor of $\mathcal{P}(\sum_i k_i)^2$, defining the *shape function* $\mathcal{S}(k_1,\, k_2,\, k_3)$:

$$\mathcal{S}(k_1,\, k_2,\, k_3) \coloneqq \frac{\left(k_1 k_2 k_3\right)^2}{(2\pi)^4 \mathcal{P}(\sum k)^2} B_\zeta(k_1,\, k_2,\, k_3) \, . \tag{1.47}$$

(Unfortunately, there is no unique convention in the literature for the prefactor)

Factoring out this scaling, we can now see how the bispectrum is suppressed by $\mathcal{O}(\epsilon,\, \eta)$ with respect to the power spectrum. We do not know the real value of the slow-roll parameters, but the bounds given by eqs. (1.44) and (1.45) make us expect a very small bispectrum, which, together with a nearly scale-invariant power spectrum are a strong indication for slow-roll single-field inflation. Both WMAP and Planck confirm the smallness of the bispectrum when fitting the shape in eq. (1.46).

Notice that in the last sentence we explicitly stated which shape of the bispectrum was fitted when quoting the amount of non-gaussianity. We must always do so, and the reason is that the many possible configurations of momenta in which the bispectrum can be evaluated (see fig. 1.4) leave room for very diverse functional forms with very different dependencies on the configurations. Given a data set containing a significant amount on non-gaussianity in a particular shape, if we try to fit a shape much different from that of the signal we would get a small amplitude for it. Being aware of this, we should be careful about making general statements of the amount of non-gaussianity in e.g. the CMB data, since this *amount*, in principle, depends on its shape. We are about to describe

a model that predicts a distinct shape for the primordial bispectrum, and so do other inflationary settings. A precise measurement of the cosmological observables which are most informative about the primordial bispectrum (specially the CMB bispectrum) can be determinant to distinguish among different inflationary scenarios.

1.3.4 Beyond the simplest model – Transient reductions in the speed of sound

Inflation occurs at high energy scales in which our knowledge of the field content, prior to cosmological constraints, is almost non-existent. The presence of multiple fields during Inflation would produce a very rich phenomenology that could also be constrained in principle by cosmological observations (different, also rich phenomenology can arise too from other alternatives to the simplest inflationary scenario, such as brief interruptions of the slow-roll regime or different choices of the vacuum).

As we argued in the last section, the predictions of slow-roll, single-field inflation are well within the constraints of CMB observations, their most direct probe. This fact motivates us to approach the presence of multiple fields during inflation under the hypothesis that they are significantly heavy. In this regime, we expect to be able to integrate out those heavy degrees of freedom and obtain an effectively-single-field theory in which the effect of the background enters through effective operators in the action, and the inflaton direction fulfils a slow-roll regime. Small excitations of the heavy degrees of freedom during inflation might produce, through the effective operators, potentially detectable deviations from the predictions of a purely single-field framework.

To achieve that, several different approaches are possible. In this section we will not discuss the details of the procedure, but instead we start from the effective action, after integrating out the heavy degrees of freedom, of the field $\pi(t, \boldsymbol{x})$, which represents displacements along the background trajectory in the effective potential. The quadratic and cubic action for this field are [13]

$$S_2 = \int \mathrm{d}^4x\, a^3 \epsilon H^2 \left(\frac{\dot{\pi}^2}{c_s^2} - \frac{1}{a^2} \left(\nabla\pi\right)^2 \right) , \tag{1.48a}$$

$$S_3 = \int \mathrm{d}^4x\, a^3 \epsilon H^2 \times \\ \times \left\{ -2Hsc_s^{-2}\pi\dot{\pi}^2 - \left(1 - c_s^{-2}\right) \pi \left[\dot{\pi}^2 - \frac{1}{a^2} \left(\nabla\pi\right)^2 \right] \right\} , \tag{1.48b}$$

where we are neglecting higher order slow-roll corrections. Notice the introduction of the speed of sound c_s, and the derived quantities u and s, defined as:

$$u \coloneqq 1 - c_s^{-2} \quad , \quad s \coloneqq \frac{\dot{c}_s}{c_s H} . \tag{1.49}$$

The action above neglects terms of higher order in u and s. The speed of sound in the effective field theory can be related to the angular velocity $\dot{\theta}$ of the field π whenever it finds a *turn* in the inflationary trajectory:

$$c_s^{-2} = 1 + \frac{4\dot{\theta}^2}{\frac{k^2}{a^2} + M_{\text{eff}}^2} \,, \tag{1.50}$$

where M_{eff}^2 is the effective mass of the mode perpendicular to the trajectory in field space. Therefore, a turn between two straight segments of the trajectory produces a momentary reduction in the speed of sound. As the action above is perturbative in terms of u and s, the reduction in the speed of sound cannot be too big nor too quick; at the same time, we keep at all times a slow-roll regime, and expect the contribution of the slow-roll corrections to be much smaller than that of the speed of sound, which must be dominating the evolution during the turns. Altogether, the theoretical framework presented here is well defined in the regime of *mild and moderately sharp transient reductions*:

$$\mathcal{O}(\epsilon, \eta) \ll \max(u, s) \ll 1 \,. \tag{1.51}$$

Now we will take a look at the possible observable consequences of this regime. To do that, we need to relate the inflaton direction π with the comoving curvature perturbation ζ:

$$\zeta = -H\pi \,. \tag{1.52}$$

In terms of the adiabatic curvature perturbation, the quadratic action of the inflationary mode is

$$S_2 = \underbrace{\int \mathrm{d}^4x\, a^3\epsilon \left[\dot{\zeta}^2 - \frac{(\nabla\zeta)^2}{a^2}\right]}_{S_{2,0}} + \underbrace{\int \mathrm{d}^4x\, a^3\epsilon(-u)\dot{\zeta}^2}_{\delta S_2} \,. \tag{1.53}$$

Notice that the first part of the action, $S_{2,0}$, corresponding to the case $c_s = 1$, is similar to the canonical slow-roll action eq. (1.28), and has the same phenomenology, producing the well known power spectrum of eq. (1.41). The new, interesting term δS_2 accounts for the reductions in the speed of sound, and produces a perturbation of the (nearly) scale-invariant power spectrum, which can be calculated using the *in-in formalism* [27, 41] for the case of a small, transient reduction in c_s, to first order in $u \equiv 1 - c_s^{-2}$ [3]:

$$\frac{\Delta\mathcal{P}}{\mathcal{P}}(k) = k \int_{-\infty}^{0} \mathrm{d}\tau\; u(\tau) \sin(2k\tau) \,, \tag{1.54}$$

where $k \equiv |\boldsymbol{k}|$, $\mathcal{P}$ is the featureless power spectrum with $c_s = 1$ defined by eq. (1.41), and τ is the conformal time. This very important expression shows

precisely how changes in the speed of sound, within the regime described here, seed features in the primordial power spectrum of curvature perturbations. Those features are given by a simple Fourier transform in conformal time, assuming $c_s = 1$ at the beginning ($\tau = -\infty$) and the end of inflation ($\tau = 0$).

We can also calculate the resulting bispectrum from the cubic action in terms of the adiabatic curvature perturbation, using the in-in formalism, at first order on u and s, and disregarding slow-roll contributions [3]:

$$\begin{aligned}\Delta B_\zeta(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3) = \\ \frac{(2\pi)^4 \mathcal{P}^2}{(k_1 k_2 k_3)^2} \Bigg\{ &- \frac{3}{2} \frac{k_1 k_2}{k_3} \left[\frac{1}{2k} \left(1 + \frac{k_3}{2k} \right) \frac{\Delta\mathcal{P}}{\mathcal{P}} - \frac{k_3}{4k^2} \frac{\mathrm{d}}{\mathrm{d}\ln k} \frac{\Delta\mathcal{P}}{\mathcal{P}} \right] + \dots \\ &+ \frac{1}{4} \frac{k_1^2 + k_2^2 + k_3^2}{k_1 k_2 k_3} \left[\frac{1}{2k} \left(4k^2 - (k_1 k_2 + \dots) - \frac{k_1 k_2 k_3}{2k} \right) \frac{\Delta\mathcal{P}}{\mathcal{P}} \right. \\ &\left. - \frac{k_1 k_2 + \dots}{2k} \frac{\mathrm{d}}{\mathrm{d}\ln k} \frac{\Delta\mathcal{P}}{\mathcal{P}} + \frac{k_1 k_2 k_3}{4k^2} \left(\frac{\mathrm{d}}{\mathrm{d}\ln k} \right)^2 \frac{\Delta\mathcal{P}}{\mathcal{P}} \right] \Bigg\} , \end{aligned} \tag{1.55}$$

where $k_i \coloneqq |\boldsymbol{k}_i|$, $k \coloneqq {}^1\!/_2 \sum_{i=1}^2 k_i$, and $\mathcal{P}$, $\Delta\mathcal{P}/\mathcal{P}$ and its derivatives are evaluated at k. The ellipsis $\dots$ means a sum over all possible permutations of $(1, 2, 3)$ on the indices of the term previous to it. Note that in the squeezed limit $k_i \ll k_j \simeq k_k$ we recover the single-field consistency condition of [32, 14].

It becomes immediately clear how a reduction in the speed of sound seeds correlated perturbations in both the power spectrum and bispectrum. In the first part of this thesis, we study how this correlation can be exploited to try to detect reductions in the speed of sound in CMB (and LSS) data.

As a final remark, note that, though we have related the reduction in the speed of sound to the angular velocity along a turn, the result above is independent of the physical origin of such reduction. Different high energy realisations of the reduction would produce different effective operators in the cubic action, making the realisations distinguishable at the level of the bispectrum [13, 2].

1.4 The Cosmic Microwave Background

1.4.1 Statistical properties – spectrum and bispectrum

We expect the temperature anisotropies in the CMB to be the most informative probe of the primordial perturbations generated by inflation. As we will see later, the statistical properties of the primordial n-point correlation functions are related to the corresponding correlations of temperature anisotropies in the CMB sky. We will leave the description of the physical effects leading from primordial to CMB anisotropies for the next section, and we start here by describing the treatment of correlation functions of the temperature anisotropies in the CMB sky.

Since the anisotropies are projected on the 2-sphere defined by all possible directions in the sky given by the unit vector $\hat{\boldsymbol{n}}$, we use as the basis of the transformation the spherical harmonics:

$$\frac{\Delta T}{T}(\hat{\boldsymbol{n}}) = \sum_{\ell m} a_{\ell m} Y_{\ell m}(\hat{\boldsymbol{n}}) \tag{1.56}$$

Notice how the harmonic-transformed perturbation, instead of a continuous function like the primordial perturbation $\zeta(\boldsymbol{k})$ was, is now a discrete series of coefficients $a_{\ell m}$, due to the finitude of the space over which the transformation is taken: a 2-sphere. The integer sub-index $\ell \geq 0$ characterises the angular periodicity of $Y_{\ell m}$, its *period* being $(2\pi/\ell)$ rad or $360^\circ/\ell$. For each ℓ, the values of the integer index $m \in [-\ell, \ell]$ represent the different relative directions in which the angular variation of $Y_{\ell m}$ can be realised.

As we did with primordial perturbations, we define the *CMB power spectrum* C_ℓ as the 2-point correlation function averaged over all possible realisations of the random fluctuations:

$$\langle a_{\ell_1 m_2} a^*_{\ell_2 m_2} \rangle \;=\; \delta_{\ell_1 \ell_2} \delta_{m_1 m_2} C_{\ell_1} \; . \tag{1.57}$$

In an analogous way to the primordial spectrum's lack of information of direction, $\mathcal{P}(\boldsymbol{k}) \equiv \mathcal{P}(k)$, the CMB power spectrum carries no information on direction either, $C_{\ell m} \equiv C_\ell$, imposed by the delta $\delta_{m_1 m_2}$.

Equivalently, one could have calculated directly the correlation between two temperature anisotropies, and averaged them over all possible realisations of the CMB sky, getting

$$\left\langle \frac{\Delta T}{T}(\hat{\boldsymbol{n}}_1) \frac{\Delta T}{T}(\hat{\boldsymbol{n}}_2) \right\rangle = \sum_\ell \frac{2\ell + 1}{4\pi} C_\ell P_\ell(\cos\theta) \; , \tag{1.58}$$

where P_ℓ are the Legendre polynomials and θ is the angle between the direction $\hat{\boldsymbol{n}}_1$ and $\hat{\boldsymbol{n}}_2$.

We define the 3-point correlation function, or *reduced CMB bispectrum* as

$$\langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \rangle \coloneqq \mathcal{G}^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3} \, b_{\ell_1 \ell_2 \ell_3} \; , \tag{1.59}$$

where we make use of the so-called *Gaunt integral*:

$$\mathcal{G}^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3} \;=\; \sqrt{\frac{(2\ell_1 + 1)(2\ell_2 + 1)(2\ell_3 + 1)}{4\pi}} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix} , \tag{1.60}$$

where the matrices represent Wigner 3-j symbols (we will show the origin of the Gaunt integral in the next section). We can also write the inverse equation to (1.59):

$$b_{\ell_1 \ell_2 \ell_3} \;=\; \sum_{m_i} \mathcal{G}^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3} \, \langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \rangle \; . \tag{1.61}$$

Notice now that the reduced CMB bispectrum has no directional information (m_i), as expected, since we assumed isotropy. The only directional information, which is geometrical, not physical, is inside the Gaunt integral, which takes the rôle of the deltas in (1.57). Those deltas would impose some constraints on $\langle a_{\ell_1 m_2} a^*_{\ell_2 m_2} \rangle$ which were not visibly present in C_ℓ. In the same way, the fact that the Gaunt integral contains those two 3-j symbols imposes some *selection rules* in $\langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \rangle$ that cannot be seen directly on $b_{\ell_1 \ell_2 \ell_3}$. In particular, the ℓ_i's must form a triangle of even perimeter, i.e.[6]

$$\frac{\ell_1 + \ell_2 + \ell_3}{2} \in \mathbb{Z} \,, \tag{1.62a}$$

$$|\ell_j - \ell_k| \leq \ell_i \leq \ell_j + \ell_k \,, \qquad \forall (i,j,k) \in \mathrm{Perm}(1,2,3) \,. \tag{1.62b}$$

The second condition is analogous to the one imposed over the wave numbers in the primordial bispectrum, (1.36): both wave numbers and harmonic indices must form triangles in order to contribute to each bispectrum. In particular, in the harmonic space, the triangle must have even perimeter.

The triangle condition determines a tetrahedral cone for the ℓ_i's, as it did for the k_i's. In the case of the wave numbers, the cone was cut into a pyramid by the existence of a maximum observable wave number K. Here, correspondingly, we assume a maximum precision L, such that $\ell_1, \ell_2, \ell_3 \leq L$. But here the lack of a vectorial condition like $\sum_i \boldsymbol{k}_i = 0$ lets the cone be extended up to the edges of the cube limited by L, forming not a pyramid but a *bi-pyramid* (see [19, fig. 2], where it is called *tetrapyd*) with apices in $(0,0,0)$ and (L,L,L), limited by the planes

$$\ell_i = \ell_j + \ell_k \,, \qquad \ell_i = L \,, \qquad \forall (i,j,k) \in \mathrm{Perm}(1,2,3) \,. \tag{1.63}$$

Notice that the observed quantity in the sky is the *angle-averaged CMB bispectrum*, defined as

$$B_{\ell_1 \ell_2 \ell_3} \coloneqq \sum_{m_i} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \rangle \,, \tag{1.64}$$

or, inversely

$$\langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \rangle = \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix} B_{\ell_1 \ell_2 \ell_3} \,, \tag{1.65}$$

and which fulfils the following relation with the reduced bispectrum

$$B_{\ell_1 \ell_2 \ell_3} = \sqrt{\frac{(2\ell_1+1)(2\ell_2+1)(2\ell_3+1)}{4\pi}} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ 0 & 0 & 0 \end{pmatrix} b_{\ell_1 \ell_2 \ell_3} \,. \tag{1.66}$$

[6] Notice that the second condition, provided that all ℓ_i's are positive numbers, needs only be checked for a particular permutation of $(1\,,2\,,3)$, which implies that it is satisfied for all of them.

1.4.2 From inflationary primordial fluctuations to CMB anisotropies

In section 1.3.1 we described the curvature perturbations created by slow-roll, single-field inflation, and proved that their modes get frozen when their respective wave length exists the Hubble sphere.[7] The very fact that they remain frozen while they are outside the Hubble sphere is what allows us to learns anything about Inflation, since we know close to nothing about the physics going on right after the end of inflation.

After the end of inflation, the Universe becomes filled with radiation (part of which decouples as matter as some point), and the comoving curvature perturbations created by Inflation source perturbations in the density of the radiation field. In slow-roll, single field inflation the only degree of freedom present is the scalar field, which sources curvature perturbations ζ in the syncronous 3-surfaces. As the only degree of freedom driving the dynamics of the Universe, at each point of the fluid, the curvature perturbation can be used as a clock $t_{\boldsymbol{x}} = t_{\boldsymbol{x}}(\zeta)$. Then, the small spatial variations can be understood as inflation being slightly ahead or behind at each point by a small quantity $\delta t(\boldsymbol{x})$. This small, continuous time shift causes the different regions to have expanded a slightly different amount – the surfaces of constant time are not homogeneous any more, or, equivalently, the surfaces of homogeneity are not flat anymore, which is another way to understand why *curvature perturbations* in the 3-space are generated.

Let us look at the density perturbation of the fluid i at some point after inflation, in a surface of constant time t. This 3-surface is nearly homogeneous, with an energy density $\bar{\rho}_i(t)$. The real, perturbed density of the fluid i, for a small time shift, can be written as

$$\rho_i(t, \boldsymbol{x}) = \bar{\rho}_i \left(t + \delta t(\boldsymbol{x})\right) \simeq \bar{\rho}_i(t) + \dot{\bar{\rho}}_i(t)\delta t(\boldsymbol{x}) \,, \tag{1.67}$$

and a similar expression for the pressure. Defining $\delta\rho_i(t, \boldsymbol{x}) \coloneqq \dot{\bar{\rho}}_i(t)\delta t(\boldsymbol{x})$ and substituting the time derivative of the density with the continuity equation (1.5b), one arrives at the equality

$$\frac{\delta\rho_i}{\rho_i + p_i} = -3H\delta t(x) \,, \tag{1.68}$$

where one should notice that the right term is independent of the species. This independence of the particular fluid is a consequence of there being a *single* degree of freedom $\delta t(\boldsymbol{x})$ sourcing the perturbations, and defines *adiabatic perturbations*.

Before continuing, a caveat is in order: in general, the linear perturbations of the different fluids, photons, baryons or CDM, are possibly coupled with each other in different ways, depending on the mechanism that generated them. They

[7] Formally, we have only proven that perturbations are conserved during inflation, but still would have to prove that they keep frozen after inflation has ended. The proof for adiabatic perturbations can be found in [40], and in an alternative formulation in [39].

can be expanded by two orthogonal components: *entropy perturbation*, which imply local variations of the relative number density, but keeping the total energy density constant (and thus the curvature of the local space-time, hence also called *isocurvature perturbations*); and *adiabatic perturbations*, in which there is no energy exchange between the different fluids, and are defined by equation (1.68). Slow-roll single-field inflation produces adiabatic perturbations, while in multiple-field inflation we expect several degrees of freedom seeding perturbations, thus creating a combination of adiabatic and isocurvature perturbations. Only because we assumed slow-roll single-field inflation, we are able to choose such a synchronous gauge that traces the perturbations in all fluids simultaneously, showing explicitly the defining property of adiabatic perturbations, eq. (1.68).

Current CMB data clearly favours *adiabatic perturbations* as the only necessary ones to describe the universe today, while restricting the size of isocurvature perturbations to be negligible [6].

The next step to relate the primordial curvature perturbations with the CMB temperature perturbations is to study the physical processes that the perturbation modes undergo after entering the Hubble sphere. The nature of those processes is well known, and can be summarised into a series of *transfer functions* $\Delta_\ell(k)$, which project the primordial curvature perturbation to the harmonic coefficients of the CMB sky as

$$a_{\ell m} \;=\; 4\pi(-i)^\ell \int \frac{d^3\boldsymbol{k}}{(2\pi)^3}\Delta_\ell(k)\zeta(\boldsymbol{k})Y_{\ell m}(\hat{\boldsymbol{k}}) \; . \tag{1.69}$$

The effect of the transfer functions in the power spectrum can be divided in two sets of contributions: those acting at scales which are super-Hubble at the time of recombination, and those that had entered the Hubble sphere earlier, divided approximately by the scale of the multipole $\ell \approx 90$, as discussed in section 1.2.1.

Of the first ones, the main contribution comes from the *Sachs-Wolfe effect*, which is the gravitational redshift that the photons suffer at the last scattering surface due to the differences in the local gravitational potential. A smaller contribution comes from the *Integrated Sachs-Wolfe effect*, which consists on the gravitational redshift that the photons undergo during their journey to us, due to the difference in energy between descending and climbing a potential in an expanding, radiation- or dark matter-dominated Universe. The relevant part of the power spectrum, $l \lesssim 90$, results in a direct processing of the primordial one, keeping flat bar a factor $\ell(\ell+1)$ coming from the projection of the flat waves in $\boldsymbol{k}$ over the spherical sky.

At sub-Hubble scales, the cosmic fluid has been able to undergo local physical processes between leaving the Hubble-sphere and the last scattering surface. Therefore, on top of the Sachs-Wolfe spectrum, we will see the effect of these processes. The cold dark matter (CDM) fluid is decoupled from an initially coupled fluid of baryons and photons. CDM dominates the gravitational potential, and

the baryons+photons fluid oscillates inside the potential wells due to the photon radiation pressure. These stationary oscillations produce, at the time of decoupling of photons and baryons in the last scattering surface, the series of *acoustic peaks* that we can see on the CMB power spectrum. These peaks of approximately equal height in the spectrum are actually damped at small scales by the fact that the decoupling did not occur instantly, and the photons underwent scattering at scales of their mean free path at that time. This last effect is called *Silk damping*.

The effect of those physical processes in the CMB power spectrum can be seen in figure 1.5(a) of section 1.4.3. For a detailed study of them we refer the reader to thorough reviews such as [17]. The transfer functions are often computed with the help of cosmological codes, such as CLASS or CAMB. Here it is enough to keep track of the parameters of the cosmological model implied in them: (1) the ones referring to the ΛCDM background, namely the value of the expansion rate today (H_0) and the densities of baryons and CDM ($\Omega_{\rm b}$ and $\Omega_{\rm CDM}$); (2) those referring to late-time effects, the optical depth due to reionisation $\tau_{\rm reio}$ being the only one that we can constrain (the rest are treated as *nuisance parameters*, see later); and (3) parameters describing experimental effects, e.g. effect of calibration, of the particular survey, treated as nuisance parameters.

The equation (1.69) can be applied to the definitions of the CMB spectrum and bispectrum, in order to relate them to their primordial counterparts eq. (1.35a) and (1.35b). After some algebra (involving integrating angles between vectors $\boldsymbol{k}_i$'s, using properties of spherical harmonics, expanding the Dirac delta in Legendre polynomials...) we arrive at the expressions:

$$\langle a_{\ell_1 m_2} a^*_{\ell_2 m_2} \rangle \;=\; \delta_{\ell_1 \ell_2} \delta_{m_1 m_2} \frac{2}{\pi} \int_0^\infty dk\, k^2 \Delta^2_\ell(k) P_\zeta(k) \;, \tag{1.70}$$

$$\begin{aligned} \langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \rangle = \mathcal{G}^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3} \left(\frac{2}{\pi}\right)^3 \iiint \mathrm{d}k_1 \mathrm{d}k_2 \mathrm{d}k_3 \, (k_1 k_2 k_3)^2 \\ \Delta_{\ell_1}(k_1) \Delta_{\ell_2}(k_2) \Delta_{\ell_3}(k_3) \, B_\zeta(k_1, k_2, k_3) \, J_{\ell_1 \ell_2 \ell_3}(k_1, k_2, k_3) \;, \end{aligned} \tag{1.71}$$

where

$$J_{\ell_1 \ell_2 \ell_3}(k_1, k_2, k_3) \;\coloneqq\; \int x^2 dx \, j_{\ell_1}(k_1 x) j_{\ell_2}(k_2 x) j_{\ell_3}(k_3 x) \;, \tag{1.72}$$

and the already mentioned *Gaunt integral* is

$$\mathcal{G}^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3} \;\coloneqq\; \int d\Omega_{\boldsymbol{x}} Y_{\ell_1 m_1}(\hat{\boldsymbol{x}}) Y_{\ell_2 m_2}(\hat{\boldsymbol{x}}) Y_{\ell_3 m_3}(\hat{\boldsymbol{x}}) \;, \tag{1.73}$$

whose solution is eq. (1.60) and whose properties we have already discussed.

1.4.3 Experimental status

The CMB temperature power spectrum measured by the Planck satellite and released in March 2013 [5] can be seen in figure 1.5(a), with the best-fit ΛCDM model on top (solid line). As we can see, the concordance between the measured power spectrum and the ΛCDM prediction is stunning.

The CMB bispectrum found by Planck can be seen in figure 1.5(b). Its small size appears to fulfil the prediction for slow-roll inflation (see sec. 1.3.3), though there are hints of an oscillatory pattern with possible primordial origin [7, sec. 7.3.3].

1.5 Constraining inflation with the CMB – parameter extraction

Now that we have proven how the properties of the perturbations generated by inflation are imprinted in the CMB, we turn to the topic of how to extract that information from the measured CMB data of a real survey. The result is given as regions of the parameter space containing the most likely values of the cosmological parameters that generated the data at hand. To achieve that result, we will need a number of statistical tools, which we discuss in this section.

1.5.1 Bayes' theorem

Let $\mathcal{M}$ be our model for the Universe, which depends on some parameters θ defined over a parameter space Θ. Let $\mathcal{D}$ be a set of data that may be well described by the model $\mathcal{M}$. We are interested in the probability distribution of the parameters of the model $\mathcal{P}(\theta \,|\, \mathcal{D}, \mathcal{M})$, conditional on having observed the data $\mathcal{D}$, which we assume has been produced by the model $\mathcal{M}$. We call it *posterior pdf*[8]. It is often extracted from *Bayes' theorem*, which states

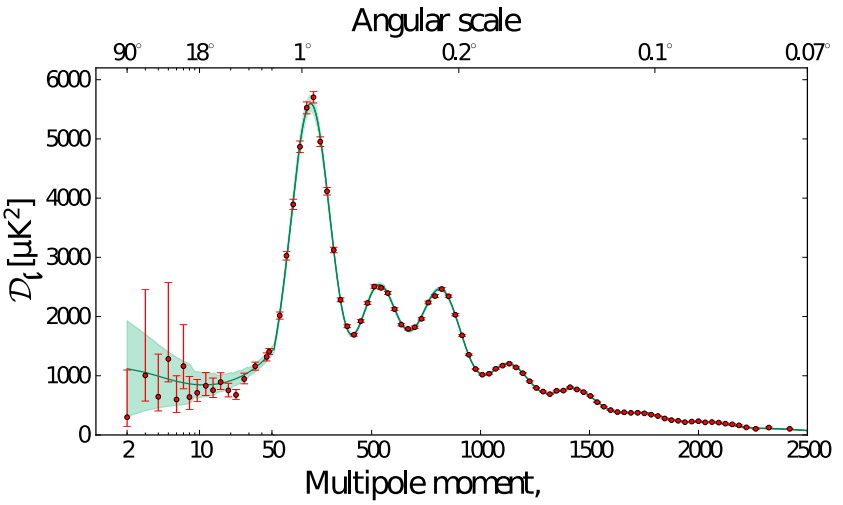
$$\mathcal{P}(\theta \,|\, \mathcal{D}, \mathcal{M}) = \frac{\mathcal{L}(\mathcal{D} \,|\, \mathcal{M}(\theta))}{\mathcal{Z}(\mathcal{D} \,|\, \mathcal{M})} \pi(\theta \,|\, \mathcal{M}) \ . \tag{1.74}$$

Here, we have introduced three more probability distributions:

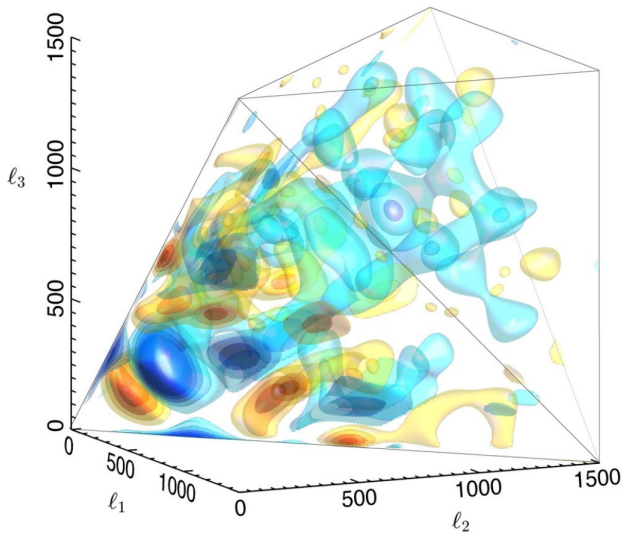
$\pi(\theta \,|\, \mathcal{M})$**,** , called **prior**, is the pdf of the values of the parameters θ within the model $\mathcal{M}$, which is assigned before knowing about the data $\mathcal{D}$.

$\mathcal{L}(\mathcal{D} \,|\, \mathcal{M}(\theta))$**,** called **likelihood**, is the pdf of all possible outcomes of the survey that produced the data $\mathcal{D}$, assuming that the hypothetical model $\mathcal{M}$ is true, and the right values of its parameters are θ. It measures the compatibility of the data with the hypothesis.

[8] The *probability density function* (*pdf*) of a random variable x is a positive definite function $f(x)$ defined such that the probability of x being sampled in the (a, b) interval is equal to $\int_a^b f(x)\,\mathrm{d}x$. The pdf must be normalised to 1 over the domain of x.



(a) Temperature power spectrum of Planck on March 2013, from [5]. D_ℓ stands for $\ell(\ell+1)C_\ell/2\pi$, and the shaded area represents the cosmic variance. The solid line is the prediction of the best-fit ΛCDM model.



(b) Reconstructed bispectrum of Planck on March 2013, from [7] (in colour).

Figure 1.5: CMB power spectrum and bispectrum from Planck's data release of March 2013.

$\mathcal{Z}(\mathcal{D}\,|\,\mathcal{M})$**,** called **marginal likelihood** or **evidence**, is the probability of all possible outcomes of the survey that produced the data $\mathcal{D}$, given that the model $\mathcal{M}$ is true, and having marginalised (i.e. integrated) over every possible value of the parameters θ.

We will describe the prior and likelihood pdf's in the following sections, and ignore the evidence for now, since when dealing with parameter extraction it only amounts to an uninteresting normalisation constant. The result of the parameter extraction process will therefore be the region of the parameter space Θ over which the probability

$$\mathcal{P}(\theta\,|\,\mathcal{D},\,\mathcal{M}) \propto \mathcal{L}(\mathcal{D}\,|\,\mathcal{M}(\theta))\,\pi(\theta\,|\,\mathcal{M})\ , \tag{1.75}$$

takes its highest values.

1.5.2 Prior

The prior pdf should encode two features:

Extension It should assign zero probability for those values of the theory which are considered not valid for the model, e.g. negative values of a mass.

Distribution Different approaches are possible, but one often tries to assign equal probability to values of the parameters among which there are no preferred ones, e.g. a uniform pdf in time for an instantaneous phenomenon whose timing we cannot predict. One can also choose a prior such that it lets the data, through the likelihood, be most informative when determining the constaints on the parameters (this priors are often called *non-informative*). For an extended review on different criteria for prior choice, see [26].

These aspects of the prior may present some problems:

Extension Some parameters, such as those controlling a perturbative expansion, have only *soft* limits, so their extension is not clear; e.g. the amplitude controlling a perturbative expansion may have as an upper limit 10^{-1}, 10^{-2}, $10^{-3}\ldots$

Distribution We will try to assign equal probabilities to unpreferred values of the parameters, but what pdf this precisely means depends on the particular parametrisation of the model, since lack of preference for a parameter θ does not mean the same as lack of preference for θ^2: a uniform pdf for θ transforms into a $\frac{1}{\theta}$ prior for θ^2.

Despite the problems stated above, fortunately different choices of the prior make a very small difference for highly predictive data which impose strong restrictions on the parameters' values. In the pessimistic case of a not-so-restricted model, where the choice of the priors is important, care must be taken in clearly stating the choice of the prior and how it affects the result, by showing how the posterior varies for different prior choices.

1.5.3 Digression – Objectivity, frankness and Bayesian statistics

Some people see the existence of prior pdf's in Bayesian statistics as a flaw of the approach which would introduce *unnecessary subjectivity*, since the prior I am assigning may not be the same that someone else may assign. Let us argue that it is not *unnecessary*, and it is not more *subjective* than any other model assumption.

Necessity. Confindence intervals are regions of the parameter space built in such a way that the probability of the underlying true value of the parameters falling in those intervals, given the data, is equal to a certain value, usually 68% or 95%, which correspond respectively to the 1- and 2-σ intervals of a gaussian pdf. Confidence intervals are built from posterior samples (likelihood samples in the frequentist approach), such that they contain some fraction of the posterior mass (i.e. the integral of the posterior over those intervals amounts to that fraction of the total integral). It is easy to prove that the shape of those confidence intervals unavoidably depends on the choice of parametrisation of the model. The frequentist approach uses the Fischer information matrix to account for that dependence. The Bayesian approach does it in a different way: different parametrisations are equivalent to different prior choices. Therefore, in the introduction of prior pdf's is necessary in order to account for an effect that is already explicitly present in the frequentist approach.

Subjectivity. Objectivity in science is about the methods applied to reach the results, and about putting up-front every assumption of the model under test. The assumptions themselves, though they must be sane from a scientific point of view, are chosen subjectively, up to the physical intuition of the scientist, and may change under certain circumstances. Again, it is being straightforward about those assumptions which gives the scientific endeavour its objectivity. This is precisely what Bayesian statistics does: discussing a prior choice is not a trick to get the results that one desires; it is being frank about an *unavoidable* choice of assumptions.

1.5.4 Likelihood

As we stated, the likelihood function $\mathcal{L}(\mathcal{D} \,|\, \mathcal{M}(\theta))$ is the probability that the data $\mathcal{D}$ is an actual realisation of the model $\mathcal{M}$. In our particular case, the data is the CMB power spectrum as measured from the map, while our model $\mathcal{M}$ is our calculation of it starting from the primordial perturbations described in section 1.3 and projecting them in the CMB sky following section 1.4. Of course, $\mathcal{M}$ may also be any modification of the ΛCDM model, like the one we test in chapters 2 and 3 of this thesis.

When comparing them both, one has to take into account the spurious effects that introduce differences between the power spectrum that we measure in the CMB and the one our model predicts. Those effects may be intrinsic to the instrument, such as the effect of the *beam window function* of the detector (real-world detectors, when pointed to one direction, give back not exactly the signal coming from that precise direction, but an integrated measurement that includes some signal coming from the surroundings of the direction to which we point), instrumental noise (some of the signal is generated by electronic noise in the detectors) or differences in calibration between different instruments of the survey; they may also be *foregrounds*, i.e. the physical effects that disturb the CMB photons during their journey towards us.

A likelihood function must take into account those effects and model them as effectively as possible. The likelihood function then takes the more exact form

$$\mathcal{L}\left(C_\ell^{\text{map}} \,|\, C_\ell^{\mathcal{M}}(\theta), \mathcal{M}^*(\eta)\right) \,, \tag{1.76}$$

where $C_\ell^{\mathcal{M}}(\theta)$ is the CMB power spectrum predicted by our model and parametrised by θ, and $\mathcal{M}^*$ models the aforementioned instrumental and foreground effects, parametrised by η.

For didactic purposes, let us formulate a very basic likelihood function for a very basic, fictional CMB survey. For starters, we will ignore foreground effects on top of the CMB signal, assume full-sky coverage on a single frequency, and correct only for the effects of the beam of the detector and its electronic noise, and the fact that CMB maps have finite pixelisation.

The simplest *data model* hypothesis that we can make is

$$\boldsymbol{d} \coloneqq \boldsymbol{s} + \boldsymbol{n} \,, \tag{1.77}$$

where the components d_i of the vector $\boldsymbol{d}$ are the measured values of the temperature anisotropy ${}^{\Delta T}/_{T}$ corresponding to the pixel i of the map, with a total number of pixels N_{pix}; we have divided this data into a linear combination of two components: the underlying CMB signal $\boldsymbol{s}$ and the amount of signal given by the noise, $\boldsymbol{n}$. According to our cosmological model, the temperature anisotropies caused by inflation are gaussian. Their covariance, $\langle {}^{\Delta T_i}/_{T} \cdot {}^{\Delta T_j}/_{T} \rangle$, is, according to eq. (1.58)

$$\langle s_i s_j \rangle = \sum_\ell \frac{2\ell+1}{4\pi} \hat{C}_\ell \, P_\ell(\alpha_{i,j}) \,, \tag{1.78}$$

where $\alpha_{i,j}$ is the angle between the centres of the pixels i and j. The power spectrum on the right hand side is directly calculated from the pixelised map. In order to extract the power spectrum corresponding to the underlying CMB signal, one has to take into account the effects of the finite beam of the expetiment and that of the pixelisation.

In any detector the signal measured in a direction $\hat{\boldsymbol{n}}$ does not just come from it only, but also mixes in some signal from adjacent directions. This means that in our map the temperature fluctuation measured as coming from $\hat{\boldsymbol{n}}$ is the result of the convolution with a *beam window function* B:

$$\left(\widehat{\frac{\Delta T}{T}}\right)(\hat{\boldsymbol{n}}) = \int_{\Omega_{\hat{\boldsymbol{r}}}} \left(\frac{\Delta T}{T}\right)^{\mathrm{CMB}} (\hat{\boldsymbol{r}})\, B\left(\cos(\hat{\boldsymbol{r}}, \hat{\boldsymbol{n}})\right) \mathrm{d}\Omega_{\hat{\boldsymbol{r}}} \,, \tag{1.79}$$

where we integrate over all possible directions $\hat{\boldsymbol{r}}$. If we make now a spherical transform, we get

$$\hat{a}_{\ell m} = \int_{\Omega_{\hat{\boldsymbol{n}}}} \left\{ \int_{\Omega_{\hat{\boldsymbol{r}}}} \left(\frac{\Delta T}{T}\right)^{\mathrm{CMB}} (\hat{\boldsymbol{r}})\, B\left(\cos(\hat{\boldsymbol{r}}, \hat{\boldsymbol{n}})\right) \mathrm{d}\Omega_{\hat{\boldsymbol{r}}} \right\} Y^*_{\ell m}(\hat{\boldsymbol{n}})\, \mathrm{d}\Omega_{\hat{\boldsymbol{n}}} = a^{\mathrm{CMB}}_{\ell m} B_\ell \,. \tag{1.80}$$

Notice how the spherical harmonics transformation turns the convolution with the beam window function into a product with the transformed beam window function B_l. A typical shape for a beam is a gaussian one with variance σ_B^2, whose spherical harmonic transform is $B_\ell = \exp\left(-1/2\, \ell(\ell+1)\sigma_B^2\right)$.

After applying this and a similar correction for the pixel window function (each pixel integrates signal from a finite area) encoded by W_ℓ [24], we get the variance of the signal as a function of the underlying CMB spectrum

$$\langle s_i s_j \rangle = \sum_\ell \frac{2\ell+1}{4\pi} C_\ell^{\mathrm{CMB}} B_\ell^2 W_\ell^2\, P_\ell(\alpha_{i,j}) \,, \tag{1.81}$$

We know the probability distribution for the signal part of the data: it follows a gaussian distribution with zero mean and the variance above. As for the electronic noise, we often assume it to be gaussian, with a variance $N_{i,j}$ (and zero mean). The sum of two gaussians is a gaussian, whose mean and variance are the sum of those of the components. Therefore, the pdf of the data in all pixels is the multivariate gaussian

$$\boldsymbol{d} \sim \mathcal{N}(0, \boldsymbol{\Sigma}) \,, \quad (\boldsymbol{\Sigma})_{i,j} \coloneqq \sum_\ell \frac{2\ell+1}{4\pi} C_\ell^{\mathrm{CMB}} B_\ell^2 W_\ell^2\, P_\ell(\alpha_{i,j}) + N_{i,j} \,. \tag{1.82}$$

This probability distribution is, of course, conditional on the underlying model $\mathcal{M}$ giving the CMB power spectrum, parametrised by some variables θ, and on the models for the beam and the noise, that we will ignore in the future. Therefore, as this multivariate gaussian assigns a probability for the particular realisation of the data given an underlying model, we have precisely the *likelihood* we wanted:

$$\mathcal{L}(\boldsymbol{d} \,|\, \mathcal{M}(\theta)) = \frac{1}{(2\pi)^{N_{\mathrm{pix}}/2} |\boldsymbol{\Sigma}|^{1/2}} e^{\frac{1}{2}\boldsymbol{d}\boldsymbol{\Sigma}^{-1}\boldsymbol{d}} \,, \tag{1.83}$$

where the dependence on the model enters though the covariance matrix $\mathbf{\Sigma}\left(\mathcal{M}\left(\theta\right)\right)$.

The approach described above is a nice, simple one, but it is also doomed: inverting the matrix takes $\mathcal{O}(N_{\rm pix}^3)$ operation: two orders come from the dimensions of the matrix, and the third one from the sum on ℓ, since the larger the resolution of the experiment is, $\ell_{\rm max}$, the finer the pixelisation needed. This approach makes sense only for low-resolution maps aimed at characterising the spectrum at low multipoles, $\ell \lesssim 50$, and even in that case the joint analysis of maps with multiple frequencies makes it computationally expensive, since we have to multiply $N_{\rm pix}$ by the number of maps. In addition to that, we still need to account for the effect of foregrounds, which introduce further computational complications. Still, the the method described here can be used as a basis for the temperature power spectrum likelihoods of Planck are formulated [5], in particular for $\ell < 50$. For higher multipoles, the *central limit theorem* can be exploited to gain some simplicity.

The low-ℓ Planck C_ℓ likelihood

The low-ℓ likelihood is based on a more sofisticated version of the method described above. It is known as *Commander*, and it is presented in [18]. Let us sketch it here. As a first step, we include different foregrounds effects as components of the total temperature perturbation:

$$\boldsymbol{d} \coloneqq \boldsymbol{s} + \boldsymbol{n} + \sum_i \boldsymbol{f}^{(i)} \,. \tag{1.84}$$

As for the CMB signal, we need a model for the foreground effects, $\mathcal{M}^*$ with parameters η. As this method is to be applied to small multipoles only, the resolution of the sky maps is downgraded to gain some efficiency.

Let us now look at the joint pdf of the CMB signal $\boldsymbol{s}$, its underlying estimated power spectrum $\hat{C}_\ell^{\rm CMB}$ (notice the difference with $C_\ell^{\rm CMB}$, given by the model $\mathcal{M}$), and the foreground effects $\boldsymbol{f}^{(i)}$, conditional to the measured temperature anisotropy $\boldsymbol{d}$, i.e. $\mathcal{P}\left(\boldsymbol{s}, \hat{C}_\ell^{\rm CMB}, \boldsymbol{f}^{(i)} \,|\, \boldsymbol{d}, \mathcal{M}, \mathcal{M}^*\right)$. Marginalising over all possible splittings between CMB signal part and foreground effects, we would get the desired likelihood $\mathcal{P}\left(\hat{C}_\ell^{\rm CMB} \,|\, \mathcal{M}\right)$. But that is a very complicated distribution with lots of parameters coming both from the model for the CMB and the model for the foregrounds, and not even using a Monte Carlo method, as the one described in the next section, is this pdf easy to map out.

On the contrarly, its conditional distributions for each of $\boldsymbol{s}$, $\hat{C}_\ell^{\rm CMB}$ or $\boldsymbol{f}^{(i)}$ are much simpler: the first one,

$$\mathcal{P}\left(\boldsymbol{s} \,|\, \hat{C}_\ell^{\rm CMB}, \boldsymbol{f}^{(i)}, \boldsymbol{d}, \mathcal{M}, \mathcal{M}^*\right) \equiv \mathcal{P}\left(\boldsymbol{s} \,|\, \boldsymbol{f}^{(i)}, \boldsymbol{d}, \mathcal{M}\right) \,, \tag{1.85}$$

is, as we discussed, exactly a Gaussian distribution with covariance (1.81). The

second,

$$\mathcal{P}\left(\hat{C}_\ell^{\rm CMB} \,|\, \boldsymbol{s}, \boldsymbol{f}^{(i)}, \boldsymbol{d}, \mathcal{M}, \mathcal{M}^*\right) \equiv \mathcal{P}\left(\hat{C}_\ell^{\rm CMB} \,|\, \boldsymbol{s}\right) \,, \tag{1.86}$$

can be easily calculated to be an *inverse gamma distribution*.[9] The last one,

$$\mathcal{P}\left(\boldsymbol{f}^{(i)} \,|\, \boldsymbol{s}, \hat{C}_\ell^{\rm CMB}, \boldsymbol{d}, \mathcal{M}, \mathcal{M}^*\right) \equiv \mathcal{P}\left(\boldsymbol{f}^{(i)} \,|\, \boldsymbol{s}, \boldsymbol{d}, \mathcal{M}^*\right) \,, \tag{1.87}$$

has no simple analytical form, but can be mapped out numerically without much difficulty. When the single conditional distributions are simple enough, we can take a Monte Carlo approach known as *Gibbs sampling*, consisting on applying a Markov Chain algorithm on each of the distributions alternatively (the reader may want to come back to this section once having read the next one): at a step t of the chain, first sample a point $\boldsymbol{s}_{t+1}$ from (1.85), then use the combination $(\boldsymbol{s}_{t+1}, \hat{C}_{\ell,t}^{\rm CMB}, \boldsymbol{f}_t^{(i)})$ as a starting point to sample a point $\hat{C}_{\ell,t+1}^{\rm CMB}$ from (1.86), and lastly, using $(\boldsymbol{s}_{t+1}, \hat{C}_{\ell,t+1}^{\rm CMB}, \boldsymbol{f}_t^{(i)})$ sample a point $\boldsymbol{f}_t^{(i)}$ from (1.87), getting the following step of the chain, $(\boldsymbol{s}_{t+1}, \hat{C}_{\ell,t+1}^{\rm CMB}, \boldsymbol{f}_{t+1}^{(i)})$; repeat until achieving convergence. Once we have enough samples, we can marginalise over $\boldsymbol{s}$ and the foregrounds and use the result to create an approximated interpolating likelihood $\mathcal{L}\left(\hat{C}_\ell^{\rm CMB} \,|\, \mathcal{M}\right)$.

The high-ℓ Planck C_ℓ likelihood

For higher multipoles, we need higher resolution maps, which makes the method above less eficient. Planck used the *CamSpec* algorithm [5, sec. 2.1], based on the *MASTER* approach [25].

Let $\tilde{a}_{\ell m}$ be the spherical transform in the pixelised map of the data $\boldsymbol{d}$, including the correction of a beam window function as in eq. (1.80). We can estimate from them a power spectrum, that we call *pseudo-C_ℓ*, as

$$\tilde{C}_\ell = \frac{1}{2\ell+1} \sum_m |\tilde{a}_{\ell m}|^2 \,. \tag{1.88}$$

This power spectrum contains contributions from the CMB and foregrounds, and a contribution from the instrumental noise that can be minimised if we use $\tilde{a}_{\ell m}$ of different maps to calculate this square (the instrumental noise is uncorrelated in the different instruments; though a small contribution is left, it is guaranteed to be unbiased, i.e. to have zero mean). It can be proven [25] that the underlying joint power spectrum of CMB and foregrounds, $\hat{C}_\ell$, can be recovered from the pseudo-C_ℓ multiplying by a convolution matrix $M_{\ell\ell'}$ that accounts for the leaking between multipoles due to the finite beam window function, the pixelisation of the map and the use of non-full sky coverage (a big chunk of the sky, containing the galactic plane up to a elevation and a number of point sources, is discarded

[9] See e.g. `http://en.wikipedia.org/wiki/Inverse-gamma_distribution` .

to minimise the effect of the anisotropic part of the foregrounds). Since the convolution matrix can be calculated explicitly (see [5, app. A.1]), we can work directly with the pseudo-C_ℓ.

Working with the $\tilde{C}_\ell$ has an important advantage. Given isotropy of the signal,[10] for equal ℓ all the $\tilde{a}_{\ell m}$ are distributed following the same pdf, with a well-defined mean and variance. On the other hand, according to eq. (1.88), $\tilde{C}_\ell$ is at each ℓ equal to the average of the $\tilde{a}_{\ell m}$ over m. For high values of ℓ (remember, here $\ell > 50$), up to a very good approximation the *central limit theorem* is fulfilled:[11] the pseudo-C_ℓ follow a simple gaussian distribution. Therefore, the likelihood can simply be written as a multivariate gaussian – the product of a gaussian for each ℓ. The covariance matrix $\mathbf{\Sigma}$ of the multivariate gaussian, giving the dependence between the different multipoles (and between different combinations of frequencies) is estimated from simulations performed on a fiducial model. Having precomputed it, the value of the gaussian likelihood can be computed in a very small amount of time.

As in the low-ℓ case, the likelihood is conditional on the models for the CMB power spectrum and the isotropic foreground effects, as well here as on the different calibration parameters for the maps used. Marginalising at the time of sampling over these *nuisance* effects one can get the desired likelihood $\mathcal{L}\left(\hat{C}_\ell^{\mathrm{CMB}} \mid \mathcal{M}\right)$.

1.5.5 Monte Carlo methods

Now that we have all the necessary ingredients to extract the posterior pdf of the parameters of the model and characterise it around its maximum, we may try to do so in the most naïve way: analytically maximising the product of prior and likelihood using a gradient method, i.e. following the direction of the function with the highest slope until we find its maximum. We will immediately find this approach to be practically unfeasible, due to problems such as

Non-analyticity of the likelihood: In the particular case of the CMB, the calculation of the CMB power spectrum from the primordial conditions, based on eq. (1.69), cannot be done analytically, but is the result of numerical solutions of a system of Boltzmann equations, often solved with the help of computers [30, 11].

Cost of calculating the likelihood: In the particular case of the CMB, the full calculation of the likelihood from the set of parameters of the model may take up to a few seconds. This discourages us from attempting a thorough sampling on a grid in the parameter space, aimed at constructing an analytic approximation to the likelihood as an interpolating function.

[10] The most anysotropic foregrounds, Via Lactea and point sources, have been masked away. The remaining ones are isotropic to a good approximation.

[11] See e.g. `http://en.wikipedia.org/wiki/Central_limit_theorem` .

Size of the parameter space: The dimensionality of the typical parameter spaces, including interesting parameters of the model and uninteresting parameters modelling the survey's shortcomings, easily goes to high values. This too prevents us from searching an interpolating approximation of the likelihood by sampling on a grid.

Complicated shape of the likelihood: Even in the optimistic case of an analytic likelihood, it is usually a very complicated function with many local maxima, in which a gradient method may get trapped.

This problems tell us about the desired features of an ideal solution: it needs a mechanism for not getting trapped in local maxima, often realised through some random jumping; it must sample as few points of the parameter space as possible, and it must scale well with the dimensionality of the parameter space.

The result of this method must be a *fair sample*, or set of pairs of points in the parameter space and their respective values of the likelihood, possibly including a relative *sampling weight*. *Fairness* of the sample means that in every region, the density of the sampling must be proportional to the probability density. If the goal is to obtain constraints for the parameters of the model, the samples only need to cover the region of the parameter space where the posterior has most of its probability mass.

We call *Monte Carlo methods* a broad family of random sampling algorithms aimed at solving this kind of problems.

1.5.6 Markov chain Monte Carlo

In this section we will start by briefly defining the general concept of a Markov chain, its expected properties, and the Metropolis-Hastings algorithm. Details and proofs can be found in [23, Ch. 1 & 4]

Let X_0, $X_1, \ldots, X_t$ be a sequence of random variables taking values on a *state space* E. A (discrete time) *Markov chain* is a sequence of such random variables X_0, $X_1, \ldots, X_t$ such that the probability of the random variable to take values in a particular subset $A \subset E$ in the next step depends *only* on the value taken in the current step:

$$\mathcal{P}\left(X_{t+1} \in A \,|\, X_0 = x_0,\, X_1 = x_1, \ldots, X_t = x_t\right) \;=\; \mathcal{P}\left(X_{t+1} \in A \,|\, X_t = x_t\right) \,. \tag{1.89}$$

A Markov chain is *time-homogeneous* or *stationary* whenever this probability does not depend on the specific order of the step for which the last property is fulfilled:

$$\mathcal{P}\left(X_{t+1} \in A \,|\, X_t = x\right) \;=\; \mathcal{P}\left(X_t \in A \,|\, X_{t-1} = x\right) \qquad \forall t \,, \tag{1.90}$$

i.e. $\mathcal{P}\left(X_{t+1} \in A \,|\, X_t = x\right)$ is independent of the order of the step, and we call it the *transition kernel*. The probability distribution of any step of a stationary

Markov chain is completely specified by its transition kernel and the probability distribution of the initial step.

A stationary Markov chain has *stationary* or *invariant distribution* ρ when, being ρ the initial probability distribution at step t,

$$\mathcal{P}\left(X_{t+1} \in A \,|\, X_t \sim \rho\right) \;=\; \rho(A) \qquad \forall t \;. \tag{1.91}$$

Thus, *the kernel of a stationary Markov chain is determined by its stationary distribution.*

As it should be obvious by now, we will try to build a Markov chain such that the stationary distribution is the pdf that we wish to sample, in this case the *posterior pdf* $\mathcal{P}\left(\theta|\mathcal{D},\mathcal{M}\right)$.

At this point, looking at eq. (1.91), it appears that the points in the chains are drawn from the stationary pdf ρ. One would be tempted to estimate the desired averages over ρ, $\langle f(X)\rangle_\rho$, using averages over the states along the chain, i.e. *ergodic* averages:

$$\bar{f}_n \;\coloneqq\; \frac{1}{n+1}\sum_{i=0}^{n} f(X_i) \;, \tag{1.92}$$

but there are two important caveats:

1. For the states of the chain to really be samples from ρ, according to eq. (1.91) all the previous states must be so, *including the initial state* X_0, which obviously cannot be sampled from ρ yet, since sampling from ρ is precisely the problem at hand. Therefore, we need to ensure that the initial state will be *forgotten* at some point, leaving us with pure samples from ρ. We achieve so by discarding a number of initial states of the chain when calculating (1.92), which we call *burn-in states*.

2. Still, the samples are not *independent* from one another, but correlated samples from ρ. This needs not be a problem, provided that the samples are drawn in the correct proportion, i.e. drawing from $A \subset E$ must be proportional to $\rho(A)$, i.e. the sample, though correlated, must be *fair*.

Ensuring this two requirements, or having *convergence* of the path-average $\bar{f}_n$ to the true expected value under the stationary pdf $\langle f\rangle_\rho$, is defined as the Markov chain being *ergodic*. In order to ensure ergodicity, one must prove three properties of the chain: that it is *irreducible*, i.e. it explores all subsets $A \subset E$ such that $\rho(A) \neq 0$, *recurrent*, i.e. an infinite chain reaches said regions infinitely often, and *aperiodic*, i.e. it does not transition cyclically through a fixed sequence of sets.

We now propose a very simple algorithm to build a Markov chain that is stationary and convergent, and which is the basis, with some improvements, of most of the widespread cosmological MCMC codes [29, 9]. It is called the *Metropolis-Hastings algorithm*.

Suppose a *proposal* pdf $\phi(Y|X_t)$, where X_t is the value at the current step. For example, $\phi(Y|X_t) \sim \mathcal{N}(X_t, \mathcal{C})$, a multivariate gaussian centred at X_t and with a fixed covariance matrix $\mathcal{C}$. Let us suppose the we draw a sample y from ϕ. Assuming ρ to be the stationary pdf of the chain, we build the quantity

$$\alpha(X_t, Y) \coloneqq \min\left(1,\, \frac{\rho(Y)}{\rho(X)}\frac{\phi(X_t \,|\, Y)}{\phi(Y \,|\, X_t)}\right) \,. \tag{1.93}$$

Notice that $\alpha \in [0, 1]$. The new point Y is accepted or rejected as the next point of the chain with probability α. If it is rejected, take $X_{t+1} = X_t$. In any case, repeat for X_{t+2}. Notice how in eq. (1.93) the ratio of the stationary distribution drives the next step towards the maxima of this pdf. Notice too how the choice of a gaussian proposal distribution implies that the second fraction has no effect.

The proposal distribution ϕ and the stationary distribution of the chain ρ is all we need to build the Metropolis-Hasting Markov chain. It is easy to see that its kernel is

$$\begin{aligned}\mathcal{P}\left(X_{t+1} \,|\, X_t\right) =& \phi(X_{t+1} \,|\, X_t)\, \alpha(X_t,\, X_{t+1}) + \\ & I(X_{t+1} = X_t)\left[1 - \int \phi(Y \,|\, X_t)\, \alpha(X_t,\, Y)\, \mathrm{d}Y\right] \,,\end{aligned} \tag{1.94}$$

where $I(cond)$ is a function valued 1 if *cond* is true and 0 otherwise.

It can be proven that this chain converges towards the stationary distribution ρ regardless of the choice of the proposal distribution ϕ. This does not mean that this is not an important choice: the convergence will be faster the closest the proposal distribution is to the stationary distribution – a bad choice of ϕ will have bad consequences in the efficiency of the sampling of ρ: if ϕ is much more concentrated than ρ, we will not reach to explore the tails of the distribution ρ; if ϕ is very spread, we will tend to draw candidate samples Y such that $\rho(Y) \ll \rho(X_y)$, away from the maxima, so we will very often reject them, resulting in a chain with many repeated steps and few different ones. Therefore, assuming a nearly-gaussian stationary pdf, one would ideally try to use a gaussian proposal pdf with a covariance matrix as close as possible to that of the stationary distribution.[12] In order to get such a covariance matrix, we can approximate it from the sample covariance of a small chain of sample of ρ.

1.5.7 Final remarks

Having described the likelihood and prior pdf's, and one of the Monte Carlo methods aimed at mapping the posterior, the next step is to actually use these tools to tackle the two most common computational tasks in statistical inference: *parameter inference* and *model selection* (in frequentist terms, equivalent to *hypothesis*

[12] In fact, it is advisable to use a covariance matrix for the proposal distribution that is a little broader than that of the stationary distribution, since the tails of the latter get better sampled this way.

testing). We will not go into much detail here, but just introduce how these two problems have been approached in the research performed for this thesis.

Parameter inference

Parameter inference is the problem of, given a model with some parameters, characterising the regions of the parameter space containing most of the posterior probability, i.e. containing the most likely underlying parameter values. These regions are called *confidence level regions* or *intervals*. Computing them requires a sampling method that produces *fair samples*, since we will estimate the probability corresponding to those regions with the proportion of the Monte Carlo samples falling inside them. Since confidence level intervals are, of course, centred around the maxima of the posterior, the sampling method used must guarantee that the regions of higher posterior value are thoroughly explored, as Markov chain Monte Carlo does.

If the resulting posterior is at least approximately gaussian, a set of nested confidence level intervals (usually 68% and 95%) summarises the posterior pdf well enough. In the opposite case, of very non-gaussian distribution with several regions of high probability, not only confidence intervals are not enough to summarise the information contained in the posterior, but also the simple Markov chain approach explained in the last section needs a unmanageably large amount of running time in order to guarantee that the sampling is fair. This is so because the chain gets *stuck* around one of the several maxima (or *modes*), and only rarely jumps between all of them, and only after a large number of *exchanges* between the modes can their relative mass probabilities be fairly represented.

The case of a non-gaussian distribution with several modes is prone to pop out whenever we attempt to constrain parameters of models which exploit small anomalies close to the signal-to-noise ratio of the data set; this is what we do in this study with CMB and LSS data. In these cases, one often prefers alternative methods to Markov chain Monte Carlo, specially aimed at mapping weirdly-shaped pdf's, such as *multi-modal nested sampling* [38, 21, 22, 20]. In the first part of this study, however, we managed to map a multi-modal posterior using solely MCMC's. The method we used is described in section 2.3.

Model selection

Model selection is the problem of assessing which one of a set of models is the most likely to have generated the data at hand, or equivalently which one is the most informative or predictive. The quantification of this predictivity is given by the *evidence*, defined in section 1.5.1, which is the marginalisation of the data likelihodd over the priors allowed by the theory, i.e. the total probability of the model being the underlying one regardless of the values of its parameters. A short review of the procedure followed in *Bayesian model selection* can be found in section 3.3.

In particular, in this study we do not aim to characterise the total evidence of the model discussed in section 1.3.4, since we do not explore the full parameter region allowed by the model, but instead explore a significant patch of it and use model selection arguments to argue about the consistency between candidate signals in CMB and LSS data. The reader is referred to chapter 3 for the complete discussion. The sampling method used in this part of the study is the aforementioned multi-modal nested sampling, which is briefly described in section 4.3.

Bibliography

[1] L. F. Abbott and M. B. Wise. Large-scale anisotropy of the microwave background and the amplitude of energy density fluctuations in the early universe. *Astrophys.J.Lett.*, 282:L47–L50, July 1984.

[2] Ana Achucarro, Jinn-Ouk Gong, Sjoerd Hardeman, Gonzalo A. Palma, and Subodh P. Patil. Effective theories of single field inflation when heavy fields matter. *JHEP*, 1205:066, 2012, 1201.6342.

[3] Ana Achucarro, Jinn-Ouk Gong, Gonzalo A. Palma, and Subodh P. Patil. Correlating features in the primordial spectra. *Phys.Rev.*, D87:121301, 2013, 1211.5619.

[4] Viviana Acquaviva, Nicola Bartolo, Sabino Matarrese, and Antonio Riotto. Second order cosmological perturbations from inflation. *Nucl.Phys.*, B667:119–148, 2003, astro-ph/0209156.

[5] P.A.R. Ade et al. Planck 2013 results. XV. CMB power spectra and likelihood. 2013, 1303.5075.

[6] P.A.R. Ade et al. Planck 2013 results. XXII. Constraints on inflation. 2013, 1303.5082.

[7] P.A.R. Ade et al. Planck 2013 Results. XXIV. Constraints on primordial non-Gaussianity. 2013, 1303.5084.

[8] P.A.R. Ade et al. Planck 2013 results. XVI. Cosmological parameters. *Astron.Astrophys.*, 2014, 1303.5076.

[9] Benjamin Audren, Julien Lesgourgues, Karim Benabed, and Simon Prunet. Conservative Constraints on Early Cosmology: an illustration of the Monte Python cosmological parameter inference code. *JCAP*, 1302:001, 2013, 1210.7183.

[10] Daniel Baumann. TASI Lectures on Inflation. 2009, 0907.5424.

[11] Diego Blas, Julien Lesgourgues, and Thomas Tram. The Cosmic Linear Anisotropy Solving System (CLASS) II: Approximation schemes. *JCAP*, 1107:034, 2011, 1104.2933.

[12] T.S. Bunch and P.C.W. Davies. Quantum Field Theory in de Sitter Space: Renormalization by Point Splitting. *Proc.Roy.Soc.Lond.*, A360:117–134, 1978.

[13] Clifford Cheung, Paolo Creminelli, A. Liam Fitzpatrick, Jared Kaplan, and Leonardo Senatore. The Effective Field Theory of Inflation. *JHEP*, 0803:014, 2008, 0709.0293.

[14] Paolo Creminelli and Matias Zaldarriaga. Single field consistency relation for the 3-point function. *JCAP*, 0410:006, 2004, astro-ph/0407059.

[15] Tamara M. Davis and Charles H. Lineweaver. Superluminal recession velocities. 2000, astro-ph/0011070.

[16] Tamara M. Davis and Charles H. Lineweaver. Expanding confusion: common misconceptions of cosmological horizons and the superluminal expansion of the universe. *Proc.Astron.Soc.Austral.*, 2003, astro-ph/0310808.

[17] S. Dodelson. *Modern Cosmology*. 2003.

[18] H. K. Eriksen, J. B. Jewell, C. Dickinson, A. J. Banday, K. M. Górski, and C. R. Lawrence. Joint bayesian component separation and cmb power spectrum estimation. *The Astrophysical Journal*, 676(1):10, 2008.

[19] J.R. Fergusson, M. Liguori, and E.P.S. Shellard. The CMB Bispectrum. *JCAP*, 1212:032, 2012, 1006.1642.

[20] F. Feroz, M. P. Hobson, E. Cameron, and A. N. Pettitt. Importance Nested Sampling and the MultiNest Algorithm. *ArXiv e-prints*, June 2013, 1306.2144.

[21] F. Feroz, M.P. Hobson, and M. Bridges. MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics. *Mon.Not.Roy.Astron.Soc.*, 398:1601–1614, 2009, 0809.3437.

[22] Farhan Feroz and M.P. Hobson. Multimodal nested sampling: an efficient and robust alternative to MCMC methods for astronomical data analysis. *Mon.Not.Roy.Astron.Soc.*, 384:449, 2008, 0704.3704.

[23] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. *Markov Chain Monte Carlo In Practice*. Chapman and Hall/CRC, 1999.

[24] K.M. Gorski, Eric Hivon, A.J. Banday, B.D. Wandelt, F.K. Hansen, et al. HEALPix - A Framework for high resolution discretization, and fast analysis of data distributed on the sphere. *Astrophys.J.*, 622:759–771, 2005, astro-ph/0409513.

[25] E. Hivon, K.M. Gorski, C.B. Netterfield, B.P. Crill, S. Prunet, et al. Master of the cosmic microwave background anisotropy power spectrum: a fast method for statistical analysis of large and complex cosmic microwave background data sets. *Astrophys.J.*, 567:2, 2002, astro-ph/0105302.

[26] Robert E. Kass and Larry Wasserman. The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, 91(435):1343–1370, 1996.

[27] L.V. Keldysh. Diagram technique for nonequilibrium processes. *Zh.Eksp.Teor.Fiz.*, 47:1515–1527, 1964.

[28] Thomas S. Kuhn. *The structure of scientific revolutions.* University of Chicago Press, Chicago, 1970.

[29] Antony Lewis and Sarah Bridle. Cosmological parameters from CMB and other data: a Monte- Carlo approach. *Phys. Rev.*, D66:103511, 2002, astro-ph/0205436.

[30] Antony Lewis, Anthony Challinor, and Anthony Lasenby. Efficient computation of CMB anisotropies in closed FRW models. *Astrophys.J.*, 538:473–476, 2000, astro-ph/9911177.

[31] M. Liguori, E. Sefusatti, J. R. Fergusson, and E. P. S. Shellard. Primordial Non-Gaussianity and Bispectrum Measurements in the Cosmic Microwave Background and Large-Scale Structure. *Advances in Astronomy*, 2010, 2010, 1001.4707.

[32] Juan Martin Maldacena. Non-Gaussian features of primordial fluctuations in single field inflationary models. *JHEP*, 0305:013, 2003, astro-ph/0210603.

[33] V. F. Mukhanov. Gravitational instability of the universe filled with a scalar field. *ZhETF Pisma Redaktsiiu*, 41:402–405, May 1985.

[34] S. Perlmutter et al. Measurements of Omega and Lambda from 42 high redshift supernovae. *Astrophys.J.*, 517:565–586, 1999, astro-ph/9812133.

[35] Adam G. Riess et al. Observational evidence from supernovae for an accelerating universe and a cosmological constant. *Astron.J.*, 116:1009–1038, 1998, astro-ph/9805201.

[36] Matts Roos and S.M. Harun-or Rashid. How flat is the universe? 2000, astro-ph/0003040.

[37] Misao Sasaki. Large Scale Quantum Fluctuations in the Inflationary Universe. *Prog.Theor.Phys.*, 76:1036, 1986.

[38] John Skilling. Nested sampling. *AIP Conference Proceedings*, 735(1):395–405, 2004.

[39] David Wands, Karim A. Malik, David H. Lyth, and Andrew R. Liddle. A New approach to the evolution of cosmological perturbations on large scales. *Phys.Rev.*, D62:043527, 2000, astro-ph/0003278.

[40] Steven Weinberg. Adiabatic modes in cosmology. *Phys.Rev.*, D67:123504, 2003, astro-ph/0302326.

[41] Steven Weinberg. Quantum contributions to cosmological correlations. *Phys.Rev.*, D72:043514, 2005, hep-th/0506236.