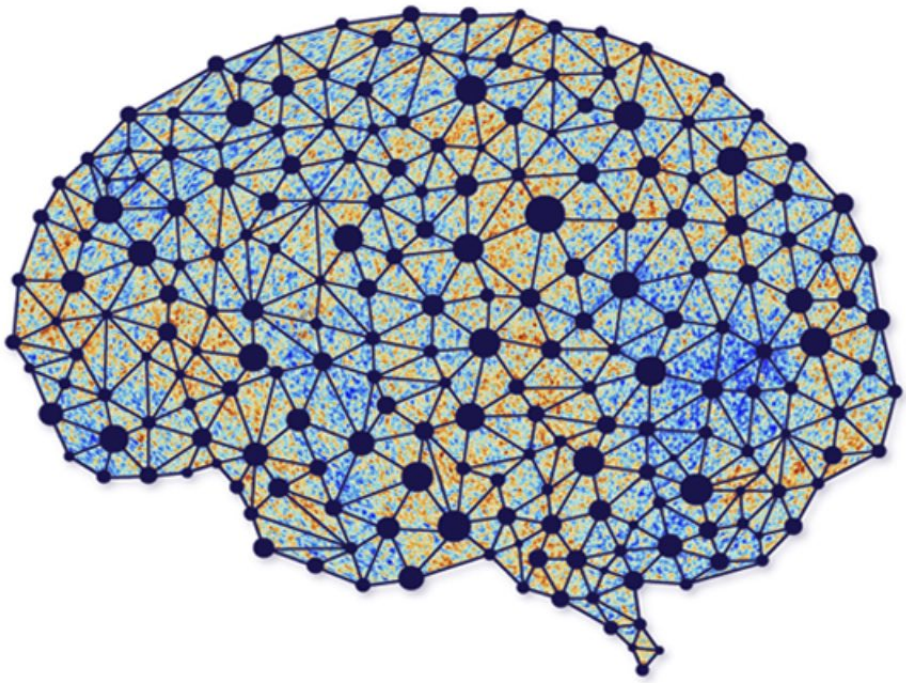
Lecture Notes on

Computational Cosmology

Moritz Münchmeyer

Physics Department, University of Wisconsin-Madison



Version: *June 27, 2024*

Contents

1 Introduction

This is a one-semester course on Computational Cosmology, aimed in particular at beginning graduate students working in cosmology. My goal in these lectures is to focus on computational and statistical methods that data oriented cosmologists need, while being brief on theoretical foundations that can be learned later on in more detail. For example, topics that are not treated in any detail here are inflation, BBN and relativistic perturbation theory. These topics are typically covered in a course on theoretical cosmology. There are many excellent texts on these topics.

These lectures are primarily about the two most important data sources for cosmology: the Cosmic Microwave Background and Large-Scale Structure (LSS) surveys. Of course there are many other useful probes of the universe, such as Supernovae, Strong Lensing and Gravitational Waves. However we won't have time to cover these in any detail this semester.

For the CMB and LSS we will cover how cosmological parameters can be measured from data using both the standard methods and more recent developments. My goal is to be broad rather than detailed, and mention many of the techniques that all cosmologists should know. I also try to provide good references for each section so you can dig deeper.

The course is organized as follows. In Unit 1, we discuss fundamentals of cosmology for those of you that are completely new to the field. If you studied theoretical cosmology before you can perhaps skip this unit. In Unit 2, we introduce common statistical and computational tools required for data analysis, such as random fields, correlation functions, likelihoods and MCMC. We finish this unit by analyzing an N-body simulation to measure its cosmological parameters. In Unit 3 we discuss the CMB, focussing on data analysis methods. We include brief discussions of several advanced topics, such as lensing and kSZ. In Unit 4, we study data analysis in large-scale structure. We both discuss the classic power spectrum analysis, and more recent developments such as simulation-based inference.

Please check out my website https://munchmeyer.physics.wisc.edu/lecture-notes/ for computational notebooks, links, and updated versions of this course. If you find any errors or typos, please drop me an email so that I can fix them (muenchmeyer@wisc.edu). If you use these notes for teaching or learning, I'd also be very happy to hear about it. If you want to build on these notes, the latex version is available on request.

I thank Utkarsh Giri for contributing the MCMC analysis of the Quijote Power Spectrum and Sai Tadepalli for developing and teaching the section on EFTofLSS. I also thank Jacob Audette for making the front page graphic.

Part I
Basics of Cosmology

In the first part of this course we will briefly review basics of cosmology. My main goal is to introduce the coordinates as well as the physical parameters that we want to measure in data later on.

Further reading

There are many excellent textbooks that go deeper into these foundations. This section is based primarily on the text books

- Daniel Baumann - Cosmology (2021) as well as his TASI lecture notes (arxiv: 0907.5424). Material similar to the textbook is also available here: http://cosmology.amsterdam/education/cosmology/.
- Dragan Huterer - A course in cosmology (2023)

I'm also using David Tong's cosmology lecture notes from

- http://www.damtp.cam.ac.uk/user/tong/teaching.html

I recommend all of David's lecture notes. Another popular textbook which we will use is

- Dodelson, Schmidt - Cosmology (2020).

2 Scales and Units

The universe is obviously enormous. In fact, while the **observable universe** is finite (due to the finite age of the universe and the finite speed of light), it is not known whether the universe is finite as a whole.

Cosmologists like to measure distances in parsec (pc). The conversion from pc to lightyears is

$$1 \text{ pc} = 3.26 \text{ ly} \tag{2.1}$$

Parsecs are the typical distance of nearby stars. A parsec is equal to the distance at which 1 AU (astronomical unit – average distance between Earth and the Sun) is seen at an angle of one arc second, which is $\frac{1}{3600}$ of a degree. The size of our galaxy is more conveniently given in kilo parsec (kpc). The Milkyway is about 30 kpc in diameter, and we are about 8kpc from the center. The distance to other galaxies is usually given in mega parsec (Mpc). The nearest spiral galaxy, Andromeda, is about 1Mpc away from us. The comoving distance (we'll explain the term "comoving" soon) to the edge of the observable universe is about 14.3 Gpc.

An order of magnitude estimate is that the observable universe contains about 100 billion galaxies and a typical galaxy contains about 100 billion stars. There is no reason to believe that our galaxy or star are particularly special in the cosmological sense.

3 Expansion of the Universe

In this section we want to understand the equations that govern the evolution of the entire universe. Cosmology can be understood in two steps:

- On large scales (i.e. after smoothing out small scale irregularities such as galaxies), **the universe is uniform**. By studying its average contents we can understand the **background expansion** of the universe. The **Cosmological Principle states: On the largest scales, the universe is spatially homogeneous and isotropic.**
- On smaller scales, there are initially small and later very large inhomogeneities (such as galaxies). The evolution of these **cosmological perturbations** on top of the background expansion is much more complicated, but tells us much of what we know about the universe.

Following the standard practice of cosmology courses, we will first discuss the uniform large-scale universe and then later discuss perturbations.

To start describing the universe mathematically we first need to define coordinates. A crucial feature of cosmology is that space-time cannot be treated statically because the universe is expanding very substantially during its history.

A large part of theoretical and computational cosmology can be done by assuming that the universe is flat. To date there is no experimental evidence for any curvature of space on large scales. We thus focus on flat expanding space-time and are brief on the generalization to curvature.

3.1 The metric

We can define a space-time through the so called **metric**. The metric can be thought of as a mathematical object that turns coordinate distances (which are a matter of coordinate choice) into physical distances (which are invariant and thus physically meaningful). For 3-dimensional Euclidean space the physical distance dl is related to Euclidean coordinate distance (dx,dy,dz) by

$$dl^2 = dx^2 + dy^2 + dz^2 = \sum_{i,j=1}^{3} \delta_{ij} dx_i dx_j, \tag{3.1}$$

where the Kronecker delta $\delta_{ij} = \mathrm{diag}(1,1,1)$ is the metric. If we were to use spherical coordinates instead we would get

$$dl^2 = dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta d\phi^2 = \sum_{i,j=1}^{3} g_{ij} dx_i dx_j, \tag{3.2}$$

where $(x_1, x_2, x_3) = (r, \theta, \phi)$ and the metric is $g_{ij} = \mathrm{diag}(1, r^2, r^2 \sin^2\theta)$.

Since Einstein we know that physics is really happening in **space-time** and that distances in time and space are not independently invariant. We instead need a metric that turns space-time coordinates $x^\mu = (ct, x_i)$ into the **invariant space-time distance (also called invariant line element)**

$$ds^2 = \sum_{\mu,\nu=0}^{3} g_{\mu\nu} dx^\mu dx^\nu \equiv g_{\mu\nu} dx^\mu dx^\nu. \tag{3.3}$$

In the specific case of special relativity where space-time is not curved (**Minkowski space**), using Euclidean coordinates, this line element is

$$ds^2 = -c^2 dt^2 + \sum_{i,j=1}^{3} \delta_{ij} dx_i dx_j \tag{3.4}$$

$$= -c^2 dt^2 + d\mathbf{x}^2 \tag{3.5}$$

and the **Minkowski metric** is $g_{\mu\nu} = \text{diag}(-1, 1, 1, 1)$. Recall from special relativity that ds can be positive, negative or null.

In general relativity, the metric depends on the position in space-time, $g_{\mu\nu}(t, \mathbf{x})$. The metric is of course coordinate dependent. To give a physical description of curvature that is independent of the choice of coordinates one needs to use the formalism of Riemann geometry. In this course we won't need much of that.

3.2 Flat expanding space time

The space-time metric of cosmology, assuming a flat space (but not flat space-time) is a simple generalization of Minkowski space, where we scale the spatial part of the metric with the time dependent **scale factor $a(t)$**:

$$ds^2 = -c^2 dt^2 + a(t)^2 d\mathbf{r}^2 \tag{3.6}$$

The spatial coordinates $\mathbf{r}$ are called the **comoving coordinates**. The comoving coordinate of an object does not change due to the expansion of space time. The comoving coordinate system expands with spacetime, as illustrated in Fig. 1. In computational cosmology we usually work with comoving coordinates (e.g. comoving galaxy positions in an N-body simulation). The scale factor is usually defined to be equal to 1 today, $a(t_{today}) = 1$. To define the coordinates we also need to set some origin O where $r = 0$ and $t = 0$.

We also define the **physical coordinates** $\mathbf{r}_{phys}(t) = a(t)\mathbf{r}(t)$. If an object has a trajectory $\mathbf{r}(t)$ in comoving coordinates and $\mathbf{r}_{\text{phys}} = a(t)\mathbf{r}$ in physical coordinates, the physical velocity of the object is

$$\mathbf{v}_{\text{phys}} \equiv \frac{d\mathbf{r}_{\text{phys}}}{dt} = \frac{da}{dt}\mathbf{r} + a(t)\frac{d\mathbf{r}}{dt} \equiv H(t)\mathbf{r}_{\text{phys}} + \mathbf{v}_{\text{pec}}, \tag{3.7}$$

where we have introduced the **Hubble parameter**

$$H \equiv \frac{\dot{a}}{a} \tag{3.8}$$

and the **peculiar velocity**

$$v_{\text{pec}} \equiv a(t)\dot{\mathbf{r}} \tag{3.9}$$

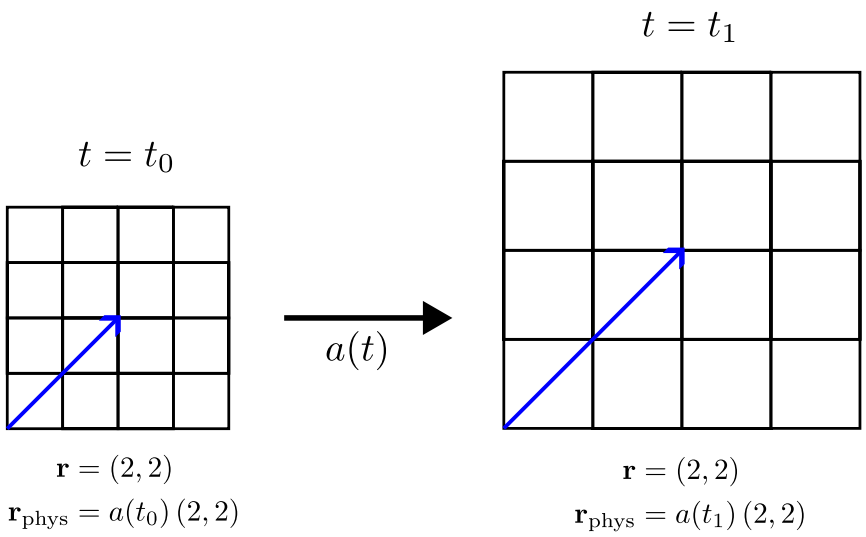


Figure 1. Comoving coordinate grid on an expanding spacetime.

The first term $H\mathbf{r}_{\text{phys}}$ is the **Hubble flow**, which is the physical velocity of the object due to the expansion of space between the origin and the object. This expression is a version of Hubble's law (though not the original one where H is time-independent). The second term, the peculiar velocity, describes the motion of the object relative to the **cosmological rest frame**. Typical peculiar velocities of galaxies are hundreds of km/s, so $\beta = \frac{v}{c} \simeq 10^{-3}$. The present day value of the Hubble parameter is[1]

$$H_0 \simeq 67.8 \text{ km } s^{-1} \text{ Mpc}^{-1} \tag{3.10}$$

This is telling us that a galaxy 1 Mpc away will be seen to be retreating at a speed of 67.8 km/s due to the expansion of space. Galaxies that are farther away than a few Mpc thus have a larger recession speed due to the Hubble flow than due to their peculiar velocities. A common definition of the Hubble parameter is given by introducing h so that $H_0 = 100\ h$ km s^{-1} Mpc^{-1} with $h \approx 0.678$.

It is often useful to write the metric in polar coordinates:

$$ds^2 = -c^2dt^2 + a^2(t)\left(dr^2 + r^2 d\Omega^2\right) \tag{3.11}$$

where

$$d\Omega^2 = d\theta^2 + \sin^2\theta d\phi^2 \tag{3.12}$$

is the metric on the unit two-sphere. This metric is useful to describe observations by an observer at the coordinate center of the universe. The radial coordinate r is called the **comoving distance** to the origin.

A further way to write the metric is by introducing **conformal time**

$$d\eta = \frac{dt}{a(t)} \tag{3.13}$$

[1]You may notice that I am primarily a CMB cosmologist (see the "Hubble tension").

Conformal time slows down with the expansion of the universe. The metric is then

$$ds^2 = a^2(\eta)\left(-c^2 d\eta^2 + dr^2 + r^2 d\Omega^2\right) \tag{3.14}$$

The scale-factor is now a time-dependent overall factor infront of a static metric. Conformal coordinates are especially useful to analyze light rays and causality.

3.3 Curved expanding space time

The generalization of the line element (3.11) to curved space-time is

$$ds^2 = -c^2 dt^2 + a^2(t)\left(\frac{dr^2}{1-\frac{kr^2}{R_0^2}} + r^2 d\Omega^2\right) \tag{3.15}$$

This general form is called the **Friedmann-Lemaitre-Robertson-Walker (FLRW)** metric. Here the constant is $k = 0$ for flat space, $k = 1$ for positively curved space and $k = -1$ for negatively curved space. R_0 is the **curvature scale**, which defines how strongly the space is curved. The three spaces are the three maximally symmetric three-spaces, i.e. they are homogeneous and isotropic. This metric is derived in most textbooks. Note that the FLRW metric is not invariant under Lorentz transformation. This means that the universe picks out a preferred rest frame, described by co-moving coordinates (physically the matter content breaks invariance under Lorentz boosts). A **closed universe (positive curvature)** has a finite volume and the angles of a triangle add up to more than 180° (like on a sphere). An **open universe (negative curvature)** has infinite volume and the angles of a triangle add up to less than 180°.

This metric is sometimes written by defining the radial coordinate

$$d\chi \equiv \frac{dr}{\sqrt{1-\frac{kr^2}{R_0^2}}} \tag{3.16}$$

which can be integrated to obtain $r = S_k(\chi)$ where

$$S_k(\chi) \equiv R_0 \begin{cases} \sinh(\chi/R_0), & k=-1 \\ \chi/R_0, & k=0 \\ \sin(\chi/R_0), & k=1 \end{cases} \tag{3.17}$$

The metric is then

$$ds^2 = -c^2 dt^2 + a^2(t)\left(d\chi^2 + S_k^2(\chi) d\Omega^2\right) \tag{3.18}$$

Note that for flat space-time $k = 0$ there is no difference between comoving distance r and radial coordinate χ. Finally, using again conformal time, this metric can be written as

$$ds^2 = a^2(\eta)\left[-c^2 d\eta^2 + \left(d\chi^2 + S_k^2(\chi) d\Omega^2\right)\right] \tag{3.19}$$

3.4 Redshift

The expansion of the universe means that light rays which travel through the universe also get stretched. This is called **cosmological redshift**. We define the dimensionless **redshift parameter**

$$z = \frac{\lambda_0 - \lambda_1}{\lambda_1} = \frac{f_1 - f_0}{f_0} \tag{3.20}$$

where λ_0 is the observed wave length and λ_1 is the emitted wavelength. It turns out that the ratio of wavelength scales as the ratio of the scale factor:

$$\frac{\lambda_0}{\lambda_1} = \frac{a(t_0)}{a(t_1)} \tag{3.21}$$

This result is intuitive: the photon wave is stretched with the expansion of space. Let's derive this result starting from the metric. In general relativity, light rays travel along null geodesics, meaning that $ds = 0$. A light ray on a radial direction (with fixed θ and ϕ) will thus obey

$$c\ dt = \pm a(t) d\chi \tag{3.22}$$

where the minus sign describes light moving towards us (i.e. as t gets larger χ gets smaller). Thus we have

$$\frac{c\ dt}{a(t)} = \pm d\chi \tag{3.23}$$

Let's consider a crest of the light wave to be emitted at time t_1 from distance χ_1. We observe the crest at time t_0 at position $\chi_0 = 0$. Thus we get the integral equation

$$\int_{t_1}^{t_0} \frac{c\, dt}{a(t)} = \int_0^{\chi_1} d\chi \qquad \text{first crest} \tag{3.24}$$

The next crest of the wave is emitted at time $t_1 + \delta t_1$ and received at time $t_0 + \delta t_0$. Thus the integral equation is

$$\int_{t_1+\delta t_1}^{t_0+\delta t_0} \frac{c\, dt}{a(t)} = \int_0^{\chi_1} d\chi \qquad \text{second crest} \tag{3.25}$$

The right hand side of these two relations are the same and thus we have

$$\int_{t_1}^{t_0} \frac{dt}{a(t)} = \int_{t_1+\delta t_1}^{t_0+\delta t_0} \frac{dt}{a(t)} \tag{3.26}$$

from which it follows that

$$\int_{t_1}^{t_1+\delta t_1} \frac{dt}{a(t)} = \int_{t_0}^{t_0+\delta t_0} \frac{dt}{a(t)}. \tag{3.27}$$

Because $a(t)$ does not change significantly in a single tick δt_1 or δt_0 it follows that

$$\frac{\delta t_1}{a(t_1)} = \frac{\delta t_0}{a(t_0)} \tag{3.28}$$

The time difference between two wave crests is $\delta t = \frac{\lambda}{c}$, with c the same at emission and reception, and thus we confirm that

$$\frac{\lambda_0}{\lambda_1} = \frac{a(t_0)}{a(t_1)}. \tag{3.29}$$

The same result can be derived more formally by considering massless particles in General Relativity (see Baumann's book Sec 2.2). With the scale factor normalized to $a(t_0) = 1$ we find from Eq. (3.20) that

$$1 + z = \frac{1}{a(t_1)} \tag{3.30}$$

Redshifts of galaxies are roughly in the range $0 < z < 10$ (at higher redshifts no galaxies have had time to form since the big bang) and the redshift of the CMB is about $z = 1100$. For example, a galaxy at red shift 2 is observed when the universe was 1/3 of its current size.

3.5 Hubble's law at close distance

To connect to Hubble's law we Taylor expand the scale factor at some time t close to the present time t_0,

$$a(t) = a(t_0) + \dot{a}|_{t=t_0}(t - t_0) + \cdots = a(t_0)[1 + H_0(t - t_0) + \cdots] \tag{3.31}$$

where $H_0 \equiv (\dot{a}/a)|_{t=t_0}$ is the Hubble parameter today, which we call the **Hubble constant** even though it is not constant in time, and $t - t_0$ is called the **lookback time**. The first order Taylor expansion is valid for nearby sources. Using (3.30) and inverting (3.31) using the binomial series $(1 + x)^\alpha \approx 1 + \alpha x$ we find

$$z = H_0(t_0 - t_1) + \ldots. \tag{3.32}$$

In the non-relativistic limit of the Doppler shift of light $z = \sqrt{\frac{1+v/c}{1-v/c}} - 1$ we have $z = \frac{v}{c}$, and this relation can be used at any velocity to define the so called **redshift velocity**. We also define a distance d to the object as $d = c(t_0 - t_1)$. Using these approximations we find the **Hubble-Lemaitre law**

$$v = zc = H_0 d. \tag{3.33}$$

which is the famous linear relation between distance and recession velocity found by Hubble. This relation is valid for $z \ll 1$. In the next section we describe the Hubble law valid at any distance.

3.6 Proper distance, Hubble's law, Hubble distance, Hubble time

So far we have discussed coordinates, now we discuss the so called **proper distance** or **instantaneous physical distance**. This is the distance between two objects at a given fixed time (imagine stopping the expansion of the universe and going there with a meter stick). Recall that in special relativity the proper length of an object is the distance between the two end points in space in a reference frame where both of these are at rest. The proper distance between two spacelike-separated events is the distance between the two events, as measured in an inertial

frame of reference in which the events are simultaneous. Here our frame of reference is the one provided by the global FLRW metric.

The proper distance d_p between the origin and a point at coordinate (r, θ, ϕ) at fixed time t is

$$d_p = \int ds = \int a(t) d\chi = a(t)\chi \tag{3.34}$$

The proper distance, and its time derivative appear in the Hubble law:

$$\dot{dp} = \dot{a}\chi = \frac{\dot{a}}{a} a\chi = H(t) d_p \tag{3.35}$$

The Hubble law in flat space-time discussed above in Eq. 3.7 agrees with this expression.

The **Hubble distance (or Hubble radius)** is defined as the distance where the recession velocity of an object without peculiar velocity becomes equal to the speed of light. This is the case when

$$\dot{d}_p = H_0 d_p = c \tag{3.36}$$

and thus

$$d_H(t_0) = \frac{c}{H_0} \tag{3.37}$$

For $H_0 = 67.8$ km s^{-1} Mpc^{-1} this gives $d_H = 4420$Mpc. This is the distance of galaxies that are currently receding at the speed of light (not at the time when their light was emitted).

Note that the Hubble constant has inverse units of time. Another common definition is thus the **Hubble time**

$$t_H = \frac{1}{H_0} = \frac{a}{\dot{a}} = 4.45 \times 10^{17} s = 14.4 Gyr \tag{3.38}$$

This turns out to be pretty close to the age of the universe, which is somewhat accidental. The Hubble time is the age the universe would have, if the expansion had been linear, which is not the case as we shall soon see.

Measuring the Hubble parameter directly is difficult. One needs to measure both the distance of objects as well as their recession speed, which are not directly observable. A primary tool to do this are SN1a supernovae, but we won't be covering this method this semester. However, the Hubble parameter can also be measured somewhat more indirectly with the CMB and with LSS.

4 Dynamics of the Homogeneous Universe

We have seen that the metric of a homogeneous isotropic universe (one with the same matter content everywhere in space) is the FLRW metric. The free quantity that we need to determine is the evolution of the scale factor $a(t)$, which depends on matter and energy content of the universe. These calculations require both general relativity and thermodynamics. Since this is not the focus of the present course, we will only discuss the results.

4.1 Cosmological fluids and equation of state

According to the cosmological principle, we want to consider homogeneous and isotropic contents of the universe, which are called **cosmological fluids** and specified by:

- **energy density** $\rho(t)$. This has units energy per volume, and thus E^4 in natural units.
- **pressure** $P(t)$. This is the flux of momentum across a surface of unit area (which is equivalent to force per area if there were a wall). The units are also E^4 in natural units. Note that positive pressure leads to gravitational attraction (i.e. wants to contract), not expand as would be the case for a balloon. This is because the kinetic energy of the particles contributes to the positive energy density which attracts gravitationally.

The relation between the energy and pressure $P = P(\rho)$ is called the **equation of state** of the fluid. Note that in GR, both energy density and pressure gravitates, i.e. they are part of the energy momentum tensor. The equation of state is calculated in (relativistic) thermodynamics.

The two main forms of cosmological fluids in the universe are **non-relativistic particles**, also called **dust** or simply **matter** and **relativistic particles** such as photons, also called **radiation**. The cosmological fluid is made up of particles which obey the relativistic relation

$$E^2 = p^2c^2 + m^2c^4 \tag{4.1}$$

The two fluids come from considering this equation in its two limits:

- **Non-Relativistic particles:** $pc \ll mc^2$. Here the energy is dominated by the mass, $E \approx mc^2$, and the velocity of the atoms is $v \approx \frac{p}{m}$. This is true for example for galaxies and galactic dust.
- **Relativistic particles:** $pc \gg mc^2$. Here the energy is dominated by the momentum, $E \approx pc$, and the velocity of the atoms approaches the speed of light $|v| \approx c$. This is true for photons at any time, as well as for massive particles at early times depending on the temperature of the universe. As we will discuss, the temperature in the past was $T = T_0/a(t)$ (where $T_0 \simeq 2.75K$) and the energy of particles in thermodynamic equilibrium is of course $\langle E \rangle = k_B T$.

For non-relativistic particles, the equation of state parameter is

$$w = \frac{P}{\rho} \approx 0 \qquad \text{matter} \tag{4.2}$$

which means that the pressure of matter is negligible compared to the energy density in its mass. For example, even though the pressure of the gas in our atmosphere is not strictly zero, the kinetic energy in the gas molecules is far lower than the energy in their rest mass, and thus the pressure contributes almost nothing to gravity.

On the other hand, for radiation (including photons, relativistic neutrinos and gravitational waves), the equation of state is

$$w = \frac{P}{\rho} = \frac{1}{3} \qquad \text{radiation} \tag{4.3}$$

The factor 1/3 comes from 3-dimensional space.

Finally we need the equation of state parameter of **dark energy**, which is

$$w = \frac{P}{\rho} = -1 \qquad \text{dark energy} \tag{4.4}$$

We'll talk more about this exotic substance later. The key point is that this substance has a negative pressure that leads to gravitational repulsion.

4.2 Solving Einstein's Equation

In this course we will not derive the equations that govern the evolution of the scale factor, which are the continuity equation and the Friedmann equation. These are derived in many textbooks, such as the one from Baumann. One starts with the field equation of general relativity, the **Einstein equation**. It plays a role similar to Maxwell's equations in electrodynamics and is given by

$$G_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu} \tag{4.5}$$

where $G_{\mu\nu}$ is called the Einstein tensor which can be expressed as a function of the metric $g_{\mu\nu}$ and its derivatives and defines the curvature of space.

On the right side of the equation is the **energy-momentum** tensor $T_{\mu\nu}$ (also called the stress-energy tensor). The energy-momentum tensor is the source term for the curvature, similar to electric charge in the Maxwell equations. Both energy density and pressure are sources of curvature.

For a perfect fluid, in the frame of the comoving observer, the energy-momentum tensor is given by

$$T^{\mu}_{\ \nu} = \begin{pmatrix} -\rho c^2 & 0 & 0 & 0 \\ 0 & P & 0 & 0 \\ 0 & 0 & P & 0 \\ 0 & 0 & 0 & P \end{pmatrix} . \tag{4.6}$$

By plugging this tensor in the Einstein equation, one can derive the equation of motion of the scale factor, the Friedmann equation. There is also a conservation law for the energy-momentum tensor given by

$$\nabla_{\mu} T^{\mu}_{\ \nu} = 0 \tag{4.7}$$

where ∇ is the so-called **covariant derivative**. This equation implies the continuity equation we discuss below.

Some books also motivate the Friedmann equation and the continuity equation by arguments from Newtonian physics, to avoid GR. This is done in the book by Huterer and the lecture notes by Tong. I encourage you to check these out.

4.3 Continuity Equation

The first equation of the dynamics of the universe which we will study is the so-called **continuity equation**

$$\dot{\rho} + 3H(\rho + P) = 0 \tag{4.8}$$

This is the expression of energy conservation in a cosmological setting. Note however that energy is a subtle concept in cosmology, due to the broken time translation invariance.

Using the equation of state $P = w\rho$ and assuming a single substance with given w we get

$$\frac{\dot{\rho}}{\rho} = -3(1+w)\frac{\dot{a}}{a} = -3(1+w)H \tag{4.9}$$

To find the relation between ρ and a we can integrate this equation to get:

$$\log\left(\frac{\rho}{\rho_0}\right) = -3(1+w)\log\left(\frac{a}{a_0}\right) \tag{4.10}$$

and thus

$$\rho(a) = \rho_0 a^{-3(1+w)} \tag{4.11}$$

where we've used the fact that $a(t_0) = 1$ and where ρ_0 is the density today (i.e. at $a = 1$).

Using their equation of state we find the following scalings for our three substances:

- **Matter** ($w = 0$):

$$\rho_m \propto \frac{1}{a^3} \tag{4.12}$$

  This is just the dilution with the volume that grows as $V \propto a^3$.

- **Radiation** ($w = 1/3$):

$$\rho_r \propto \frac{1}{a^4} \tag{4.13}$$

  Radiation is not only diluted with the volume but in addition there is a linear redshift effect on the wave length and thus on the energy $E = \frac{hc}{\lambda}$.

- **Dark energy** ($w = -1$):

$$\rho_\Lambda = \text{const.} \tag{4.14}$$

  Dark energy has a constant energy density. It does not dilute with the expansion of space. A universe where $\rho_\Lambda \neq 0$ will always ultimately be dominated by dark energy. There are also more complicated dark energy models where dark energy is not the cosmological constant. In these, the equation of state can deviate from $w = -1$ and can also be time dependent. So far there is no evidence for such models.

The different substances dilute differently with an expanding universe, and thus their mutual importance changes. This is a crucial result in cosmology. In addition, note that total energy is not conserved. This is due to the broken time translation invariance in an expanding universe.

4.4 Friedmann equation

The continuity equation tells us $\rho(a)$ but it is not enough to determine $a(t)$ or $\rho(t)$ for a given collection of homogeneous fluids. For this we need the famous **Friedmann Equation**. The dynamics of the scale factor is dictated by the energy density $\rho(t)$ through the Friedmann equation

$$H^2 \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3c^2}\rho - \frac{kc^2}{R_0^2 a^2} \tag{4.15}$$

where R_0 is the curvature scale, and, as in the FLRW metric, k is either -1, 0, or +1 determining the curvature of space, and G is Newton's gravitational constant given by

$$G \approx 6.67 \times 10^{-11}\ \mathrm{m^3\,kg^{-1}\,s^{-2}} \tag{4.16}$$

The Friedmann Equation, continuity equation, and equation of state together form a closed set of equations that determines the background evolution of the universe.

By taking the time derivative of the Friedmann equation and using the continuity equation one can derive a further useful equation which is called the **acceleration equation** or the **second Friedmann Eq.** or the **Raychaudhuri equation**. It gives the acceleration rate of the scale factor as

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3c^2}(\rho + 3P) \tag{4.17}$$

4.5 Solutions to the Friedmann equation for a single fluid in flat space

The Friedmann equation is easy to solve if we consider a flat universe $k = 0$ with only a single type of fluid. From the continuity Eq. we had

$$\rho(t) = \rho_0 a^{-3(1+w)} \tag{4.18}$$

Plugging this into the Friedmann Eq. for $k = 0$ we get

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3c^2}\frac{\rho_0}{a^{3(1+w)}} = \frac{D^2}{a^{3(1+w)}} \tag{4.19}$$

where we defined the constant $D^2 = \frac{8\pi G\rho_0}{3c^2}$. To solve this equation we take the square root

$$\frac{\dot{a}}{a} = \frac{D}{a^{3/2(1+w)}} \tag{4.20}$$

and then integrate this equation:

$$\int_0^a da'\, a'^{\frac{1}{2}(1+3w)} = D\int_0^t dt' \tag{4.21}$$

where we picked time $t = 0$ to be the time of the big bang where $a(t = 0) = 0$. This leads to

$$\frac{1}{\frac{3}{2}(1+w)} a^{\frac{3}{2}(1+w)} = D\,t \tag{4.22}$$

The common convention is that today at time t_0 the scale factor is $a(t_0) = 1$. This time is given by

$$t_0 = \left(D\frac{3}{2}(1+w) \right)^{-1} \tag{4.23}$$

Plugging this definition of t_0 in our solution we can write it as

$$a(t) = \left(\frac{t}{t_0} \right)^{2/(3+3w)} \tag{4.24}$$

Let's now consider the three types of fluids:

- **Matter**: For $w = 0$ we get

$$a(t) = \left(\frac{t}{t_0} \right)^{2/3} \tag{4.25}$$

  This is known as the Einstein-de Sitter universe. It can be used to approximate our current universe if we neglect dark energy. In this universe the Hubble constant today is

$$H_0 = \frac{\dot{a}}{a}\Big|_{a=1} = \frac{2}{3}\frac{1}{t_0} \tag{4.26}$$

  With $H_0 \simeq 70$ km s^{-1} Mpc^{-1} this gives an age of the universe of

$$t_0 \simeq 10^{10}\text{yrs} \tag{4.27}$$

  It turns out that there are stars that are older than that, which shows that our universe does not contain only matter.

- **Radiation**: For $w = 1/3$ we get

$$a(t) = \left(\frac{t}{t_0} \right)^{1/2} \tag{4.28}$$

  and the relation between H_0 and t_0 is now

$$t_0 = \frac{1}{2}H_0^{-1} \tag{4.29}$$

- **Dark energy**: The dark energy density (or **vacuum energy density**) ρ_Λ is related to the so-called **cosmological constant** by

$$\rho_\Lambda = \frac{\Lambda c^2}{8\pi G} \tag{4.30}$$

  so that the Friedmann equation with $k = 0$ reads

$$H^2 = \frac{\Lambda}{3} \tag{4.31}$$

  and thus

$$\frac{\dot{a}}{a} = \sqrt{\frac{\Lambda}{3}} \tag{4.32}$$

  which is solved by

$$a(t) = A \exp^{\sqrt{\Lambda/3}\,t} \tag{4.33}$$

  This shows that for constant energy density we get exponential expansion. This space-time is called **deSitter space**, and is used in inflation. Note that our previous calculation Eq.(4.24) fails here because there is no time when $a = 0$ (no big bang).

4.6 Critical density

From the Friedmann equation Eq.(4.15), there is a certain density ρ for which the universe would be flat, i.e. $k = 0$:

$$\rho_{\rm crit} = \frac{H^2 3c^2}{8\pi G} \tag{4.34}$$

The Hubble parameter is time dependent so the critical density also varies. Today, the critical energy density is

$$\rho_{\rm crit,0} = \frac{H_0^2 3c^2}{8\pi G} \tag{4.35}$$

and the critical mass density is thus

$$\frac{\rho_{\rm crit,0}}{c^2} \simeq \times 10^{-26} {\rm kg\,m^{-3}} \tag{4.36}$$

which is about one hydrogen atom per cubic meter. The subscript 0 for today is often dropped in the literature.

A very useful and common definition is the density of all fluids together relative to the critical density, called the **density parameter**:

$$\Omega_{TOT}(t) = \frac{\rho_{TOT}(t)}{\rho_{\rm crit}(t)} \tag{4.37}$$

Our constraints on Ω today are roughly $\Omega_{TOT} = 0.999 \pm 0.002$. Of course, curvature could still exist but be smaller than that. Before the discovery of dark energy, several measurements pointed to $\Omega \sim 0.3$.

It's important to note that a flat universe will remain flat forever. To see this we re-write the Friedmann equation as

$$1 - \Omega_{TOT}(t) = -\frac{kc^2}{R_0^2 a^2 H^2} \tag{4.38}$$

from which for $\kappa = 0$ it follows that $\Omega_{TOT}(t) = 1$.

However flatness is dynamically unstable. If the density is just slightly above or below the critical density, the universe will become more curved quickly. This poses the question of why our universe was so flat to begin with that it is still flat today. This can be explained by cosmological inflation as we will discuss later.

4.7 General solution to the Friedmann equation

We now reinstate the curvature term in the Friedmann equation and consider a mix of fluids. The Friedmann equation is then

$$H^2 = \frac{8\pi G}{3c^2} \sum_{w=m,r,\Lambda} \rho_w \; - \frac{kc^2}{R^2 a^2} \tag{4.39}$$

The three fluids have individual density parameters:

$$\Omega_m = \frac{\rho_m(t)}{\rho_{\rm crit}(t)} \qquad \Omega_r = \frac{\rho_r(t)}{\rho_{\rm crit}(t)} \qquad \Omega_\Lambda = \frac{\rho_\Lambda(t)}{\rho_{\rm crit}(t)} \tag{4.40}$$

It is useful to write the curvature on an equal footing as the fluids by defining

$$\rho_k = -\frac{3kc^4}{8\pi G R_0^2 a^2} \tag{4.41}$$

with critical density

$$\Omega_k = \frac{\rho_{k,0}}{\rho_{\text{crit},0}} = -\frac{kc^2}{R_0^2 H_0^2} \tag{4.42}$$

We can then write the Friedmann Eq. as

$$\frac{H^2}{H_0^2} = \frac{\Omega_r}{a^4} + \frac{\Omega_m}{a^3} + \frac{\Omega_k}{a^2} + \Omega_\Lambda \tag{4.43}$$

The Ω are related as

$$\Omega_k = 1 - \Omega_M - \Omega_R - \Omega_{DE} \equiv 1 - \Omega_{TOT} \tag{4.44}$$

Recall that $\Omega_{TOT} > 1$ for the closed universe case and $\Omega_{TOT} < 1$ for the open universe case.

We can now calculate $a(t)$ for an arbitrary universe. Taking the square root of the Friedmann equation

$$\frac{da}{adt} = H_0 \sqrt{\Omega_r a^{-4} + \Omega_m a^{-3} + \Omega_k a^{-2} + \Omega_\Lambda} \tag{4.45}$$

which we can integrate to obtain

$$t(a) = H_0^{-1} \int_0^a \frac{da'}{\sqrt{\Omega_r a'^{-2} + \Omega_m a'^{-1} + \Omega_k + \Omega_\Lambda a'^2}} \tag{4.46}$$

which can be evaluated numerically. The age of the universe today is given by setting $a = 1$.

Note that Ω_m, Ω_r and Ω_Λ change over time, but in a flat universe $\Omega_{TOT} = 1$ does not change change, as we saw above. In the same way a closed universe stays closed and an open universe stays open (Ω_{TOT} changes but not its sign).

4.8 Lambda-CDM and its parameters

The main success of cosmology is to establish the so-called **standard model of cosmology**, or **Lambda-CDM** model (we will also write ΛCDM). This model fits an amazing amount of different cosmological observations with stunning efficiency. In fact, only 6 parameters are required to fit all this cosmological data (we don't include known physical constants such as masses of known particles in the counting). The Lambda-CDM ("Lambda cold dark matter") model has three components: (cold, non-relativistic) matter, radiation and the dark energy. We have already met 3 of the 6 ΛCDM parameters. Their current best fit values from the Planck CMB satellite are

- **Matter density** $\Omega_m = 0.310 \pm 0.007$. This is the combined density of cold dark matter and baryons.
- **Baryon density** $\Omega_B h^2 = 0.0224 \pm 0.0002$ (i.e. $\Omega_B \approx 0.05$). Baryons are a part of matter (the rest being dark matter $\Omega_{CDM} \approx 0.26$). We need both of these components to fit observations for reasons that we will discuss later.

- **Hubble constant** $H_0 = (67.9 \pm 0.7)$ km s^{-1} Mpc^{-1}. There is currently a famous $\sim 5\sigma$ disagreement of different measurements called the **Hubble tension** which we will talk about later.

In addition, spatial curvature is not detected, $\Omega_k = 0.001 \pm 0.002$. Radiation Ω_r is negligible today. The current density in photons is about $\Omega_\gamma = 5 \times 10^{-5}$. Neutrinos today are not relativistic anymore and their density is about $\Omega_\nu = 3.4 \times 10^{-5}$. The density of dark energy of $\Omega_\Lambda = 0.6847 \pm 0.0073$ follows from the other values. Note that the selection of 3 parameters is not unique, for example one could switch Ω_Λ and H_0 or replace H_0 by the age of the universe.

The above parameters define the background expansion of the universe. Another two parameters of ΛCDM are not about the background expansion but rather about the small inhomogeneities (perturbations) that seeded structure formation at the beginning of the universe. They are called

- **amplitude of primordial perturbations** $\boldsymbol{A_s}$ and
- **spectral index of primordial perturbations** $\boldsymbol{n_s}$.

We'll discuss these in the section about inflation. Finally a sixth parameter of ΛCDM is the so-called

- **Optical depth** τ. This parameter describes how transparent the universe is for CMB light. Physically it depends on how many free electrons there are, which depends on the process of reionization.

This parameter is required to fit CMB data.

Much of this course will be about how these and other parameters can be measured with data from large-scale structure and the CMB. Other parameters that we routinely fit to data either have a standard model expectation (such as N_{eff}) or are currently compatible with zero (such as the tensor-to-scalar ratio r) and are thus not included in the counting of 6 Lambda-CDM parameters. There is of course no guarantee that 6 parameters are sufficient to fit any future data, and much of cosmology, as in particle physics, is about searching for physics beyond the standard model (of cosmology).

4.9 Matter-Radiation Equality

Over much of the history of the universe one component of the fluids dominated over the others. As we have seen from the Ω values today, currently we are in a Λ dominated era but matter is not yet negligible, while radiation is four orders of magnitude smaller. From the scaling $\rho_m \propto a^{-3}$ and $\rho_r \propto a^{-4}$ it is clear that at earlier times matter and radiation dominated over dark energy, and that at the earliest times radiation dominated over the other two components. We now want to determine the time of matter-radiation equality.

It turns out that neutrinos were relativistic at the time of matter-radiation equality. Thus we include them in the radiation density

$$\Omega_r = \Omega_\gamma + \Omega_\nu \approx 8.4 \times 10^{-5}, \tag{4.47}$$

To find the redshift where matter and radiation were equal, we equate their densities:

$$\begin{aligned}
\rho_m(z_{eq}) &= \rho_r(z_{eq}) \\
\rho_{m,0}(1+z_{eq})^3 &= \rho_{r,0}(1+z_{eq})^4 \\
\rho_{\text{crit},0}\Omega_m(1+z_{eq})^3 &= \rho_{\text{crit},0}\Omega_r(1+z_{eq})^4 \\
z_{eq} &= \frac{\Omega_M}{\Omega_R} - 1 \approx 3250
\end{aligned}$$

Therefore the equality of the two happened when the scale factor was about 3000 times smaller than today. Using $a(t) = (t/t_0)^{2/3}$, valid during matter domination, gives $t_{eq} = 70.000$ yrs. A more accurate calculation using Eq.(4.46) gives $t_{eq} = 50.000$ yrs.

The growth of perturbations (such as those in CMB) depends sensitively on which component is dominating the universe. Radiation pressure suppresses structure growth. We will study this topic later.

We may also ask about matter-dark energy equality. Using a similar calculation one finds that this happened about 4 billion years ago, which is relatively recently in cosmological terms.

4.10 Numerical examples

To understand the expansion of the universe better, I recommend that you make some plots of various quantities. This can be done for example using the Boltzmann solvers CAMB or CLASS, or the library AstroPy. Here are some ideas what to plot:

- Evolution of the different fluids. Plot Ω_X for m, r, Λ as well as their sum as a function of the scale factor from $a = 10^{-5}$ to $a = 100$ on a log-x plot. Do the same as a function of time.
- Evolution of $H(a)$ and $aH(a)$. Plot these functions for the same range of the scale factor. Mark the radiation, matter and DE dominated regions. Do the same as a function of log time and linear time.
- Evolution of the scale factor $a(t)$.

5 Early Universe Thermodynamics

At early times the universe is well approximated by a **hot fluid in thermal equilibrium**. The further we go back in time, the higher the temperature of the universe. This section covers the so called **hot big bang**. Prior to that there was the period of inflation which we cover next.

As the universe gets hotter (i.e. we go backwards in time), the properties of the fluid and the relevant physical particles change dramatically. For example, at some point the universe was too hot to form atoms, nuclei and even nucleons, since their collisions would rip them apart immediately. Up to temperatures that we have probed with particle colliders ($\sim$ 1 TeV), the fundamental physics is known through the standard model, and we can thus in principle calculate what happens (the strong force is difficult to calculate because of the strong coupling, and even for the other forces some precision calculations are still being improved). These calculations

generally match observations very well. Beyond that energy scale, theorists have come up with different models. Of course, this is an opportunity to probe physics beyond the standard model.

I want to give only a very brief overview of this material. While important, these topics have been worked out in detail and have been put into code packages that can be used without detailed understanding for most practical purposes.

5.1 Overview

We can get an idea of the hot big bang simply from the following facts:

- The temperature in the past was $T = T_0/a(t)$. This is because for radiation (which dominates the thermodynamics even in the matter dominated universe) we have $\lambda \propto a$ as we have derived and the temperature of a black body scales as $\lambda_{\text{peak}} = \frac{\text{const.}}{T}$ (Wien's law). We have already calculated $a(t)$.
- The average energy of a particle in thermal equilibrium is $\langle E \rangle = k_B T$ (up to a constant factor counting degrees of freedom).
- The temperature of the universe now is $T_0 \simeq 2.75K$, which is the temperature of the Cosmic Microwave Background. A clump of gas without any source of heat (also no gravitational heating) will be at this temperature.
- We know the masses of particles and binding energies of bound states. For example the binding energy of electrons in atoms is of order eV. If the kinetic energy exceeds the binding energy, the bound state will be broken up. The binding energy of nuclei is of order 1 MeV and the binding energy of nucleons is of order 1 GeV. This tells us very roughly at what temperature these bounds states form. The actual temperatures are lower because of the tail of the Boltzmann distribution.

These facts correctly suggest a thermal history that is illustrated in Fig. 2. The key events in the thermal history of the universe are also listed in table 1. To understand it in more detail we need to review some thermodynamics.

5.2 Statistical mechanics description of the universe

In the very early universe (after inflation), the rate of interactions in the primordial plasma was very high and the universe was in a state of thermal equilibrium. This simple state is the beginning of the so-called hot big bang. Later, some particles drop out of thermal equilibrium, and we will need the Boltzmann equation to describe them.

To define the state of matter in statistical mechanics, we use the **phase-space distribution function** $f(\mathbf{x}, \mathbf{p}, t)$. It gives the probability density that a particle is found at a particular position x with momentum p at time t. That is, the number of particles N in a phase space element is is given by

$$N(\mathbf{x}, \mathbf{p}, t) = f(\mathbf{x}, \mathbf{p}, t) \times \frac{(\Delta x)^3 (\Delta p)^3}{(2\pi)^3} \tag{5.1}$$

In the present unit we are mostly concerned with homogeneous and isotropic matter, and thus the phase space distribution function does not depend on position and only on the magnitude

Event	**Temperature**	**Energy**	**Time**
Inflation	$< 10^{28}$K	$< 10^{16}$GeV	$> 10^{-34}$s
Dark matter decouples	?	?	?
Baryogenesis (matter-antimatter asymmetry, GUT?, quark-gluon plasma)	?	?	?
EW phase transition (symmetry breaking due to Higgs)	10^{15}K	100GeV	10^{-11}s
Hadrons form (protons, neutrons) from quark-gluon plasma. QCD phase transition.	10^{12}K	150MeV	10^{-5}s
Neutrinos decouple (weak interaction)	10^{10}K	1MeV	1s
Big Bang Nucleosynthesis BBN: sets element abundances	10^{9}K	100keV	200s
Atoms form (helium, hydrogen)	3400K	0.30eV	$260,000$yrs
Photons decouple (transparent universe)	2900K	0.25eV	$380,000$yrs
First stars	50K	4meV	100million yrs
First galaxies	20K	1.7meV	1billion yrs
Dark energy	3.8K	0.33meV	9billion yrs
Today	2.7K	0.24meV	13.8billion yrs

Table 1. Key events in the histroy of the universe (adapted from Baumann table 1.2)

of momentum: $f(p,t)$. To treat inhomogeneities we will later need the full $f(\mathbf{x},\mathbf{p},t)$ and its differential equation, the Boltzmann equation.

5.3 Thermal equilibrium

5.3.1 Thermal equilibrium vs decoupling

The basis to understand the early universe is **thermal equilibrium**. To be in thermal equilibrium, particles need to interact and we must have waited for long enough so that their thermal bath has become uniform. For example, a typical situation would be a particle X that can annihilate with its anti-particle $\bar{X}$ into 2 photons, and in return 2 photons can pair-create an $X+\bar{X}$ pair:

$$X + \bar{X} \rightleftharpoons 2\gamma \tag{5.2}$$

The interaction rate Γ (per particle) for this process is

$$\Gamma = n\sigma v,$$

with n the number density of particles, σ their cross-section, and v their velocity (all three are in general a function of temperature). Note that this is the interaction rate per particle (which is why it is linear in n), not the total interaction rate per volume. Γ has units of inverse time. In an expanding universe, it turns out (from the Boltzmann equation) that particles can be in thermal equilibrium if $\Gamma \gg H$, that is the interaction rate is much larger than the Hubble rate. To understand this better remember that the age of the universe is roughly $t_{\rm age} \simeq H^{-1}$, and we want the typical interaction time to be much smaller than the age of the universe. To

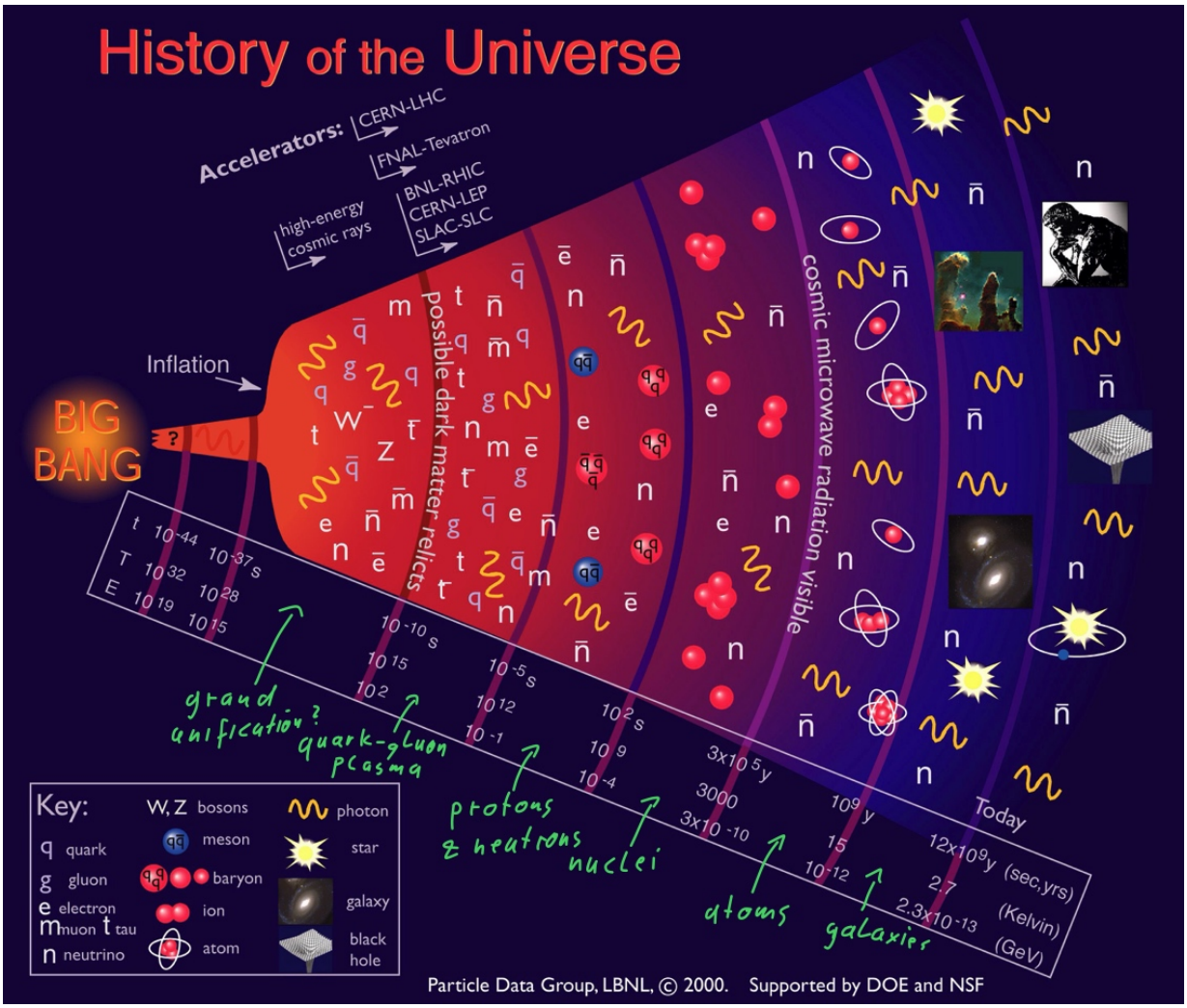


Figure 2. Thermal history of the universe (plot from the Particle Data Group).

summarize, particles fall out of thermal equilibrium when their interaction rate drops below the Hubble expansion rate of the universe. At that moment, the particles stop interacting with the rest of the thermal bath, which is called **decoupling**, and a **relic abundance** is created. Both the creation and the annihilation of such relic particles is negligibly small after decoupling.

We will first assume that we are in thermal equilibrium, but later discuss beyond equilibrium phenomena.

5.3.2 Basics of equilibrium thermodynamics

To simplify notation we will be using natural units with $c = 1$, $\hbar = 1$ and $k_B = 1$ (see e.g. Baumann Appendix C or Huterer 1.7). Particles in thermal equilibrium are either **Bose-Einstein distributed (bosons)** or **Fermi-Dirac distributed (fermions)**. The distribution function is given by

$$f(p,T) = \frac{1}{e^{(E(p)-\mu)/T} \pm 1} \tag{5.3}$$

where the $-$ sign is for bosons and the $+$ sign is for fermions. It gives the probability that a particle chosen at random has the momentum p. The distributions depend on the **temperature** $\boldsymbol{T}$ and the **chemical potential** $\boldsymbol{\mu}$ (which can depend on temperature and thus on time in an expanding universe). The chemical potential describes the response of a system to a change in particle numbers. Since for photons $\mu = 0$ and for particle-antiparticle pairs $\mu_X = -\mu_{\bar{X}}$ we can ignore the chemical potential in much of the following discussion. The chemical potential is important when the particle number changes, for example during recombination where the number of free electrons changes.

From the distribution functions we can calculate the important thermodynamic quantities.

- The **number density of particles** is

$$n(T) = \int \frac{g}{(2\pi)^3} d^3p f(p,T) \tag{5.4}$$

  where g is the number of internal degrees of freedom of the particle (e.g. number of spin states).

- The **energy density** is

$$\rho(T) = \frac{g}{(2\pi)^3} \int d^3p f(p,T) E(p) \tag{5.5}$$

- The **pressure** is

$$P(T) = \frac{g}{(2\pi)^3} \int d^3p f(p,T) \frac{p^2}{3E(p)} \tag{5.6}$$

  where $E(p)$ is the relativistic energy of the particles. Note that in the ultra-relativistic case $E = p$ and thus $P = \frac{1}{3}\rho$ as expected.

In thermal equilibrium we can have several different particle species with masses m_i and chemical potential μ_i but at the same temperature T.

5.3.3 Number density and energy density of particles

As we discussed, for our purpose we can set the chemical potential to zero. The integrals above are evaluated for example in Baumann's book.

In the **relativistic limit** one gets that

$$n = \frac{\zeta(3)}{\pi^2} g T^3 \times \begin{cases} 1 \text{ (bosons)} \\ \frac{3}{4} \text{ (fermions)} \end{cases} \tag{5.7}$$

where $\zeta(3) \approx 1.202$ is the Riemann zeta function, and for the energy density

$$\rho = \frac{\pi^2}{30} g T^4 \times \begin{cases} 1 \text{ (bosons)} \\ \frac{7}{8} \text{ (fermions)} \end{cases} . \tag{5.8}$$

Note the scaling with temperature and the fact that bosons and fermions only differ in a constant factor. A typical use of this result is to calculate the number density and energy density of photons

today, given the observed temperature of the CMB, $T_0 \approx 2.73$ K:

$$n_{\gamma,0} = \frac{2\zeta(3)}{\pi^2} T_0^3 \approx 410 \text{ photons cm}^{-3},$$
$$\rho_{\gamma,0} = \frac{\pi^2}{15} T_0^4 \approx 4.6 \times 10^{-34} \text{g cm}^{-3}.$$

In terms of the critical density, the photon energy density is then

$$\Omega_\gamma h^2 \approx 2.5 \times 10^{-5}. \tag{5.9}$$

as mentioned earlier.

For **non-relativisitic particles** ($m \gg T$), the result for bosons and fermions is the same. The integral gives

$$n = g \left(\frac{mT}{2\pi}\right)^{3/2} e^{-\frac{m}{T}} \tag{5.10}$$

The exponential suppression is called **Boltzmann suppression**. Physically it means that particles and anti-particles still annihilate when the temperature becomes low, but they are no longer created in pair production. This means that when the temperature of the universe falls below the particle mass and the particle is still in thermal equilibrium (i.e. $\Gamma \gg H$), then the particle's abundance and energy density drop rapidly.

For non-relativistic particles one also gets

$$\rho = mn \tag{5.11}$$

and

$$P = nT \tag{5.12}$$

which is the ideal gas law $PV = Nk_bT$ with $k_B = 1$ and thus $P \sim 0$ since $T \ll m$.

5.3.4 Effective degrees of freedom

The last aspect of thermal equilibrium that we want to mention is the **effective number of (relativistic) degrees of freedom**. A general formula for the energy density that includes all species/particles is (generalizing Eq.5.8):

$$\rho_R \equiv \sum_i \rho_i = \frac{\pi^2}{30} g^* T^4. \tag{5.13}$$

where the parameter g^*, called the effective number of relativistic degrees of freedom, is a weighted sum of the multiplicity factors of all particles. This factor is defined as

$$g^*(T) = \sum_{i \in \text{bosons}} g_i \left(\frac{T_i}{T}\right)^4 + \frac{7}{8} \sum_{i \in \text{fermions}} g_i \left(\frac{T_i}{T}\right)^4. \tag{5.14}$$

Here we are allowing the possibility that the species have a different temperature T_i from the photon temperature T, hence the power-law factors in this definition of g^*. The effective number of relativistc degrees of freedom are plotted in Fig.3. Above about 100GeV all particles of the

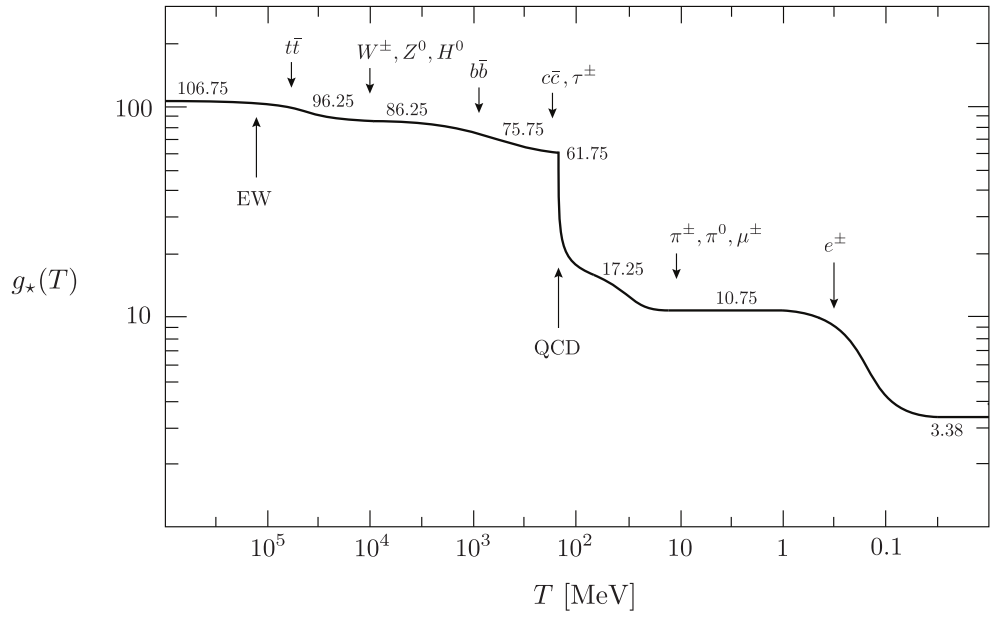


Figure 3. Evolution of effective number of relativistic degrees of freedom assuming the Standard Model particle content. The EW and QCD phase transitions are also indicated. From Daniel Baumann's cosmology lectures.

standard model are relativistic. Considering all quarks, leptons, gauge bosons, gluons and the Higgs (with their helicity or spin and their anti-particles) this adds up to $g^* = 106.75$ The fractional number is possible due to the 7/8 prefactor. As the universe cools down, the heavier particles drop out of this sum. In the end we are left with 3.38 relativistic degrees of freedom. Of these, 2 are for the photon (2 polarisations) and the rest is for neutrinos. The counting of relativistic degrees of freedom for neutrinos is subtle. In particular the neutrino temperature today (~ 1.9 K) is not the same as that of the photons (~ 2.7 K) because they decouple from the thermal bath before electrons and protons annihilate (which heats the thermal bath). Also today neutrinos are not relativistic anymore.

Several physical observables are sensitive to the effective number of relativistic species, and this is an avenue to detect new physics. We usually parametrise the "extra" degrees of freedom by the **effective number of neutrino species N_{eff}**. Constraints on N_{eff} come from

- Element abundances: BBN is sensitive to the expansion rate, which is sensitive to N_{eff}.
- The CMB power spectrum and so-called CMB spectral distortions also constrain N_{eff} at a later time.

The constraint from Planck is $N_{\text{eff}} = 2.99 \pm 0.34$. The theory expectation from the standard model is not exactly 3 but rather 3.046 which is due to the fact that neutrinos deviate a bit from a Fermi-Dirac distribution due to the energy dependence of the weak interaction.

5.4 Beyond Equilibrium: The integrated Boltzmann equation

To study processes that are not in thermal equilibrium we need the Boltzmann equation. The (integrated) Boltzmann equation for a homogeneous particle species n_i is given in general by:

$$\frac{1}{a^3}\frac{d(n_i a^3)}{dt} = C_i[\{n_j\}] \tag{5.15}$$

The left hand side is just the conservation of particle number if the right hand side is zero. The right hand side is the **collision term** that describes the interaction with all other particle species n_j. The collision term includes **cross-sections between particles**, which is where the standard model of particle physics and QFT scattering amplitude calculations come in. Solving this equation goes beyond this course material.

We want to again point out one important non-equilibrium phenomenon: the **freeze out**. The terms decoupling and freeze-out are closely related. Decoupling means that interactions effectively stop and freeze-out means the creation of a relic density. Above we found that when the temperature of the universe falls below the particle mass (thus we are in the non-relativistic regime) and the particle is still in thermal equilibrium, then the particle's abundance and energy density are exponentially suppressed in m/T. We have also discussed that for the particles to be in thermal equilibrium, we need their interaction rate to be larger than the expansion rate $\Gamma \gg H$. However if the particle drops out of thermal equilibrium before the Boltzmann suppression kicks in, i.e. $\Gamma < H$, we say that it "freezes out". In this case some **relic density** of the massive particle remains which is constant in comoving volume and does not change (unless the particle decays, such as neutrons).

If we had time to study the integrated Boltzmann equation in more detail, we would in particular examine:

- The formation of the light elements during the Big Bang nucleosynthesis (BBN). This is one of the big successes of the standard model of cosmology, making predictions for the abundance of elements that agree very well with data (with the possible exception of the lithium problem).
- The production of dark matter (which likely has a relic density set in the early universe).
- The decoupling of neutrinos in the early universe, when the weak interaction becomes too weak to couple them to the thermal bath.
- The neutron freeze-out which set the initial neutron to proton ratio. Remember that free neutrons decay.
- The period of recombination where electrons and nuclei form neutral atoms and the universe becomes transparent.
- Baryogenesis, the somewhat unknown process that led to the matter-antimatter asymmetry observed today.

5.5 Beyond Homogeneity: The Einstein-Boltzmann equations

So far in this unit we have been considering the homogeneous universe. Of course, the universe is only interesting because it is not homogeneous. Here I want to outline what the full set of equations are that govern the universe, without assuming homogeneity. The relevant equations with inhomogeneity are still the Einstein equation and the Boltzmann equation, which of course are coupled to eachother. The Boltzmann equation for an inhomogeneous anisotropic fluid with phase space density $f(\mathbf{x}, \mathbf{p}, t)$ is schematically

$$\frac{df_a(\mathbf{x}, \mathbf{p}, t)}{dt} = C[\{f_b(\mathbf{x}, \mathbf{p}, t)\}], \tag{5.16}$$

where f_a is the phase space density of particle a and we have other particles $\{f_b\}$ which also have Boltzmann Equations. Note that unlike the integrated Boltzmann equation for number densities we saw above, this is an equation for the full phase space density. The phase space density goes into the energy momentum tensor of Einsteins equation. For completeness, the energy momentum tensor for a given phase space distribution function $f(x, p, t)$ is

$$T^{\mu}_{\nu}(x, t) = \frac{g}{-\det[g_{\alpha\beta}]} \int \frac{dP_1 dP_2 dP_3}{(2\pi)^3} \frac{P^{\mu} P^{\nu}}{P^0} f(\mathbf{x}, \mathbf{p}, t)$$

(Dodelson Eq 3.20) where the degeneracy factor g counts the internal states. The EM tensor of course defines the metric through the Einstein equation

$$G_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu} \tag{5.17}$$

These equations can be solved analytically in **relativistic cosmological perturbation theory**. To do so, one expands the metric in perturbations around FLRW, and the particle content in perturbations around the average density. Relativistic cosmological perturbation theory is in particular required to calculate the Cosmic Microwave Background. In fact, the full Boltzmann equation is not required for an analytic treatment, since one can make a fluid approximation on large enough scales. However, on scales where the mean free path length of the photon becomes important (during recombination) such an approximation must break down. In practice, we thus solve the Einstein-Boltzmann equations of the early universe numerically, with codes such as CAMB and CLASS. Even the numerical solution starts with a perturbative ansatz, which gives the **linearized Einstein-Boltzmann equations**. As we will discuss more, linear perturbation theory is enough to calculate the CMB to excellent precision.

On the other hand, in the late universe, perturbations are non-linear. This is the domain of structure formation. Fortunately, in this domain relativistic effects are small and we can work with **Newtonian perturbation theory** and **Newtonian simulations**. In summary, in cosmology one very rarely needs perturbation theory that is both (general) relativistic and non-linear. We will learn much more about perturbations in the CMB and in large-scale structure during this course.

6 Inflation

To complete our overview of the evolution of the universe, we need to discuss the earliest (highest energy) epoch of the universe which we can currently understand, the period of **cosmological**

inflation. Unlike the hot big bang, inflation is still somewhat speculative. It makes predictions that we can verify with observations, but these predictions are not so unique that we would consider the theory to be proven. There is however, in the opinion of most (but not all) cosmologists, no competitive theory that would be equally attractive. In fact, inflation is such a good framework to set up the initial conditions of the universe that it is almost treated as a fact by many cosmologists. It does in particular the following things for us:

- It **makes the universe flat** even if it started out not being flat. There is some debate and ongoing research about this question (e.g. does inflation even start in an inhomogenous universe), but the majority opinion seems to be that it works.
- It **solves the horizon problem**, which is that we find thremal equilibrium and correlations between parts of the universe that would not be causally connected without inflation.
- It sets up the **horizon exit for the later re-entry of perturbations**, which is crucial to explain the matter and CMB power spectrum (turnover and BAO phases).
- It **gives a natural mechanism to generate the initial inhomogeneities** of the universe from quantum perturbations, and provides a framework to calculate their statistical properties.
- It solves the so-called magnetic monopole problem, which depends on a speculative GUT theory and may thus not exist. We will not cover this problem.

In my opinion it is near certain that accelerated expansion and a quantum origin of primordial perturbations really happened in the universe. Whether this happened through a weakly coupled slowly rolling scalar field, as is the case in most inflation models, is less certain. Whether anything happened "before inflation" and whether this question makes sense is not known, and likely won't be answered before we have a complete non-perturbative theory of quantum gravity.

Inflation is treated beautifully in all the references that we pointed out for this unit, and I will compress the material very significantly.

6.1 The flatness problem

We have already developed all the tools to understand the flatness problem. We know from experimental data that $|\Omega_k| < 0.01$. We also know that curvature scales as $\rho_k \propto \frac{1}{a^2}$, matter scales as $\rho_m \propto \frac{1}{a^3}$ and radiation $\rho_r \propto \frac{1}{a^4}$. From this it is easy to estimate how flat the universe must have been at the earliest time that we trust our understanding of physics. If you chose this time to be the electroweak phase transition (100 GeV, $z = 10^{15}$), one finds that $|\Omega_k(t_{EW})| < 10^{-30}$ (see e.g. Tong's lecture notes Sec. 1.5.1 for this calculation). The flatness problem is the question why the universe was so flat to begin with. We would like a physical theory that sets up the initial conditions of the universe so that it is "naturally" very flat. Inflation does this for us.

6.2 The horizon problem

The CMB has the same temperature in all directions of the sky. This suggests that all parts of the CMB sky should have been in thermal equilibrium prior to recombination, which requires

causal contact at that time. More than that, as we will see, different parts of the CMB sky have small temperature perturbations which are correlated. To establish a correlation, clearly causal contact is also required. It turns out that in the hot big bang picture which we developed so far, parts of the CMB that are further than about 1 degree apart in angle were not in causal contact prior to recombination without inflation. This is the horizon problem which inflation solves. Let's now put this into equations.

6.2.1 Comoving and physical particle horizon

If we are sitting at a point in comoving coordinates, say at $\mathbf{x} = 0$, at a time t, the comoving distance from which we can receive light is limited by the speed of light and age of the universe. This light cone also limits the patch of the universe that can causally influence us. This finite distance is called the **comoving particle horizon**. To evaluate it, we start with the metric in the form Eq.(3.18) or Eq.(3.19) for a radial trajectory:

$$ds^2 = a^2(\eta)\left[-c^2 d\eta^2 + d\chi^2\right] \tag{6.1}$$

$$ds^2 = -c^2 dt^2 + a^2(t) d\chi^2 \tag{6.2}$$

We then integrate this equation for a light ray $ds = 0$. If the Big Bang "started" with the singularity at $t_i \equiv 0$ then comoving particle horizon at time t is:

$$d_h^{comov}(t) \equiv \chi = c\int_0^t \frac{dt'}{a(t')} = c(\eta - \eta_i) \tag{6.3}$$

where the h stands for "horizon". Recall our definition of conformal time $\eta = \int dt/a(t)$. For a scale factor that goes to zero at $t_{BB} = 0$ (i.e. matter or radiation dominated) we can set $\eta_i = 0$ and the comoving horizon is just η. The size of the **physical particle horizon** is (using Eq.(3.34))

$$d_h^{phys}(t) = a(t) d_h^{comov}(t) = a(t) c \int_{t_{BB}}^t \frac{dt'}{a(t')} \tag{6.4}$$

The particle horizon can be nicely illustrated in a spacetime diagram. To understand that, we start from the metric Eq.(6.1). We see that a light ray is given by $\chi(\eta) = \pm c\,\eta + \text{const.}$ and thus can be drawn as a 45° angle in the χ-$c\eta$ plane. The resulting diagram is called a **spacetime diagram**. The spacetime diagram for the particle horizon is shown in Fig. 4. One often also defines the **event horizon** which is the forward (rather than backwards) light-cone and tells us what events we can influence in the future.

In a flat universe with $a(t) = \left(\frac{t}{t_0}\right)^{\frac{2}{3(1+w)}}$ and $w = \text{const.}$, the physical particle horizon today is:

$$d_{\rm h}^{phys} = \frac{2}{1+3w} H_0^{-1}. \tag{6.5}$$

which is equal to H_0^{-1} up to an order 1 factor.

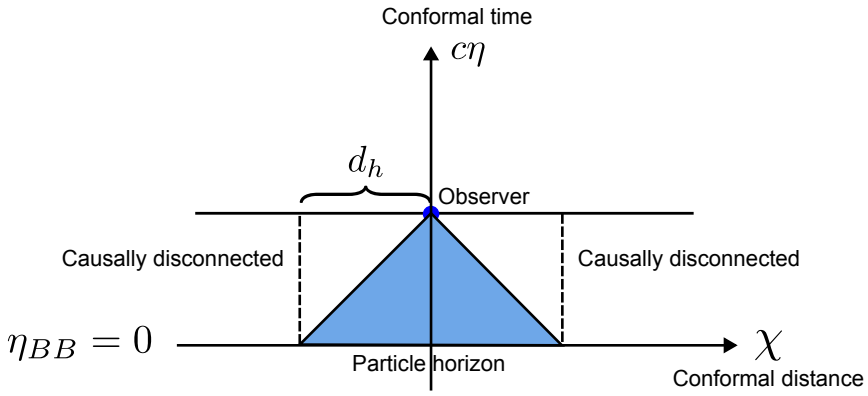


Figure 4. Spacetime diagram of the particle horizon for a given observer.

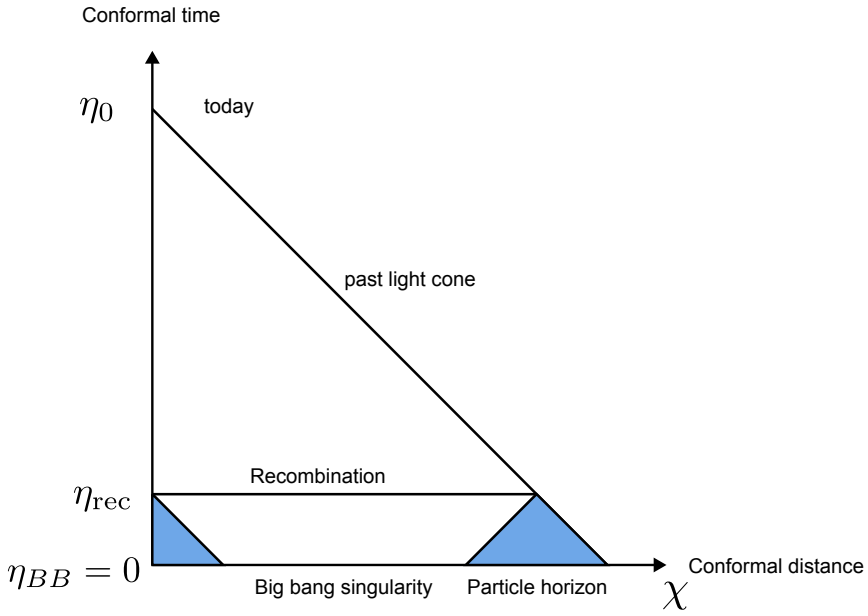


Figure 5. Particle horizon for CMB perturbations without inflation. CMB regions we see today in the sky were causally disconnected in the past.

6.2.2 Particle horizon and the CMB

It turns out that the particle horizon at the time of the CMB was much smaller than our particle horizon today. This is illustrated in Fig.5. The CMB is a particularly clear example, but the horizon problem exists also for the matter and galaxy distribution.

Let's calculate the size of the (physical) particle horizon at recombination and today. For a purely matter-dominated universe (radiation does not substantially change the conclusion), with

$$a(t) = \left(\frac{t}{t_0}\right)^{\frac{2}{3}}, \tag{6.6}$$

the particle horizon at time t is defined by

$$d_h(t) = c\,a(t)\int_0^t \frac{dt'}{a(t')} = 3ct \tag{6.7}$$

Let's write this in terms of red shift:

$$d_h(t) = 3\,c\,t = 3\,c\,a^{3/2}\,t_0 = 3\,c\,(1+z)^{-3/2}\,t_0 = 2\,c\,(1+z)^{-3/2}\,H_0^{-1} \tag{6.8}$$

where we used $1+z = \frac{1}{a}$ and $H_0 = \frac{2}{3t_0}$.

We already discusseed that recombination happened around $z \approx 1100$. We would like to know how large the particle horizon at recombination is in todays physical distance. From recombination to today, the distance scale $d_h(z)$ has been stretched by the expansion of the universe to $\frac{a_0}{a(t)}d_h(z) = (1+z)d_h(z)$.

We can compare this to the particle horizon today, which is

$$d_H(t_0) = \frac{2c}{H_0}. \tag{6.9}$$

The distance $d_H(z)$ today subtends an angle on the sky that's given by the fraction of these two distances:

$$\theta \approx \frac{(1+z)d_H(z)}{d_H(t_0)} \approx \sqrt{\frac{1}{1100}} \approx 0.03 \text{ rad} \implies \theta \approx 1.7^\circ. \tag{6.10}$$

Thus, assuming a matter distributed universe, patches of the sky separated by more than $\approx 1.7^\circ$ had no causal contact at the time the CMB was formed. Radiation does not change this estimate substantially.

6.3 Inflationary expansion

Both the horizon problem and the flatness problem can be solved by a sufficiently long time of **accelerated expansion** prior to the hot big bang. Let's see why this happens.

An accelerating phase ($\ddot{a} > 0$), assuming power-law expansion for now, is given by

$$a(t) \sim t^n \tag{6.11}$$

with $n > 1$. Alternatively, we could have exponential expansion

$$a(t) \sim e^{H_{\text{inf}}t} \tag{6.12}$$

with constant H_{inf} which also accelerates.

The comoving particle horizon is

$$d_h(t) = c\int_0^t \frac{dt'}{a(t')} \tag{6.13}$$

It is finite only if the integral converges. This was the case for a matter (or radiation) dominated universe, as we saw in above. But, for $a(t) \sim t^n$ we have

$$\int_0^t \frac{dt'}{t'^n} \to \infty \tag{6.14}$$

if $n > 1$. As t' approaches 0 we get more and more contributions to d_h so it diverges. Recall from Eq.(6.3) that $d_h = c(\eta - \eta_i)$. We see that an early accelerating phase buys us conformal time and allows all regions of the universe to have been in causal contact in the inflationary past. For inflation, the natural choice is to use time coordinates so that inflation starts at $\eta_i = -\infty$ (since the lower end of the integral leads to the divergence) and ends at conformal time $\eta_{\rm f} = 0$. Patching the spacetime of inflation to the spacetime of the later universe together we get Fig. 6, which shows that light cones now overlap. For accelerated power law expansion there is still a big bang at $t = 0$ where $a(t) = 0$, but there is an infinite amount of conformal time after that. However we should not expect that our equations hold for $a(t)$ smaller than the Planck length, since we don't know non-perturbative quantum gravity.

Inflation models naturally generate exponential expansion rather than power law expansion, i.e.

$$a(t) \propto e^{H_{\rm inf} t} \tag{6.15}$$

Unlike the power law acceleration we considered above, for exponential expansion there is no big bang in the past since $a(t) > 0$ at all times. This means that there is no natural choice for $t = 0$ and our time integral can go from $t_i = -\infty$ to t_f, which again makes the comoving horizon diverge, if exponential expansion went on infinitely long. The Hubble parameter has dimension of energy (or inverse time) so the exponent is dimensionless as it should be. Exponential expansion means that a patch of space time of physical size d_i grows to a size $d_f = d_i\, e^{H_{\rm inf} T}$ in time T. We define the **numer of e-folds of inflation** by $N = H_{\rm inf} T$.

Let's estimate how many e-folds we need to solve the horizon problem, by inflating a causally connected patch before inflation to the size of our current universe. Before inflation started, the physical particle horizon had some value d_i that was causally connected. A natural scale is the physical Hubble distance Eq.(3.37) $d_i = c H_{\rm inf}^{-1}$. After exponential expansion, this connected patch has the physical size $d_f = e^N d_i$. Then, due to the expansion of the universe since the end of inflation, the patch grows to

$$d_{now} = \frac{d_f}{a_{\rm inf}} = \frac{e^N d_i}{a_{\rm inf}} = \frac{e^N c}{H_{\rm inf}\, a_{\rm inf}} \tag{6.16}$$

where $a_{\rm inf}$ is the scale factor at the end of inflation, the beginning of the ordinary evolution of the universe. We want d_{now} to be much larger than the Hubble horizon today, i.e. $d_{now} \gg c H_0^{-1}$. It thus follows that

$$e^N > \frac{H_{\rm inf}}{H_0} a_{\rm inf} \tag{6.17}$$

Most of the relative expansion since the end of inflation happened during the radiation era, in which $H \propto \frac{1}{a^2}$. Thus we have $\frac{H_{\rm inf}}{H_0} = \frac{1}{a_{\rm inf}^2}$ from which we get

$$e^N > \left(\frac{H_{\rm inf}}{H_0}\right)^{1/2} = a_{\rm inf}^{-1} \tag{6.18}$$

In this relation H_0 is known but $H_{\rm inf}$ or equivalently $a_{\rm inf}$ are not. A possible value of H during inflation could be 10^{14}GeV or below, so let's use this value as an example. The Hubble constant

today is $H_0 \sim 10^{-18}\, s^{-1}$. Let's convert this to GeV via $E = \hbar\omega$ where $\hbar \sim 10^{-15}$eVs. This gives $H_0 \sim 10^{-33}\text{eV} = 10^{-42}\text{GeV}$. Thus we have

$$e^N > \left(\frac{10^{14}\text{GeV}}{10^{-42}\text{GeV}}\right)^{1/2} = 10^{28} \tag{6.19}$$

Solving for N this gives the often quoted estimate that we need around 60 e-folds of inflation. We can also estimate how long inflation needs to last from $N = H_{\text{inf}}T$. With $H_{\text{inf}} = 10^{14}\text{GeV} \sim 10^{38}s^{-1}$ we get that $T \sim 10^{-36}s$. So inflation can be extremely brief. However from the discussion presented here, there is no limit for how long inflation can last in principle, we only set a lower limit. In particle physics models of inflation there can be an upper limit. The length of inflation is also connected to the subject of **eternal inflation**. In some models inflation never ends globally.

Accelerated expansion also solves the flatness problem. To drive accelerated expansion as in Eq.(6.11) one needs an **inflation energy density** that goes as

$$\rho_{\text{inf}} \sim \frac{1}{a^{(2/n)}} \tag{6.20}$$

with $n > 1$ (this follows from Eq.(4.24) and Eq.(4.11)). This clearly dilutes more slowly than curvature, radiation and matter. In fact, in the case of exponential expansion, ρ_{inf} does not dilute at all, like dark energy, as we have seen. This means that after a long period of inflation, curvature, matter and radiation will all have diluted away to negligible amounts and the universe is empty except for ρ_{inf}.

While this solves the flatness problem, we are left with a new problem: why is the universe not empty. What is missing is a mechanism that **ends inflation** and converts the inflationary energy density into ordinary (relativistic) matter and radiation. This mechanism exists and is called **reheating**. Interestingly, reheating is somewhat natural in a quantum field theory of inflation, i.e. there are simple models that have this behavior. In summary, the matter and radiation in our universe is believed to have been created with the energy in the field that drove inflation.

6.4 The field theory of inflation

As physicists we would like a theory that "explains" the accelerated expansion (and its end) in terms of fundamental particles/fields and a Hamiltonian or Lagrangian for them. This will at the same time allow us to quantize the theory. It turns out that exponential expansion can be achieved by having a homogeneous scalar field that "slowly rolls down" a very flat potential. An example potential is illustrated in Fig. 7 (but many other potential shapes can also work). Most inflation models use a scalar field, the **inflaton**, and there is a sense (EFTofInflation) in which more complicated inflation models can also be described by a scalar degree of freedom. For those of you who had QFT, here is the Lagrangian of a scalar field, the inflaton ϕ, coupled to gravity:

$$S = \int \mathbf{d}^4x\sqrt{-g}\left[\frac{1}{2}R + \frac{1}{2}g^{\mu\nu}\partial_\mu\phi\,\partial_\nu\phi - V(\phi)\right] = S_{\text{EH}} + S_\phi\,. \tag{6.21}$$

The action (6.21) is the sum of the gravitational Einstein-Hilbert action, S_{EH}, and the action of a scalar field with canonical kinetic term, S_ϕ. The potential $V(\phi)$ describes the self-interactions

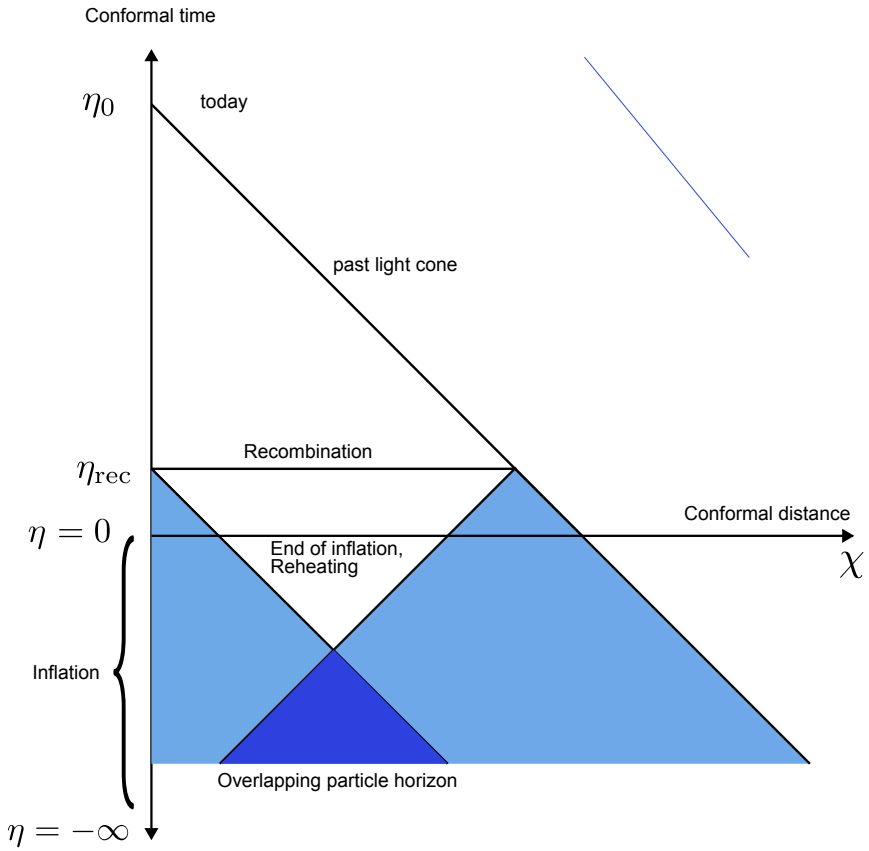


Figure 6. Particle horizon for CMB perturbations with inflation (adapted from 0907.5424).

of the scalar field (in addition there can be derivative self-interactions). Assuming the FRW metric for $g_{\mu\nu}$ and restricting to the case of a homogeneous field $\phi(t, \mathbf{x}) \equiv \phi(t)$, the scalar energy-momentum tensor

$$T^{(\phi)}_{\mu\nu} \equiv -\frac{2}{\sqrt{-g}} \frac{\delta S_\phi}{\delta g^{\mu\nu}} \tag{6.22}$$

takes the form of a perfect fluid (see e.g. Baumann's book for the math) with

$$\rho_\phi = \frac{1}{2}\dot{\phi}^2 + V(\phi)\,, \tag{6.23}$$

$$p_\phi = \frac{1}{2}\dot{\phi}^2 - V(\phi)\,. \tag{6.24}$$

The resulting equation of state

$$w_\phi \equiv \frac{p_\phi}{\rho_\phi} = \frac{\frac{1}{2}\dot{\phi}^2 - V}{\frac{1}{2}\dot{\phi}^2 + V}\,, \tag{6.25}$$

shows that a scalar field can lead to negative pressure ($w_\phi < 0$) and accelerated expansion ($w_\phi < -1/3$, see Eq.(4.24)) if the potential energy V dominates over the kinetic energy $\frac{1}{2}\dot{\phi}^2$. This is why the field needs to be rolling slowly. Inflation ends when the field rolls into a steeper

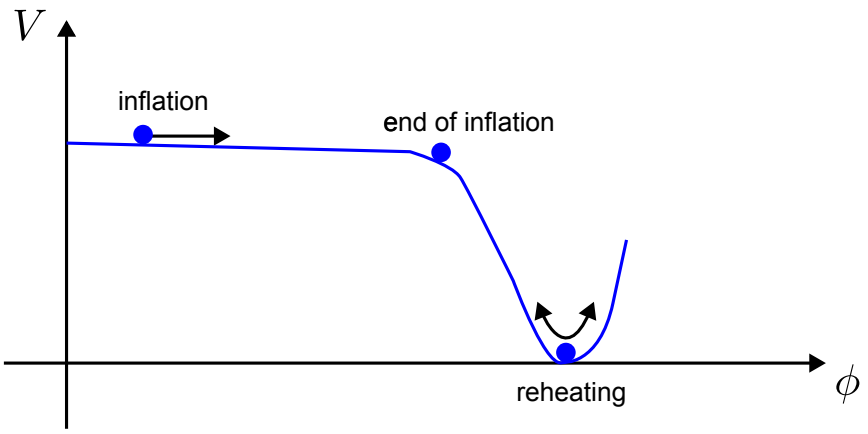


Figure 7. A typical potential for slow-roll inflation with a scalar field.

part of the potential as show in the Figure. Finally the field rolls into a minimum and starts oscillating around it. During this phase of oscillation, the inflaton acts like pressure-less matter ($w_\phi = 1$ above) and decays into other particles (those of the standard model if this is all there is). This process is called reheating and is very model dependent and very complicated. It turns out that predictions for cosmology don't depend much on reheating, all we need is that the inflaton energy is ultimately getting transformed into a thermal bath of standard model particles.

6.5 The quantum field theory of inflation

Having a field theory of inflation, we can now quantize it. Doing this goes beyond this course material. I want to make only a few comments:

- If you had QFT, you know that the standard method to quantize a field theory for which you know the Lagrangian or Hamiltonian is to promote the field $\phi(\mathbf{x}, t)$ and its conjugate momentum $\pi(\mathbf{x}, t)$ to operators and impose canonical commutation relations between them. This is correct here too. Schematically

$$\left[\hat{\phi}(\mathbf{x}, t), \hat{\pi}(\mathbf{x}', t)\right] = i\delta^3(\mathbf{x} - \mathbf{x}') \tag{6.26}$$

  where the delta function enforces locality.

- The quantization of inflation leads to exactly the kind of primordial perturbations we need to seed the structure formation of the universe. As you know, in quantum mechanics there is a fundamental uncertainty on quantities which means that the inflaton field ϕ cannot be exactly homogeneous but rather must have small "quantum wiggles" in it. A heuristic way to think about this is through Heisenbergs uncertainty principle in the form $\Delta t\, \Delta E \sim 1$ (where we set $\hbar = 1$). The time scale is set by the Hubble time $\Delta T \sim H_{\rm inf}^{-1}$ and the energy fluctuations are set by the fluctuations in ϕ, i.e. $\Delta E \sim \delta\phi$. The uncertainty relation predicts that we should see fluctuations of size $\delta\phi \sim H_{\rm inf}$.

- Because the potential is so flat during inflation, interaction terms such as ϕ^3 are very small. Inflation is thus an almost free (i.e. linear) field theory. The quantization of inflation thus

leads to a collection of (nearly) uncoupled quantum harmonic oscillators. Said differently, the Fourier modes of the inflaton field act like independent harmonic oscillators.

- Inflation does include and requires perturbative quantum gravity. Well below the Planck scale, we can quantize gravity in the sense of an effective field theory that integrates out the unknown UV physics of the ultimate quantum gravity theory. Depending on the details of the model, in particular its energy scale, inflation can be more or less sensitive to unknown quantum gravity "UV physics".

6.6 Primordial perturbations from inflation

In this unit we have focussed on the homogeneous evolution of the universe, but inflation cannot be discussed sufficiently without talking about perturbations, so let's get started with them.

6.6.1 Curvature perturbations from inflation

In cosmology (due to the cosmological principle), quantities such as the inflaton field (or the energy density etc.) can be split into their homogeneous background value and perturbations around it:

$$\phi(\mathbf{x}, t) = \phi^{hom}(t) + \delta\phi(\mathbf{x}, t) \tag{6.27}$$

As we shall see later, perturbations are best discussed in Fourier space, because these Fourier modes evolve almost independently (i.e. they are not coupled in the free theory). We therefore express perturbations as

$$\delta\phi(\mathbf{x}, t) \to \delta\phi(\mathbf{k}, t) \tag{6.28}$$

where $\mathbf{k}$ is the **comoving wave vector**.

Inflaton perturbations from quantum fluctuations lead to **curvature perturbations** in the metric and equivalently **density perturbations** in the energy density. To discuss these perturbations precisely would take us too far into relativistic perturbation theory. This topic is complicated in particular because of different possible gauge (coordinate) choices. To describe the scalar curvature field of the universe, you will most frequently encounter:

- The **comoving curvature perturbation $\mathcal{R}$**. Under some gauge conditions this is the curvature that a local observer would observe. This quantity is also conserved on superhorizon scales (see next section).

- The **curvature perturbation on uniform density hypersurfaces ζ**. This is often used in inflation calculations.

- The **Newtonian potential Φ**. The metric in **Newtonian gauge** is

$$ds^2 = a^2(\eta)\left[-(1+2\Phi)d\eta^2 + (1+2\Phi)\delta_{ij}dx^i dx^j\right] \tag{6.29}$$

  With some gauge subtleties on superhorizon scales, Φ is related to the energy density by Poisson's equation $\nabla^2\Phi \propto \rho$.

The distinction of these is not important in this course. The first main point to take away here is that we need **a single scalar field to describe the scalar curvature perturbations of the universe** (which are those induced by scalar density perturbations). This does not include tensor perturbations which are gravitational waves. Scalar and tensor perturbations together make up the full metric. So far there is no experimental evidence of primordial tensor perturbations.

It turns out that the inflaton perturbations $\delta\phi$ generate curvature perturbations as

$$\mathcal{R} \approx -\frac{H\delta\phi}{\dot{\phi}}, \tag{6.30}$$

where $\dot{\phi}$ is the inflaton speed. We won't derive this equation. The second main point to take away is thus that we can calculate the curvature perturbations $\mathcal{R}$ in a given inflation model, and that they are sourced by inflaton quantum perturbations $\delta\phi$.

6.6.2 Horizon exit and re-entering

To understand the physics of inflation, we need one more concept, the comoving Hubble radius. We have discussed the comoving and physical particle horizon. There is also the comoving and physical hubble radius (sometimes called the Hubble horizon). The physical Hubble radius is $d_H^{phys} = \frac{1}{H}$ and we discussed it in (Eq.(3.37)). The **comoving Hubble radius** is

$$d_H^{comov} = \frac{1}{aH} \tag{6.31}$$

It gives the size of the Hubble radius in comoving coordinates. An accelerating phase ($\ddot{a} > 0$) is equivalent to a shrinking Hubble radius

$$\ddot{a} > 0 \qquad \Leftrightarrow \qquad \frac{d}{dt}\left(\frac{1}{aH}\right) < 0 \tag{6.32}$$

We see that the comoving Hubble radius is shrinking during inflation, but growing during ordinary cosmological evolution (matter and radiation). You can think of the Hubble radius as the "size of the currently causally connected patch at time t", while the particle horizon tells us about the "size of the causally connected patch when considering the entire past".

The comoving Hubble radius is crucial to understand the behavior of **cosmological perturbations**. For these perturbations $\mathcal{R}(\mathbf{k}, t)$ we can study their equations of motion and find that their behavior (i.e. whether they grow, stay constant or get smaller with time) depends crucially on their size compared to the comoving Hubble radius. We can divide perturbations into two classes, by comparing their comoving wave number with the comoving Hubble horizon:

$$\textbf{subhorizon perturbations:} \quad k \gg aH\,. \tag{6.33}$$

$$\textbf{superhorizon perturbations:} \quad k < aH\,. \tag{6.34}$$

The size of cosmological perturbations compared to the Hubble radius as a function of time is illustrated in Fig.8.

We can now summarize what happens to cosmological perturbations:

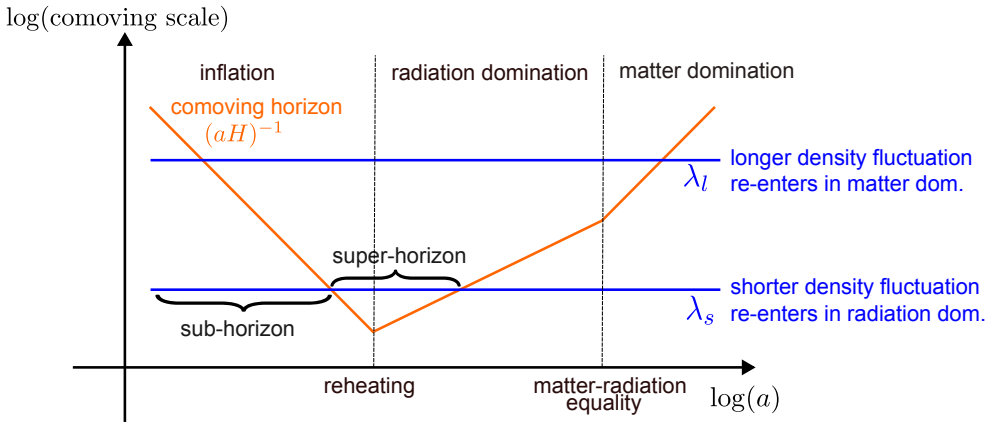


Figure 8. Horizon exit and re-entering of perturbations. On the y-axis is the comoving scale of the perturbation $1/k$ and the scale of the comoving Hubble radius $1/(aH)$. The modes that leave the horizon the latest (the smallest wave length λ) re-enter the horizon first (last out, first in).

- Quantum fluctuations create perturbations $\phi(\mathbf{k}, t)$ during inflation on subhorizon scales. Inflaton perturbations generate curvature perturbations as discussed above.
- The curvature perturbation $\mathcal{R}_\mathbf{k}$ exits the horizon during inflation and stops evolving. This can be proven in relativistic perturbation theory. Horizon exit is also connected to the fact that these perturbations **classicalize** (i.e. get a concrete value that we observe, like in a measurement). Note that the curvature perturbation does not change due to re-heating, which is why the details of reheating don't matter for cosmological predictions.
- At some point after the end of inflation the curvature perturbation $\mathcal{R}_\mathbf{k}$ re-enters the horizon and starts evolving again. Later we will see that the time when perturbations re-enter the horizon (during radiation domination or matter domination) is crucial for their amplitude.

We have not yet discussed how perturbations evolve in time. This depends on whether they are subhorizon or superhorizon and whether they evolve during radiation domination or during matter domination, or later during Lambda domination. The study of subhorizon and superhorizon evolution of perturbations is required to understand the qualitative properties of the matter power spectrum. We will briefly get back to this later in the course.

6.6.3 Primordial power spectrum

After inflation we are left with small curvature perturbations which seed the structure formation of the universe. We will discuss the statistics of these perturbations, and the meaning of their power spectrum, in more detail soon. However, let's summarize their properties and parameters here. The **primordial scalar curvature fluctuations** $\mathcal{R}$ are to very good approximation Gaussian and thus fully described by the following power spectrum:

$$\langle \mathcal{R}_\mathbf{k} \mathcal{R}_{\mathbf{k}'} \rangle = (2\pi)^3 \, \delta(\mathbf{k} + \mathbf{k}') \, P_\mathcal{R}(k) \,, \qquad \Delta^2_\mathrm{s} \equiv \Delta^2_\mathcal{R} = \frac{k^3}{2\pi^2} P_\mathcal{R}(k) \,. \tag{6.35}$$

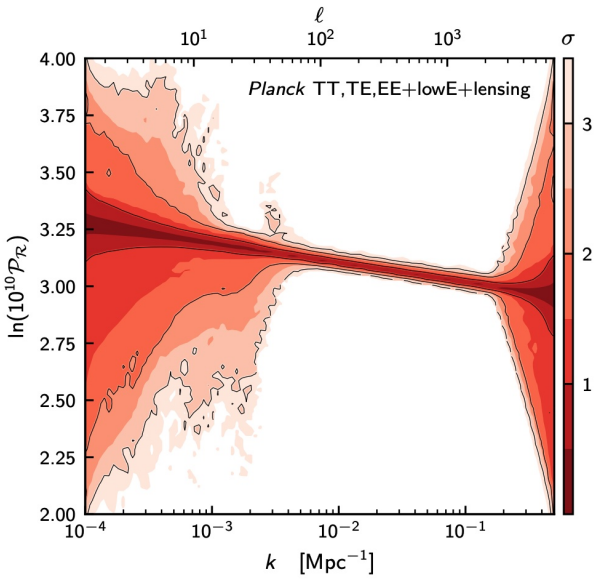


Figure 9. Reconstruction of the primordial power spectrum from Planck (1807.06211). The k scales where the power spectrum is constrained by the CMB are limited by the size of the observable universe to the left and by the small-scale damping of the CMB to the right. One can clearly see that $n_s \neq 1$.

Here, $\langle \dots \rangle$ defines the ensemble average of the fluctuations. The power spectrum is often approximated by a power law form

$$\Delta^2_{\rm s}(k) = A_{\rm s}(k_\star) \left(\frac{k}{k_\star}\right)^{n_{\rm s}(k_\star)-1+\frac{1}{2}\alpha_{\rm s}(k_\star)\ln(k/k_\star)} , \tag{6.36}$$

where $k_\star$ is an arbitrary reference or pivot scale. The scale-dependence of the power spectrum is defined by the scalar spectral index (or tilt)

$$n_{\rm s} - 1 \equiv \frac{d\ln\Delta^2_{\rm s}}{d\ln k} , \tag{6.37}$$

where scale-invariance corresponds to the value $n_{\rm s} = 1$. We may also define the running of the spectral index by

$$\alpha_{\rm s} \equiv \frac{dn_{\rm s}}{d\ln k} . \tag{6.38}$$

The free parameters we want to measure are thus

- The **primordial spectral amplitude A_s**. Planck measured $A_s \sim 2.1 \times 10^{-9}$ for $k_p = 0.05$ Mpc^{-1}.
- The **primordial spectral index n_s**. Planck measured $n_s = 0.966 \pm 0.005$.
- The **running of the spectral index α_s** (although no running has been detected).

This parametrization of the primordial power spectrum is very useful in practice. However, one can also directly reconstruct the power spectrum without a parametrization, as shown in Fig. 9.

In the same way inflation predicts small **primordial tensor fluctuations**. These are the **primordial gravitational waves** and are also given by a Gaussian power spectrum for the two polarization modes. For these one can measure in particular

- The **primordial tensor amplitude A_t** or the **tensor-to-scalar amplitude $r = \frac{A_t}{A_s}$**.
- The **primordial tensor spectral index n_t**.

Primordial gravitational waves have not been detected so there are only bounds on these parameters. The current constraint is about $r < 0.05$, i.e. tensor modes are less than 5% of the scalar modes. This bound will be significantly improved by upcoming CMB experiments. A detection of non-zero r is perhaps the best chance for a big fundamental physics discovery in the coming decade.

Finally, both scalar and tensor perturbations are not expected to be precisely Gaussian. At the very least, the coupling to gravity which is a non-linear theory, leads to some mode-coupling between perturbations. More than that, the inflaton potential as well as so-called derivative interactions also lead to **primordial non-Gaussianity**. The most obvious way to look for primordial non-Gaussianity is to search for a non-zero 3-point function

$$\langle \mathcal{R}_{\mathbf{k_1}} \mathcal{R}_{\mathbf{k_2}} \mathcal{R}_{\mathbf{k_3}} \rangle \neq 0 \tag{6.39}$$

This three-point function is called the **bispectrum**. Non-Gaussianity can come in many different bispectrum shapes (as well as higher N-point functions). The most famous bispectrum amplitude parameter is

- the **amplitude of local non-Gaussianity f_{NL}**

In many inflation models primordial non-Gaussianity is too small to be detected any time soon but there are also well-motivated scenarios where a detection could be around the corner. We will get back to this topic, which is a main research topic of mine, in more detail later in this course.

Here we have discussed the initial conditions in the curvature field. This curvature field is then converted into **initial conditions for matter and radiation**. In most models, the initial conditions for the fluids δ_r, δ_{CDM} and δ_{baryons} are all the same (up to an overall amplitude), since they are seeded by the same curvature field. Such initial conditions are called **adiabatic initial conditions**. There could in principle also be perturbations where the different fluids have different perturbations. Such perturbations are called **isocurvature perturbations**. Currently there is no experimental evidence for isocurvature perturbations and the most straight forward models of inflation and re-heating don't generate them. In this course, as in most cosmological analyses, we will assume adiabatic initial conditions.

Part II
Introduction to Computation and Statistics in Cosmology

In this section we introduce some of the main computational tools and data types used in cosmology. A practical goal will be to be able to analyze a dark matter simulation, extract its power spectrum, and run MCMC to determine its cosmological parameters. We also want to learn how to Fisher forecast experimental sensitivity, and compare it to the result in our simulation analysis. Analyzing a dark matter simulation comes without the practical complications of a real CMB or galaxy survey. We will discuss these real world complication in later units.

Further reading

The general references of Part 1 all contain some material on statistics and data analysis, in particular

- Dodelson, Schmidt, chapter 14
- Huterer, chapter 10.

Lecture notes or reviews that are specifically about data analysis in cosmology include:

- Heavens - Statistical techniques in cosmology. arxiv:0906.0664
- Verde - A practical guide to Basic Statistical Techniques for Data Analysis in Cosmology. arxiv:0712.3028
- Trotta - Bayesian Methods in Cosmology. arxiv:1701.01467
- Leclercq, Pisani, Wandelt - Cosmology: from theory to data, from data to theory. arxiv:1403.1260

7 From Initial Conditions to Observed Data

Let's summarize what we learned in the introductory unit. After inflation, through the process of reheating, we are left with **curvature perturbations $\boldsymbol{\mathcal{R}}$** which were generated by the quantum fluctuations of an inflaton field Φ. Because quantum fluctuations are stochastic, they have to be described by a probability density $\mathcal{P}(\{\mathcal{R}\})$. As we discussed, this distribution is (almost) Gaussian, and, to our current experimental sensitivity, depends only on two parameters:

$$\mathcal{P}^{A_s,n_s}_{\text{Gauss}}(\{\mathcal{R}\}) \tag{7.1}$$

These **initial conditions of the universe** then evolve forward in time, according to the standard model of cosmology or its extensions. For example, the **distributions of matter in the sky $\boldsymbol{\delta_m}$** (which can be probed e.g. by mapping galaxies δ_g) is given by some complicated

function $\mathcal{F}$ of the initial conditions which depends on the ΛCDM parameters (as well as other physical constants).

$$\mathcal{R} \xrightarrow{\mathcal{F}^{\Lambda}(\mathcal{R},t)} \delta_m(t) \tag{7.2}$$

The function (or simulation) that connects the initial conditions to whatever we observe in the data is sometimes called the **forward model** and it depends on physical parameters Λ that we want to measure such as Ω_m or the mass of neutrinos m_ν. By measuring δ_m, we can learn both about primordial parameters such as A_s, n_s and the parameters that influence the time evolution which we called Λ here. Roughly speaking, the function $\mathcal{F}$ is known exactly on large scales, approximately known on intermediate scales, and computationally intractable on small scales. A typical course on theoretical cosmology would now spend some weeks with calculating the function Eq.(7.2) analytically in **cosmological perturbation theory**, which amounts to **solving the Euler and Poisson equations** perturbatively. Instead, we will focus on how to perform data analysis and just use results from perturbation theory where needed. We will get back to (non-relativistic) perturbation theory in Sec. 21.

In some modern analyses in cosmology one tries to **reconstruct the initial conditions** $\mathcal{R}(\mathbf{x})$ directly from data such as the galaxy density $\delta_g(\mathbf{x})$. However, in the vast majority of analyses, we don't aim to reconstruct the initial conditions directly, but only their statistical parameters such as A_s, n_s, together with the parameters of cosmological time evolution Λ. This makes sense because the theory of the initial conditions only makes predictions for statistical parameters. For example, no theory can predict where in space a galaxy will form, but we can predict statistical properties of the galaxy field. For the same reason we don't usually have to analyze the volumetric data $\delta_g(\mathbf{x})$ directly but instead only **summary statistics** of this data.

The most important summary statistic (which in the Gaussian case carries all the information) is the **power spectrum** of the field. In many cases, we will measure the observed power spectrum P_g^{obs} of the galaxy data, and compare it to the theoretical power spectrum $P_g^{\text{theo}}(\Lambda, A_s, n_s)$ which depends on cosmological parameters. By adjusting these parameters so that P_g^{theo} matches P_g^{obs} we arrive at a measurement of our cosmological parameters. What we said here for galaxy density measurements, is also true for all other data sources that probe the matter and radiation distribution of the universe, in particular the Cosmic Microwave Background (CMB). The CMB is a particularly clean probe of cosmology because, as we shall see, it is linear in the initial conditions. Schematically, the "forward model" of the CMB is the linear mapping

$$\mathcal{R}_{\mathbf{k}} = T^{\Lambda}(k)\Theta_{\mathbf{k}}^{CMB} \tag{7.3}$$

where $\Theta_{\mathbf{k}}^{CMB}$ are the Fourier modes of the CMB temperature perturbations and $T(k)$ is the so-called **linear transfer function** which depends on cosmological parameters Λ. On the other hand, for the non-linear galaxy field, the Fourier modes are coupled to eachother in a complicated way.

In the present section we will develop the tools to analyze the matter distribution through the power spectrum in a simulated cosmological volume. This setup is already enough to write interesting papers in cosmology. In later units, we will use the same tools to analyze realistic data from the CMB and galaxy surveys, which comes with many interesting complications. We

will then also discuss how to go beyond the power spectrum to extract even more information from cosmological data.

8 Overview of Observed Data

Here is a list of the main sources of data that cosmologists have available.

- **Primary CMB anisotropies**. The primary CMB is the jewel of cosmological data. This is because it has perfectly understood physics, with a linear map to the initial conditions. Our best constraints on primordial physics come from the primary CMB. On the other hand, it cannot directly probe late time physics such as dark energy. For primordial physics, the only limitation is the **number of independent modes**. Modes here means either independent pixels or independent Fourier modes. First, the CMB is a 2d probe, while e.g. a galaxy survey is a 3d probe. Second, because of the free streaming length of photons, primary CMB anisotropies are damped away on small scales. This limits the number of available modes in the CMB to roughly

$$\mathcal{N}_{\mathcal{CMB}} \sim \ell_{\rm max}^2 \sim (2500)^2 \tag{8.1}$$

  where $\ell_{\rm max}$ is the maximum multipole scale, as we will see later. The **Baryon Acoustic Oscillations** in the power spectrum of the CMB reveal cosmological parameters such as Ω_m and Ω_B. The CMB is also **polarized**. While so-called E-mode polarization has been measured and roughly doubles the information in the CMB, cosmologist look for primordial **B-mode polarization** which would reveal the presence of **primordial gravitational waves**.

- **Secondary CMB anisotropies**. Two things happen to photons on the way from recombination to us. First, all photons are gravitationally lensed by the intervening matter. From the observed CMB one can reconstruct the so-called **lensing potential**, which is a weighted radial integral over the matter density on the line of sight. In this way, the CMB can also be used to probe physics that happens at later times in the universe, such as the "clumping" of non-relativistic neutrinos due to their non-zero mass. Second, a part of the CMB photons (a few percent) will hit a free electron and get re-scattered. Depending on the radial velocitiy of the electron, the photon will either gain or lose energy. This is the **Sunyaev-Zeldovich (SZ) effect**. The SZ effect can for example be used to probe the temperature of gas in clusters.

- **Large-scale structure (LSS) with galaxy surveys**. The distribution of galaxies probes the initial conditions of the universe as well as later time physics such as dark energy and neutrino masses. Galaxies are arranged in a **cosmic web of voids, filaments, walls and clusters**. The advantage over the CMB is that this is a 3-dimensional probe, and that it is not affected by the CMB damping scale, so that we can in principle probe far more modes. The disadvantage is that the smaller modes are very non-linear and hard to model. There is

however a redshift dependent scale of gravitational collapse, where primordial information should be entirely erased. The number of modes is

$$\mathcal{N}_{\mathcal{LSS}} \sim \left(\frac{k_{\max}}{k_{\min}}\right)^3 \tag{8.2}$$

so it goes cubic rather than quadratic since it is a volumetric probe. The resulting number depends very sensitively on the experiment and theoretical assumptions, which we will revisit later. Roughly speaking, current experiments have less independent accessible modes than the CMB but future experiments will have more. As in the case of the CMB, light from galaxies is also lensed. This lensing distorts the image of galaxies, which is called **cosmic shear or galaxy weak lensing**. Weak lensing probes the same cosmological volume as the galaxy positions, but it probes **all matter** (including dark matter) rather than only **luminous matter**, which gives somewhat different information. Using large-scale structure, one can measure the **Baryon Acoustic Oscillations** in the power spectrum. These provide a **standard ruler** that can measure distances, and thus the expansion history of the universe.

- **Large-scale structure (LSS) with intensity mapping**. The universe can of course not only be probed by identifying galaxies in the visible spectrum but in general by mapping any sort of radiation. In particular, one can map known **emission and absorption lines** of **both atoms and molecules** in the universe. There are many different such lines that I won't review here. A current exciting experimental front is **21cm intensity mapping** which looks for the 21cm spin-flip transition line of neutral hydrogen. The universe contains plenty of neutral hydrogen. The hardware for a **21 interferometer** is in principle cheap, only requiring a set of antennas and a supercomputer to correlate them. Achieving the required frequency resolution for precise redshifts is easy. However, due to extremely large foregrounds, this technique is not yet quite ready for cosmology. Even further in the future, it may be possible to do **21 cm intensity mapping of the dark ages**, the time before the first galaxies formed. In principle, there is an gigantic amount of primordial information hidden there ($\mathcal{N} \sim 10^{18}$). At the time scale of several decades it may be possible to access this information. A different intensity mapping technique, that is already in use, is **Lyman-α mapping** which looks for the **Lyman-α forest**, absorption lines in the emission of distant quasars due to neutral hydrogen in the intergalactic medium.

These are the data sources for which we are developing tools in this course. They have in common that they probe the universe as a density field. There is a different category of probes which looks at individual objects. Some of the main probes here are:

- **Type 1a Supernova Distance Measurements**. The discovery of dark energy was made possible by measuring distances (rather than redshifts) using type 1a supernovae. Their key property is that they have a known brightness (**standard candle**), so one can measure the so called **luminosity distance**. Type 1a SN thus probe the expansion history of the universe.

- **Strong lensing** (of quasars and galaxies by galaxies and galaxy clusters). If there is a dense enough chunk of matter infront of a cosmological light source, one can get multiple images or Einstein rings. From these strongly lensed images one can obtain a measurement of the lens profile (thus probing the dark matter profile) and, if the light source is time variable, one can get time delay measurements. These time delay measurements can be used to **measure the Hubble constant**.

- **Gravitational waves from compact objects.** A very recent addition to cosmological data are gravitational waves from black hole or neutron star mergers. These were discovered by LIGO in 2015. This is the first time that we observe the universe by something other than electromagnetic waves. For cosmology, it is particularly interesting that these blackholes can be used as **standard sirens**. One can reconstruct their absolute "loudness" from the shape of the gravitational wave pulse, and then estimate their distance from the observed loudness. In addition to LIGO's interferometric detector, there is also now strong evidence for gravitational waves from NANOGrav's pulsar timing, which may have discovered a gravitational wave background created by supermassive blackholes.

The list above is not meant to be complete, but covers the most important data sources for cosmology (rather than for the large field of multi-messenger astrophysics, which studies individual sources).

9 Random Fields in Cosmology

As we have discussed, the universe starts with a random field of initial conditions, which comes from quantum fluctuations during inflation. This random field then evolves in time, generating both the CMB perturbations and galaxy density perturbations. Much of cosmology is thus about calculating and measuring statistical properties of random fields such as their power spectrum, and their evolution in time. Our next goal is thus to study random fields.

In this section we will describe random fields in 3+1 dimensional space. These are the coordinates in which a super-observer would observe the universe, who observes all of space at equal time. In reality we can only observe the universe on the light cone, i.e. the farther away the object is the farther back we also look in time. Observations on the light cone will be described in Part IV of these lectures. Further we will be working in co-moving coordinates, so that the background expansion as a function of time is factored out.

9.1 Random scalar fields in Euclidean space

We will start by discussing random scalar fields in Euclidean 3D space. Typical scalar fields are

- The density $\rho(\mathbf{x})$ of a continuous field such as dark matter.
- The number density $n(\mathbf{x})$ of a discrete tracer such as galaxies where $n(\mathbf{x}) = \delta N(\mathbf{x})/\delta V$.
- (Over-)density fields based on these quantities

$$\delta(\mathbf{x}) = \frac{\rho(\mathbf{x}) - \bar{\rho}}{\bar{\rho}} = \frac{\rho(\mathbf{x})}{\bar{\rho}} - 1 = \frac{n(\mathbf{x})}{\bar{n}} - 1 \tag{9.1}$$

By construction, the spatial average of δ vanishes,

$$\langle \delta(x,t) \rangle = 0 \tag{9.2}$$

In this section we will primarily think about non-relativistic cosmology (such as galaxy surveys), so that the pressure can be ignored. However, the statistical techniques we develop are equally important for relativistic cosmology. We will get back to relativistic physics in the unit about the CMB.

In the following I will use the notation $f(\mathbf{x})$ for a general scalar field. In general $f(\mathbf{x})$ also depends on time as $f(\mathbf{x}, t)$ but in this section we won't need explicit time dependence (and it is easy to add to the notation).

9.1.1 PDF of random fields

The scalar fields $f(\mathbf{x})$ are **random fields**. That means that they are drawn from some probability density function (PDF) which we will write as

$$\mathcal{P}[f] \tag{9.3}$$

To be precise, here $f(\mathbf{x})$ is a continuous function so that $\mathcal{P}$ is a **probability density functional**. Continuous functions are appropriate for analytic calculations, but simulations as well as data analysis must discretize space. We will study discrete coordinate below which you will encounter in numerical examples.

As far as we know the universe is **statistically homogeneous** and **statistically isotropic**, which means that on average it looks the same in all places and all directions. Mathematically this can be expressed by defining a translation operator

$$\hat{T}_{\boldsymbol{a}} f(\mathbf{x}) \equiv f(\mathbf{x} - \boldsymbol{a}), \tag{9.4}$$

and a rotation operator

$$\hat{R} f(\mathbf{x}) \equiv f(R^{-1}\mathbf{x}), \tag{9.5}$$

where R is a rotation matrix. Given these operators the isotropic and homogeneous field PDF obeys

$$\mathcal{P}[f(x)] = \mathcal{P}[\hat{T}_a f(x)] \tag{9.6}$$

and

$$\mathcal{P}[f(x)] = \mathcal{P}[\hat{R} f(x)] \tag{9.7}$$

for any translation or rotation.

9.1.2 Position space correlation functions

The field PDF, which can often be parametrized as a function of a few cosmological parameters, encodes all there is to know about the random field. Often we don't work with the field PDF directly but with statistics that are easier to work with. All of these can in principle be calculated from the field PDF. In position space the most basic such statistic is the (position space)

correlation function. Correlation functions of fields are expectation values of products of fields at different spatial points. The two point correlator is

$$\xi(\mathbf{x},\mathbf{y}) \equiv \langle f(\mathbf{x}) f(\mathbf{y})\rangle = \int Df\, \mathcal{P}(f) f(\mathbf{x}) f(\mathbf{y}), \tag{9.8}$$

where the integral is a functional integral (or path integral) over field configurations. This is the usual definition of an expectation value in statistics.

By statistical homogeneity, the correlation function can only depend on the difference of the positions $\mathbf{x}+\mathbf{r}$ and $\mathbf{x}$ and statistical isotropy enforces dependence on the magnitude only. In this case the correlation function is given by

$$\langle f(\mathbf{x}) f(\mathbf{x}+\mathbf{r})\rangle = \xi(|\mathbf{r}|) = \xi(r) \tag{9.9}$$

The proof of this intuitive statement can easily be found in textbooks. The correlation function of galaxies and other observable fields can be measured and is used to probe properties of the universe.

9.1.3 Fourier Space

We often work in **Fourier space** (also called **momentum space**) rather than position space. The main reason is that on large scales or at early times, the perturbations of the universe evolve linearly. This means that Fourier modes evolve independently, rather than coupling to another. Recall that Fourier space is used to solve Linear Homogeneous Differential Equations with Constant Coefficients. Fourier transforms can diagonalize such differential equations, turning them into algebraic equations in Fourier space. This is because differentiation in real space corresponds to multiplication with k in Fourier space. This sort of differential equations appear when we linearize the Euler, Poisson and continuity equation for small perturbations. We will do this math later. For now, let's discuss Fourier space.

We are using the following conventions for the continuous Fourier transform.

$$f(\mathbf{k}) = \int d^3x \exp^{-i\mathbf{k}\cdot\mathbf{x}} f(\mathbf{x}) \tag{9.10}$$

and

$$f(\mathbf{x}) = \int \frac{d^3k}{(2\pi)^3} \exp^{i\mathbf{k}\cdot\mathbf{x}} f(\mathbf{k}) \tag{9.11}$$

A nice discussion of other consistent Fourier conventions is in appendix A of 0907.5424.

Cosmology needs Fourier space (also called k-space or momentum space) as much as position space (also called x-space or configuration space), so let's review some properties:

- If $f(\mathbf{x})$ is dimensionless, the Fourier modes $f(\mathbf{k})$ have dimension [length]^3.
- If the position space fields is real we have $f(\mathbf{k}) = f^*(-\mathbf{k})$. This can be shown by Fourier transforming $f(\mathbf{x}) = f^*(\mathbf{x})$.

- Under spatial translation, the Fourier transform gets a phase factor

$$\hat{T}_a f(\mathbf{k}) = \int d^3\mathbf{x}\, f(\mathbf{x}-\boldsymbol{a}) e^{-i\mathbf{k}\cdot\mathbf{x}} \tag{9.12}$$

$$= \int d^3\mathbf{x}'\, f(\mathbf{x}') e^{-i\mathbf{k}\cdot\mathbf{x}'} e^{-i\mathbf{k}\cdot\boldsymbol{a}} \tag{9.13}$$

$$= f(\mathbf{k}) e^{-i\mathbf{k}\cdot\boldsymbol{a}}. \tag{9.14}$$

  where $\mathbf{x}' = \mathbf{x} - \boldsymbol{a}$.

- The Fourier space representation of the nabla operator is given by $\nabla \to i\mathbf{k}$.

We will often use the Dirac delta function identity

$$\delta_D(\mathbf{k}-\mathbf{k}') = \frac{1}{(2\pi)^3} \int \mathbf{d}^3x\, e^{\pm i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}}. \tag{9.15}$$

The delta function has the dimension of the inverse of its argument, thus here in 3d it has dimension $[k^{-3}] = [\text{length}]^3$. This is also the orthogonality relation for plane waves in an infinite volume. In the other direction the delta function is

$$\delta_D(\mathbf{x}-\mathbf{x}') = \frac{1}{(2\pi)^3} \int \mathbf{d}^3k\, e^{\pm i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')}. \tag{9.16}$$

Using these delta function definitions you can check that the Fourier transform of the Fourier transform returns the original function as it must.

9.1.4 Power spectrum

The famous power spectrum is the 2-point function in Fourier space:

$$\langle f(\mathbf{k}) f^*(\mathbf{k}')\rangle = \int d^3x d^3x' e^{-i\mathbf{k}\cdot\mathbf{x}} e^{i\mathbf{k}'\cdot\mathbf{x}'} \langle f(\mathbf{x}) f(\mathbf{x}')\rangle \tag{9.17}$$

$$= \int d^3r d^3x' e^{-i\mathbf{k}\cdot\mathbf{r}} e^{-i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}'} \xi(r) \tag{9.18}$$

$$= (2\pi)^3 \delta_D(\mathbf{k}-\mathbf{k}') \int d^3r e^{-i\mathbf{k}\cdot\mathbf{r}} \xi(r) \tag{9.19}$$

$$= (2\pi)^3 \delta_D(\mathbf{k}-\mathbf{k}') P(k) \tag{9.20}$$

where, in the second line, we introduced $\mathbf{r} \equiv \mathbf{x}-\mathbf{x}'$ and then performed the integral over $\mathbf{x}'$ which gives us a Dirac delta function. We see that different Fourier modes are uncorrelated. This is a consequence of translation invariance. The power spectrum can also be written in the equivalent form

$$\langle f(\mathbf{k}) f(\mathbf{k}')\rangle = (2\pi)^3 \delta_D(\mathbf{k}+\mathbf{k}') P(k) \tag{9.21}$$

(note the change of signs and conjugates) due to the reality condition.

The power spectrum $\mathcal{P}(k)$ and the correlation function $\xi(r)$ are related by the 3-dimensional Fourier transform. We can simplify this relation as follows. Using spherical coordinates, $\mathbf{k}\cdot\mathbf{r} =$

$kr\cos\theta$ we have

$$P(k) = \int d^3r\, e^{-i\mathbf{k}\cdot\mathbf{r}}\xi(r) \tag{9.22}$$

$$= \int_0^{2\pi} d\phi \int_{-1}^{1} d(\cos\theta) \int_0^{\infty} dr\, r^2 e^{-ikr\cos\theta}\xi(r) \tag{9.23}$$

$$= 2\pi \int_0^{\infty} dr \frac{r^2}{ikr}\left[e^{ikr} - e^{-ikr}\right]\, \xi(r) \tag{9.24}$$

$$= \frac{4\pi}{k}\int_0^{\infty} dr\, r\sin(kr)\xi(r) \tag{9.25}$$

$$= 4\pi \int_0^{\infty} dr\, r^2 j_0(kr)\xi(r). \tag{9.26}$$

where

$$j_0(x) = \frac{\sin x}{x} \tag{9.27}$$

is a spherical Bessel function of order zero. These functions are frequently encountered in cosmology. In the other direction, one can express ξ in terms of the power spectrum as

$$\xi(r) = \int \frac{d^3r}{(2\pi)^3}\, e^{i\mathbf{k}\cdot\mathbf{r}} P(k) \tag{9.28}$$

$$= \int \frac{dk\, k^2}{2\pi^2} j_0(kr)\, P(k). \tag{9.29}$$

The power spectrum has the dimension [length]^3. It is often useful to define the **dimensionless power spectrum** by multiplying with k^3

$$\Delta^2(k) = \frac{k^3}{2\pi^2} P_{\mathcal{R}}(k) \tag{9.30}$$

which we encountered before in Eq.(6.35). There are different conventions around for the π and the 2 factor.

9.2 Gaussian Random Fields

The statements we made above are correct for any homogeneous isotropic random field. However a **Gaussian Random Field (GRF)** is particularly important in cosmology. Inflationary physics predicts an (almost) Gaussian Random field, and on large scales, the universe remains Gaussian through its evolution. Let's now see how a GRF is defined.

9.2.1 GRFs in Position Space

A vector $\mathbf{f} = [f_1, \ldots, f_N]$ of random variables is called Gaussian, if the joint probability density function (PDF) is a multivariate Gaussian

$$P(\mathbf{f}) = \frac{1}{\sqrt{(2\pi)^N |C|}} \exp\left[-\frac{1}{2} f_i C_{ij}^{-1} f_j\right] \tag{9.31}$$

where the positive definite, symmetric $N \times N$-matrix $C_{ij} = \langle f_i f_j \rangle$ is called the covariance matrix. A random field $f : \mathbb{R}^3 \to \mathbb{R}$ is a Gaussian random field (GRF) if for arbitrary collections of

field points $(x_1, ..., x_N)$ the variables $[f(x_1), ..., f(x_N)]$ are joint Gaussian variables. Since any N-point function can be calculated from the field PDF, the GRF is **fully defined in terms of its covariance matrix**, which is the 2-point function. As we see from the PDF, a Gaussian random field is not neccessarily homogeneous and isotropic. To make it so, we need to enforce that the covariance matrix is

$$C_{ij} = \langle f_i f_j \rangle = \xi(|\mathbf{x}_i - \mathbf{x}_j|) \tag{9.32}$$

Here we have discretized the PDF, i.e. we wrote $\mathbf{f}$ as a finite dimensional vector rather than an infinite dimensional function. In principle, for the continuous fields that we have discussed so far, we should express the GRF using a Gaussian functional which is schematically

$$F[f(\mathbf{x})] \propto \exp\left(-\frac{1}{2}\int d^3x d^3y\, f(\mathbf{x}) C(\mathbf{x}-\mathbf{y}) f(\mathbf{y})\right) \tag{9.33}$$

In practice we don't usually need this continuous expression.

9.2.2 GRFs in Fourier space

The field PDF has the same mathematical form when expressed in momentum space, where the vector $\mathbf{f} = [f_1, \ldots, f_N]$ is over the Fourier modes. In this case, for a homogeneous isotropic field, the covariance matrix is

$$\langle f_{\mathbf{k}_i} f^*_{\mathbf{k}_j} \rangle \propto \delta_{ik} P(k) \tag{9.34}$$

with Kronecker delta δ_{ik}. The covariance matrix for a homogeneous field is thus **diagonal in momentum space**, i.e. the covariance matrix between different Fourier modes is zero. Note that the Fourier modes $f_{\mathbf{k}}$, unlike the position space field, are complex numbers. We will get back to the precise definition of a Gaussian random field on a finite volume (i.e. with discrete Fourier modes) shortly, including the proportionality factor.

9.3 Power Law Power Spectra

9.3.1 Power Laws

A typical power spectrum

$$\langle f(\mathbf{k}) f^*(\mathbf{k}') \rangle = (2\pi)^3 \delta_D(\mathbf{k} - \mathbf{k}') \mathcal{P}(k) \tag{9.35}$$

is given by a power law,

$$P(k) = A\, k^n \tag{9.36}$$

where n is called the **spectral index**. The corresponding dimensionles power spectrum is

$$\Delta^2(k) \propto A\, k^{n+3} \tag{9.37}$$

Some special cases are $n = 0$ which is called white noise and $n = 1$ which is called the Harrison-Zeldovich power spectrum, as we will see below.

9.3.2 Potential and Density Power Spectra

To discuss typical power spectra we need to discriminate between the power spectrum of the gravitational potential Φ and the power spectrum of resulting matter perturbations δ_m, which are related by the Poisson equation. The primordial gravitational potential Φ is closely related to the primordial curvature perturbation $\mathcal{R}$ as we discussed in Sec. 6.6.1. We are switching notation to Φ now instead of $\mathcal{R}$ because the following is also valid in Newtonian gravity which works with a Newtonian gravitational potential Φ. The Poisson equation for matter in an expanding space-time is

$$\nabla^2\Phi(\mathbf{x}) = \frac{4\pi G}{c^2}a^2\bar{\rho}\delta(\mathbf{x}) \tag{9.38}$$

which in momentum space is

$$-k^2\Phi(\mathbf{k}) = \frac{4\pi G}{c^2}a^2\bar{\rho}\delta(\mathbf{k}) \tag{9.39}$$

Thus the power spectra of the two quantities will be related by

$$P_\Phi(k) \propto k^{-4}P_\delta(k) \tag{9.40}$$

The relation between the density power spectrum and the primordial potential power spectrum is thus

$$P_\delta(k) \propto k^4 P_\Phi(k) \propto k\Delta^2_\Phi(\mathbf{k}) \tag{9.41}$$

If the dimensionless primordial power spectrum is constant (rather than k dependent), i.e. if the primordial curvature perturbations have the same amplitude on all scales, then the density power spectrum is

$$P_\delta(k) \propto k \qquad\qquad P_\Phi(k) \propto k^{n-4} = k^{-3} \tag{9.42}$$

This is called a **Harrison-Zeldovich Power Spectrum**. The potential fluctuations are said to be **scale-invariant** primordial fluctuations. This is the case $n_s = 1$ (remember that $n_s \sim 0.96$, so it's close).

We also sometimes need the variance of the field (also called the zero-lag correlation function) given by

$$\sigma_f^2 \equiv \langle f^2(x)\rangle = \xi_f(0) = 1/(2\pi)^3 \int d^3k\, P_f(k)\,. \tag{9.43}$$

which can be written as

$$\sigma_f^2 \equiv \int d\ln k\, \Delta_f^2(k)\,, \tag{9.44}$$

where

$$\Delta_f^2(k) \equiv \frac{k^3}{2\pi^2}P_f(k)\,. \tag{9.45}$$

The dimensionless power spectrum is thus the contribution to variance per log wave number. If the dimensionless power spectrum has a peak at some k_* then fluctuations in f are dominated by wavelengths of order $\frac{2\pi}{k_*}$. Note that the integral Eq.(9.44) is divergent in the large k limit unless the field is smoothed at some scale so that the power spectrum goes to zero. We will get back to the smoothing of fields.

9.3.3 Illustrating Power Law Power Spectra in 2d

We'd like to get some intuition for how Gaussian density fields look like that have power law spectra. We will illustrate these fields in 2d (as appropriate for the CMB).

Let's have a look at the position space correlation function for arbitrary dimension d:

$$\langle f(\mathbf{x})f(\mathbf{y})\rangle = \int \frac{d^dk}{(2\pi)^d}\int \frac{d^dk'}{(2\pi)^d} e^{-i\mathbf{k}\cdot\mathbf{x}-i\mathbf{k}'\cdot\mathbf{y}}\langle f(\mathbf{k})f(\mathbf{k}')\rangle \tag{9.46}$$

$$= \int \frac{d^dk}{(2\pi)^d} e^{-i\mathbf{k}\cdot(\mathbf{x}-\mathbf{y})} P_f(k). \tag{9.47}$$

We again consider a power law of form

$$P(k) = A\,k^n \tag{9.48}$$

For any dimension d, for $n = 0$, we find that

$$\langle f(\mathbf{x})f(\mathbf{y})\rangle \sim \delta_D^d(\mathbf{x}-\mathbf{y}) \tag{9.49}$$

Which means that all pixels are uncorrelated. This is called white noise. This is illustrated in Fig.10 top left.

If we decrease n, for example $P(k) = A\,k^{-1}$ the correlation between points increases, i.e. nearby points become more likely to have a similar value (Fig.10). There is a special value of n for which the field becomes **scale invariant**, i.e. the correlation between any two points becomes independent of distance. In dimension d this is $n = -d$, so in 3d it is $n = -3$ as we have already seen above and in 2d it is $n = -2$. To show this, we rescale the correlation function by a factor λ

$$\langle \Phi(\lambda x)\Phi(\lambda y)\rangle = \int \frac{d^dk}{(2\pi)^d} e^{-i\lambda k\cdot(x-y)} k^{-d} \tag{9.50}$$

$$= \int \frac{d^dk'}{(2\pi)^d} \lambda^{-d}\, e^{-ik'\cdot(x-y)} \left(\frac{k'}{\lambda}\right)^{-d} \tag{9.51}$$

$$= \langle \Phi(x)\Phi(y)\rangle \tag{9.52}$$

where we changed variables to $\mathbf{k}' = \lambda\mathbf{k}$. If we go more negative with n than the scale invariant value, then the universe becomes more inhomogeneous on larger scales (i.e. we see larger perturbations if we zoom out). This would be inconsistent with the cosmological principle which wants the universe to be homogeneous on large scales..

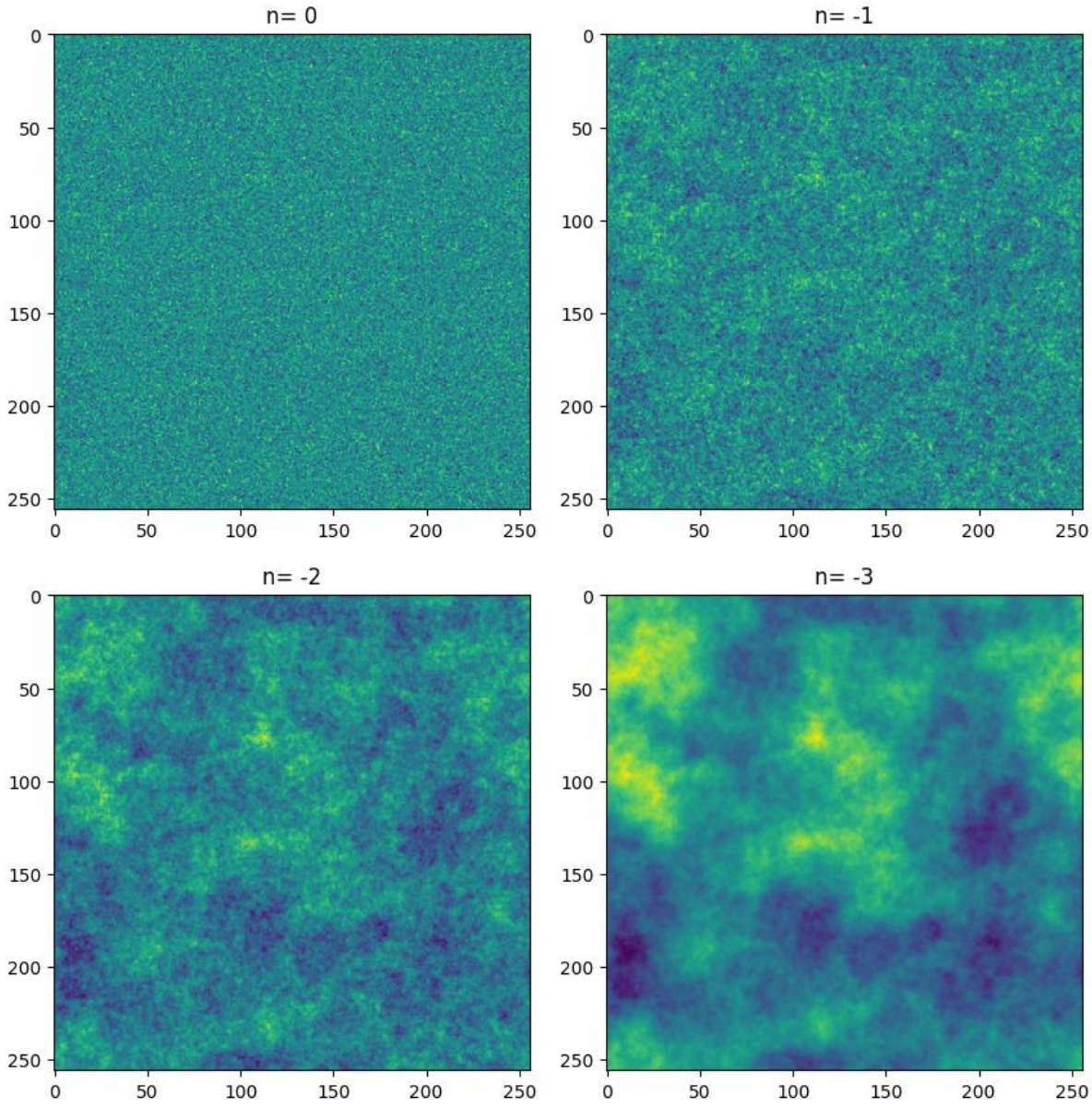


Figure 10. 2d Gaussian random fields with power spectrum $P(k) = A\,k^n$ for various n. For $n = 0$ we get white noise and for lower n we get progressively more correlation. The scale invariant case is $n = -2$ in 2d. This plot was made with the Pylians library. The script is provided with the course material.

9.4 Matter Power Spectrum and Boltzmann Codes

Apart from power-law power spectra, the most important power spectrum in this course is perhaps the **matter power spectrum**. It can be written approximately as two different power laws.

9.4.1 Transfer function and Growth function

On large scales the density perturbations of the universe evolve linearly. That means that the evolution of perturbations $\delta(\mathbf{k})$ can be described by a **transfer function** T as follows:

$$\delta_m(k,t) \propto T(k)\, D(a(t))\, \delta_m(k,t_i) \tag{9.53}$$
$$\propto T(k)\, D(a(t))\, k^2\, \Phi(k,t_i) \tag{9.54}$$

where t_i is the initial time, taken just after inflation. The function $D(a(t))$ is called the **growth function**. There are various possible conventions for T and D but the key point is that the time and k dependence factorizes.

It follows that the power spectrum evolves as

$$P_m(k,t) = T^2(k)\, D^2(a(t))\, P_m(k,t_i) \tag{9.55}$$

For non-relativistic matter the transfer function is

$$T(k) \sim 1 \qquad \text{for } k < k_{\rm eq} \tag{9.56}$$
$$T(k) \sim \text{constant} \times k^{-2} \qquad \text{for } k > k_{\rm eq} \tag{9.57}$$

The transfer function thus crucially depends on their comoving k compared to the wave number

$$k_{eq} = (aH)_{eq} \tag{9.58}$$

of the mode that entered the Hubble horizon at the time of matter-radiation equality. Modes larger than this ($k < k_{eq}$) enter the horizon in the matter dominated era and modes smaller than this ($k > k_{eq}$) will have entered during radiation domination. The form of the transfer functions comes from the fact that radiation domination stops (or more precisely slows to logarithmic) growth of perturbations as we will discuss later.

If we start with the power-law spectrum $P \sim k^n$, then it subsequently evolves to

$$P(k) = \begin{cases} k^n & \text{for } k < k_{\rm eq} \\ k^{n-4} & \text{for } k > k_{\rm eq} \end{cases}$$

with the turnover near $a_k \approx a_{k_{\rm eq}} \sim 0.01\,\text{Mpc}^{-1}$. As we have discussed, for a Harrison-Zeldovich power spectrum we have $n = 1$, while in we measure $n = 0.96$. The linear matter power spectrum, scaled to today, is shown in Fig. 11.

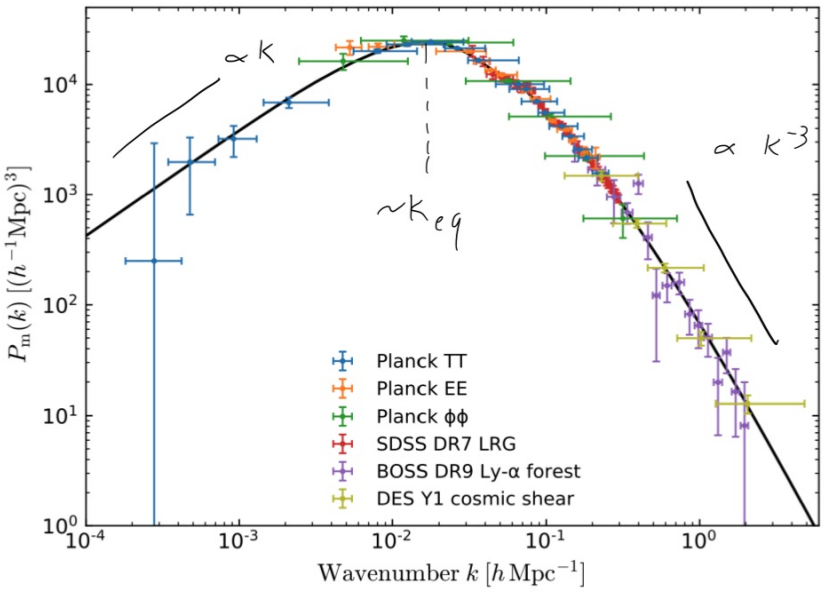


Figure 11. The linear matter power spectrum scaled to $z = 0$ from Planck 2018 CMB data and various galaxy surveys. We see the power spectrum turnover at k_{eq}.

9.4.2 Boltzmann Solvers

The polynomial approximation to the matter power spectrum is of course not exact, especially around the turnover which happens rather gradually. In practice, cosmologists calculate the power spectrum of Gaussian fields such as the matter density and the CMB temperature with so-called **Boltzmann codes** which solve the Einstein-Boltzmann equations, briefly discussed in Sec. 5.5, numerically. There are currently two main codes that are used by the community:

- **CAMB**. https://camb.info/. CAMB, written in Fortran, has been the community standard for a long time. It comes with a nice python wrapper and documentation https://camb.readthedocs.io/en/latest/ and a demo notebook https://camb.readthedocs.io/en/latest/CAMBdemo.html.
- **CLASS**. https://lesgourg.github.io/class_public/class.html. This is a more recent C++ implementation which is becoming increasingly popular. If you need to modify (rather than just run) a Boltzmann code you may be better off with CLASS. There is also a useful extension to calculate non-linear power spectra at smaller scales using EFTofLSS called Class-PT. We will get back to this topic in the unit on LSS. There is also a python wrapper https://github.com/lesgourg/class_public/wiki/Python-wrapper.

Both of these codes can for example generate the black theory curve in Fig. 11. In general these codes are only correct in the linear regime (CMB, LSS for $k \lesssim 0.1$ Mpc^{-1} at $z = 0$). However they have some extensions to calculate power spectra in the non-linear regime. These are based on results from non-linear perturbation theory or N-body simulations.

9.5 Random scalar fields in discrete coordinates

While analytic work is usually done with continuous distributions, numerical work usually uses a discrete data representation. For example, the 3d matter distribution can be represented as a box of 3d pixels. Such a finite box also has a finite set of discrete Fourier modes. We work in 3d but adapting to 2d is straight forward.

9.5.1 Fourier conventions

We now work in a finite pixelized box with side length L, with K grid points per side length and grid length $H = L/K$. The box volume is then $V_{\rm box} = L^3$ and the pixel volume is $V_{\rm pix} = H^3 = V_{box}/N_{pix}$ where $N_{pix} = K^3$. Our Fourier conventions are then:

$$f(\mathbf{x}) = \frac{1}{V_{\rm box}} \sum_{\mathbf{k}_i} f(\mathbf{k}) e^{i\mathbf{k}_i \cdot \mathbf{x}} \tag{9.59}$$

$$= \int_{V_k} \frac{d^3k}{(2\pi)^3} f(\mathbf{k}) e^{i\mathbf{k}_i \cdot \mathbf{x}} \tag{9.60}$$

$$f(\mathbf{k}) = V_{\rm pix} \sum_{\mathbf{x}_i} f(\mathbf{x}_i) e^{-i\mathbf{k}\cdot\mathbf{x}_i} \tag{9.61}$$

$$= \int_{V_{box}} d^3x\, f(\mathbf{x})\, e^{-i\mathbf{k}\mathbf{x}} \tag{9.62}$$

Here we also introduced $V_k = (\frac{2\pi}{H})^3$, the volume of the Fourier space cube. Fourier conventions are non-uniform in the literature, but I am using probably the most common one here. The best discussion of the discrete Fourier transfrom in cosmology that I have found is in Donghui Jeong's PhD thesis appendix A. The discrete mode orthogonality condition with our conventions is

$$\sum_{\mathbf{x}_i} e^{i\mathbf{k}\mathbf{x}_i} \left(e^{i\mathbf{k}'\mathbf{x}_i}\right)^* = \sum_{\mathbf{x}_i} e^{i(\mathbf{k}-\mathbf{k}')\mathbf{x}_i} = N_{pix}\delta_{\mathbf{k}\mathbf{k}'} \tag{9.63}$$

The larger K the more high frequency modes we can resolve. The lowest Fourier mode which covers one side length with one whole mode is called the **fundamental mode**

$$k_f = \frac{2\pi}{L} \tag{9.64}$$

The total **set of Fourier modes** is

$$\mathbf{k}_i \in (n_x, n_y, n_z) k_f \tag{9.65}$$

where (n_x, n_y, n_z) is a set of whole numbers that runs from $-K/2$ to $K/2$. The finite number of Fourier modes leads to **cosmic variance** as we will discuss further shorty.

The power spectrum is given by

$$\left\langle f(\mathbf{k}) f(\mathbf{k}')^* \right\rangle = V_{\rm box} P_f(k)\, \delta_{\mathbf{k}\mathbf{k}'} \tag{9.66}$$

A few more comments:

- In our conventions, for a dimensionless $f(\mathbf{x}_i)$ the Fourier modes have again dimension [length]^3. The power spectrum also has dimension [length]^3. The Kronecker delta is dimensionless.
- For discrete modes the reality condition reads again $f_{-\mathbf{k}} = f^*_{\mathbf{k}}$.
- If your data is not periodic there will be "spurious transfer of power" (aliasing) in your FT. We'll address this when we talk about experimental masks.
- The highest frequency that we can resolve is the **Nyquist frequency** of the grid given by

$$k_{Ny} = \frac{K}{2} k_f = \frac{K\pi}{L} = \frac{\pi}{H} \tag{9.67}$$

9.5.2 Gaussian random field in discrete coordinates

For a Gaussian random field, of course our discrete Fourier modes are drawn from a Gaussian distribution. Since they are complex numbers, let's understand precisely what that means. This will also suggest how we can generate such a field in code.

We split modes into their real and imaginary parts as $f(\mathbf{k}) = a(\mathbf{k}) + ib(\mathbf{k})$. The reality of f requires $f_{-\mathbf{k}} = f^*_{\mathbf{k}}$ and hence the real and imaginary parts of $f_{\mathbf{k}}$ must satisfy the constraints $a_{-\mathbf{k}} = a_{\mathbf{k}}$ and $b_{-\mathbf{k}} = -\boldsymbol{b}_k$. For a homogeneous and isotropic Gaussian processes these modes are drawn from:

$$p(a_{\mathbf{k}}, b_{\mathbf{k}}) = p(a_{\mathbf{k}}) p(b_{\mathbf{k}}) \tag{9.68}$$

$$= \frac{1}{\sqrt{\pi}\sigma_k} \exp\left(-\frac{a^2_{\mathbf{k}}}{\sigma^2_k}\right) \frac{1}{\sqrt{\pi}\sigma_k} \exp\left(-\frac{b^2_{\mathbf{k}}}{\sigma^2_k}\right) \tag{9.69}$$

where the variance is equal to $\sigma^2_k/2$ and is the same for both independent variables a_k and b_k. For the expectation value of the product of Fourier coefficients we get

$$\langle f_{\mathbf{k}} f_{\mathbf{k}'} \rangle = \langle a_{\mathbf{k}} a_{\mathbf{k}'} \rangle + i(\langle a_{\mathbf{k}} b_{\mathbf{k}'} \rangle + \langle a_{\mathbf{k}'} b_{\mathbf{k}} \rangle) - \langle b_{\mathbf{k}} b_{\mathbf{k}'} \rangle = \sigma^2_k \delta_{\mathbf{k},-\mathbf{k}'},$$

where we have taken into account that $a_{-\mathbf{k}} = a_{\mathbf{k}}$ and $b_{-\mathbf{k}} = -b_{\mathbf{k}}$ and that the two random variables a and b are uncorrelated. One can also change variables from a, b to polar coordinates r, ϕ and find that the PDF of r is a Rayleigh distribution and the PDF of the phase is constant:

$$p(r) = \frac{2r}{\sigma^2} e^{-\frac{r^2}{\sigma}} \qquad\qquad p(\phi) = \frac{1}{2\pi}. \tag{9.70}$$

Comparing with the above we have $\sigma^2_k = V_{\rm box} P_f(k)$.

9.5.3 Implementing Fourier Transforms

The way to calculate the Fourier transform is of course the famous **Fast Fourier Transform** algorithm. In numpy this can be done in n dimensions with numpy.fftn and numpy.rfftn. Numerical methods are explained in detail in the famous book "Numerical Recipes" (NumReps) which goes in detail over FFTs. A main choice one can make is whether to use a complex FFT or a real FFT called RFFT. The RFFT enforces the reality condition $f_{-\mathbf{k}} = f^*_{\mathbf{k}}$ to be more memory efficient (it saves only half as many modes) but the code can look somewhat less elegant than with the complex FFT.

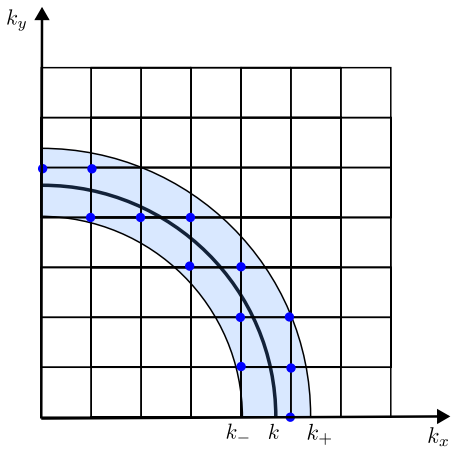


Figure 12. Discrete Fourier grid and discrete modes (blue points) contributing to a wavenumber bin (blue shaded region) centered around k.

9.6 Power spectrum estimation

9.6.1 Power spectrum estimator

We estimate the power spectrum in bins, i.e. spherical shells of width dk corresponding to an interval in wavevector magnitude $[k^{\pm}] = [k^-, k^+) = [k - \frac{dk}{2}, k + \frac{dk}{2})$ centered at k by averaging the square of all the modes in this bin:

$$\hat{P}(k) = \frac{1}{N_k V_{box}} \sum_{k_i \in [k^-, k+)} f(k_i) f^*(k_i),$$

where N_k is the number of cells in the k-bin. The modes are illustrated in Fig.12.

This estimate assumes statistical isotropy (i.e. the power spectrum depends only on the magnitude of the wave vector. It is easy to see that this estimator is unbiased:

$$\langle \hat{P} \rangle = \frac{1}{N_k V_{box}} \sum_{k_i \in [k^-, k+)} \langle f(\mathbf{k}_i) f^*(\mathbf{k}_i) \rangle \tag{9.71}$$

$$= \frac{1}{N_k V_{box}} \sum_{k_i \in [k^-, k+)} V_{\rm box} P_f(k_i) \tag{9.72}$$

$$= P_f(k) \tag{9.73}$$

where we have used Eq.(9.66). In the last step, for a finite bin width, in principle we should average the theory power spectrum over the same modes, but in practice for narrow bins this is not neccessary (i.e. all modes in the bin have almost the same theoretical power spectrum). Power spectrum estimation is no longer as easy when statistical isotropy is broken by the experiment (i.e. we only observe a part of the sky) but we will deal with this difficulty in later chapters.

We note that here we estimate the power spectrum, which is defined as an expectation value over the PDF of the random field (as in Eq.(9.8) but in Fourier space) from a single universe. This is possible because the modes are independent so they are all drawn from the same PDF, whether in the same universe or in different universes.

It is also useful to calculate the number of modes in a power spectrum bin analytically (rather than numerically from the FT grid). The number of modes is given by

$$N_k = \frac{V_{\rm shell}}{V_f} = \frac{4\pi k^2 dk}{V_f} = \frac{4\pi k^3 {\rm dln}k}{V_f} \tag{9.74}$$

where V_f is called the volume of the fundamental cell $V_f = \frac{(2\pi)^3}{V_{box}}$. This means that in logarithmic bins (in which we often plot the power spectrum), the number of modes goes as k^3.

9.6.2 Cosmic Variance

We want to calculate the variance of the power spectrum estimator, which tells us how well we can measure the power spectrum from a given cosmological volume. Recall that the variance of a random variable X is

$$V[x] = \langle (X - \langle X \rangle)^2 \rangle \tag{9.75}$$

$$= \langle X^2 \rangle - (\langle X \rangle)^2 \tag{9.76}$$

For the power spectrum estimator we thus have

$$V[\hat{P}(k)] = \langle \hat{P}^2(k) \rangle - \langle \hat{P}(k) \rangle^2 = \frac{1}{N_k^2 V^2} \sum_{\mathbf{k}_i, \mathbf{k}_j \in [k\pm]} \langle f(\mathbf{k}_i) f(-\mathbf{k}_i) f(\mathbf{k}_j) f(-\mathbf{k}_j) \rangle - P^2(k) \tag{9.77}$$

To make progress, we need an important theorem for Gaussian fields called **Wick's theorem** (which also appears in QFT). Wick's theorem states that the higher order correlation functions of a Gaussian random field of mean zero can be expressed as certain products of the two point function. This implies that the 3-point function $\langle f(\mathbf{k}_1) f(\mathbf{k}_2) f(\mathbf{k}_3) \rangle$ of such a field must vanish, as do all odd N-point function. On the other hand for a 4-point functions, as we have in our calculation, Wick's theorem states:

$$\langle f_1 f_2 f_3 f_4 \rangle = \langle f_1 f_2 \rangle \langle f_3 f_4 \rangle + \langle f_1 f_3 \rangle \langle f_2 f_4 \rangle + \langle f_1 f_4 \rangle \langle f_2 f_3 \rangle \tag{9.78}$$

A general discussion of Wick's theorem can be found in some cosmology text books. Using this relation in our calculation we get

$$\langle \hat{P}^2(k) \rangle - \langle \hat{P}(k) \rangle^2 = \frac{1}{N_k^2} \sum_{\mathbf{k}_i, \mathbf{k}_j \in [k\pm]} P(k_i) P(k_j) + \frac{2}{N_k^2} \sum_{\mathbf{k}_i \in [k\pm]} P^2(k_i) - P^2(k) \tag{9.79}$$

$$= \frac{2}{N_k} P^2(k) \tag{9.80}$$

Thus the relative error on the power spectrum is given by:

$$\frac{\Delta P}{P} = \sqrt{\frac{2}{N_k}} \tag{9.81}$$

The factor of 2 comes from the modes being complex (thus in the k-sphere we double counted) and the $\sqrt{N}$ may be familiar from the error bar in a histrogram. From Eq.(9.74) we see that the error scales with the box volume as $V_{box}^{-(1/2)}$. So we need four times the cosmological volume to reduce

the error bar by a factor of 2. Remember that this calculation is only correct in the Gaussian case. For the smaller scale non-linear power spectrum the variance also gets a contribution from the so-called connected 4-point function which cannot be reduced to 2-point functions by Wick's theorem.

The error that results from a finite number of Fourier modes in a given cosmological volume is called **cosmic variance**. Since the observable universe is also finite in size, we will never be able to measure the power spectrum on large scales precisely. This is reflected in the error bars in Fig. 11.

9.6.3 Power spectrum estimation with experimental noise

A common situation in cosmology is that we can only measure the field that we are interested in up to some noise. Let's assume that we measure the galaxy density field δ_g up to some noise n, i.e. that

$$\delta_g^{obs}(\mathbf{k}) = \delta_g(\mathbf{k}) + n(\mathbf{k}) \tag{9.82}$$

where $n(\mathbf{k})$ is a Gaussian field of noise. These quantities have the power spectrum

$$\langle \delta_g(\mathbf{k}) \delta_g^*(\mathbf{k}') \rangle = (2\pi)^3 \delta_D(\mathbf{k} - \mathbf{k}') P_g(k) \tag{9.83}$$

$$\langle n(\mathbf{k}) n^*(\mathbf{k}') \rangle = (2\pi)^3 \delta_D(\mathbf{k} - \mathbf{k}') N(k) \tag{9.84}$$

$$\langle \delta_g(\mathbf{k}) n^*(\mathbf{k}') \rangle = 0 \tag{9.85}$$

The last equation means that they are mutually uncorrelated. This is a good approximation in many situations. If we now calculate the expectation value of the observed power spectrum we get

$$\langle \hat{P}^{obs} \rangle = \frac{1}{N_k V_{box}} \sum_{k_i \in [k^-, k+)} \langle \delta_g^{obs}(\mathbf{k}_i) \delta_g^{obs*}(\mathbf{k}_i) \rangle \tag{9.86}$$

$$= P_g(k) + N(k) \tag{9.87}$$

That means that to measure the true power spectrum P_g, we need to subtract the noise power spectrum from the observed spectrum. We can also calculate the variance

$$V[\hat{P}^{obs}(k)] = \frac{2}{N_k} (P_g(k) + N(k))^2 \tag{9.88}$$

In the case of a galaxy survey, the noise is to good approximation given by the comoving number density $\bar{n}_g$ as

$$N_g = \frac{1}{\bar{n}_g}. \tag{9.89}$$

This is called **Poisson noise** or **shot noise**. If P is much larger than N then the error is **cosmic variance domianted** while if N is larger than P the error is noise dominated. For a CMB experiment the noise power spectrum depends on the angular resolution and sensitivity of the CMB detector. For both CMB and galaxy surveys, on large scales the error is always cosmic variance dominated.

10 Basics of Statistics

We now discuss how to measure and forecast cosmological parameters. Above we have learned how to

- calculate a theoretical power spectrum, such as the power spectrum P_m of the matter density, using for example CAMB
- estimate the observed power spectrum $\hat{P}^{obs}$ from the data, for example the matter distribution δ_m^{obs} with the estimator Eq. (9.86).

To measure cosmological parameters Λ from the power spectrum, schematically one adjusts the parameters Λ so that P_m matches $\hat{P}^{obs}$ up to noise. While we are primarily considering the power spectrum here, the same methodology can be used for other summary statistics such as the 3-point function. In this section we discuss how this parameter fitting works in detail, and also how we can forecast parameter sensitivity without having taken any data. Let's start with reviewing some concepts from statistics.

10.1 Estimators

While we have already used the concept, let's define what an **estimator** is. If a random variable x is characterized by a PDF $p(\mathbf{x}|\lambda)$ dependent on a parameter λ, then an estimator for λ is a function $\mathcal{E}(\mathbf{x})$ used to infer the value of the parameter. If a given dataset $\{\mathbf{x}_{\rm obs}\}$ is drawn from the distribution $p(\mathbf{x}, \lambda)$, then $\hat{\lambda} = \mathcal{E}(\mathbf{x}_{\rm obs})$ is the estimate of the parameter λ from the given observations. We often use a "hat" over the variable to indicate an estimator. Since $\mathcal{E}$ is a function of a random variable, it is itself a random variable. A random variable obtained as a function of another set of random variables is often referred to as a **statistic**.

An estimator for a parameter λ is **unbiased** if its average value is equal to the true value of the parameter:

$$\langle\hat{\lambda}\rangle = \lambda. \tag{10.1}$$

We want our estimator to be unbiased. However, biased estimators can also be useful, since it can be possible to "unbias" them.

After unbiasedness, the second key property of an estimator is its expected error or **variance**. The variance is given by

$$\begin{aligned} \mathrm{Var}[\hat{\lambda}] &= \langle(\hat{\lambda} - \langle\hat{\lambda}\rangle)^2\rangle & (10.2)\\ &= \langle\hat{\lambda}^2\rangle - \langle\hat{\lambda}\rangle^2 & (10.3)\end{aligned}$$

and the error is given by the square root of the variance

$$\sigma_\lambda = \sqrt{\langle(\hat{\lambda} - \langle\hat{\lambda}\rangle)^2\rangle}, \tag{10.4}$$

We try to find an estimator that is unbiased and that has as small an error as possible. One can often show which estimator will have the smallest possible error bar. Such an estimator is called an **optimal estimator**. We already saw an optimal estimator, the one for the power spectrum in Eq. (9.86), although we did not prove optimality.

If we have several estimators, we are also interested in their **covariance**

$$\mathrm{Cov}[\hat{\lambda}_i, \hat{\lambda}_j] = \langle (\hat{\lambda}_i - \langle \hat{\lambda}_i \rangle)(\hat{\lambda}_j - \langle \hat{\lambda}_j \rangle) \rangle \tag{10.5}$$

$$= \langle \hat{\lambda}_i \hat{\lambda}_j \rangle - \langle \hat{\lambda}_i \rangle \langle \hat{\lambda}_j \rangle \tag{10.6}$$

From the covariance, one can also calculate their **cross-correlation** (which is between -1 and 1) as

$$\mathrm{Corr}[\hat{\lambda}_i, \hat{\lambda}_j] = \frac{\mathrm{Cov}[\hat{\lambda}_i, \hat{\lambda}_j]}{\sqrt{\mathrm{Cov}[\hat{\lambda}_i, \hat{\lambda}_i]\mathrm{Cov}[\hat{\lambda}_j, \hat{\lambda}_j]}} \tag{10.7}$$

which tells us whether the estimators are correlated, anti-correlated or uncorrelated ($\mathrm{Corr} = 0$).

10.2 Likelihoods, Posteriors, Bayes Theorem

The central concept to connect data to theory is the **likelihood** function. The likelihood is **the probability of measuring data d given a model M with parameters $\boldsymbol{\lambda}$.** We write it as

$$\mathcal{L}(\mathbf{d}|\boldsymbol{\lambda}, M) \tag{10.8}$$

where the line | is read as "given". It is often possible to write down the likelihood function analytically. The likelihood does not tell us what model and model parameters are likely given the data (rather it answers the opposite question). It is the **posterior** probability

$$\mathcal{P}(\boldsymbol{\lambda}, M|\mathbf{d}) \tag{10.9}$$

that measures parameters for us. From now on I will drop the label M, since a likelihood and a posterior are always only defined assuming some model (e.g. Lambda-CDM), and are different if you assume a different model. To connect the posterior and the Likelihood we need **Bayes theorem**:

$$\mathcal{P}(\boldsymbol{\lambda}|\mathbf{d}) = \frac{\mathcal{L}(\mathbf{d}|\boldsymbol{\lambda})\mathcal{P}(\boldsymbol{\lambda})}{\mathcal{P}(\mathbf{d})} \tag{10.10}$$

Here we also have the

- The **prior** $\mathcal{P}(\boldsymbol{\lambda})$ of the parameters in model M before the data is analyzed. The choice of the prior can be somewhat tricky but often flat or Gaussian works.
- The **evidence** $\mathcal{P}(\mathbf{d})$ which is the probability of seeing the data $\mathbf{d}$ under any parameters $\boldsymbol{\lambda}$ of the model. The evidence can also be written as

  $$\mathcal{P}(\mathbf{d}) = \int \mathcal{L}(\mathbf{d}|\boldsymbol{\lambda})\mathcal{P}(\boldsymbol{\lambda}) d\boldsymbol{\lambda} \tag{10.11}$$

  The evidence can be difficult to calculate numerically because it is often a huge multidimensional integral. However in many cases we do not need to evaluate it, since it only depends on the data and thus does not change our measurement of model parameter $\boldsymbol{\lambda}$. The evidence is however useful for model selection as we will discuss later.

10.3 Gaussian Likelihoods

Both the likelihood and the posterior are of course probability distributions. It turns out that a Gaussian likelihood is often a good approximation of the data (while the posterior is often not Gaussian). Consider first the simple case of a single Gaussian random variable. Imagine you want to measure a persons's weight w (following Dodelson's example). To get an error bar, you will measure the weight m times. In each measurement, you get data d_i which is given by the true value plus some (Gaussian) noise: $d_i = w + n_i$. If our measurements are independent, then the likelihood is

$$\mathcal{L}(\{d_i\}_{i=1}^m | w, \sigma_w) = \frac{1}{(2\pi\sigma_w^2)^{m/2}} \exp\left(-\frac{\sum_{i=1}^m (d_i - w)^2}{2\sigma_w^2}\right) \tag{10.12}$$

which is the product of the likelihoods of the individual measurements. The parameters of our model here are the true weight w and the variance of the data σ^2. To find the maximum likelihood estimator for our parameters, in this simple case we can do the maximization analytically. We are assuming a flat prior here, so that the prior does not change our estimate. Taking the derivative we get

$$\frac{\partial \mathcal{L}}{\partial w} = \frac{1}{\sigma_w^2 (2\pi\sigma_w^2)^{m/2}} \left(\sum_{j=1}^m (d_j - w)\right) \exp\left(-\frac{\sum_{i=1}^m (d_i - w)^2}{2\sigma_w^2}\right) \tag{10.13}$$

which we has its maximum at

$$\frac{\partial \mathcal{L}}{\partial w} = 0 \;\Rightarrow\; \sum_{j=1}^m (d_j - w) = 0 \tag{10.14}$$

The maximum likelihood estimator is then

$$\hat{w} = \frac{1}{m} \sum_{i=1}^m d_i \tag{10.15}$$

which one can guess of course. In the same way we can calculate the σ_w^2 estimator from

$$\frac{\partial \mathcal{L}}{\partial \sigma_w^2} = \mathcal{L} \times \left[\frac{-m}{2\sigma_w^2} + \frac{\sum_{i=1}^m (d_i - w)^2}{2\sigma_w^4}\right] \tag{10.16}$$

Setting this to zero gives

$$\hat{\sigma_w^2} = \frac{1}{m} \sum_{i=1}^m (d_i - w)^2. \tag{10.17}$$

which is the well-known estimator of the variance. Taking into account that w is also estimated from the same data one gets $m \to m - 1$.

We can also calculate the variance (error) of our two estimators using Eq.(10.4). The answer is

$$\mathrm{Var}[\hat{w}] = \frac{\sigma_w^2}{m} \tag{10.18}$$

and

$$\mathrm{Var}[\hat{\sigma_w^2}] = \frac{2}{m}(\sigma_w^2)^4 \tag{10.19}$$

These calculations are written out in Dodelson's textbook.

Often we are interested in measuring only a subset of the parameters, while other parameters are considered **nuisance parameters**. In the weight example, we may be interested in measuring w but do not have knowledge of σ_w. Then, given the full posterior $P(w, \sigma_w|d_i)$, we can calculate the the marginalized posterior

$$P(w|d_i) = \int_0^\infty d\sigma_w P(w, \sigma_w|d_i). \tag{10.20}$$

10.3.1 Power spectrum likelihood

An important example of an (approximately) Gaussian likelihood is the likelihood of the power spectrum as a function of cosmological parameters $\mathbf{p}$:

$$\ln \mathcal{L}(\{\hat{P}(k)\}|\boldsymbol{\lambda}) = -\frac{1}{2}\sum_{k_i,k_j}\left(\hat{P}(k_i) - P^{theo}(k_i,\boldsymbol{\lambda})\right)\left(\mathrm{Cov}^{-1}\right)_{k_i,k_j}\left(\hat{P}(k_j) - P^{theo}(k_j,\boldsymbol{\lambda})\right) + \text{const.} \tag{10.21}$$

Here we have dropped the determinant term of the Gaussian, assuming that the covariance matrix does not depend on cosmological parameters (a common assumption) and included off-diagonal terms between different modes of the power spectrum (which are needed in the non-linear regime and when including masks). The covariance matrix (defined in Eq.(10.5)) in a real experiment can usually not be calculated analytically but is estimated using simulations. We'll talk more about this in the large-scale structure unit. We also note that, while the covariance matrix is in principle parameter dependent, it is usually better to evaluate it at fixed fiducial values. This has to do with the power spectrum likelihood not being exactly Gaussian, see arxiv:1204.4724 for details. For this reason we have also dropped the determinant term of the Gaussian.

10.3.2 Gaussian Field likelihood

The likelihood for the power spectrum is an example of a **likelihood for a summary statistic**. In cosmology we also use **likelihoods for fields**. We have already seen the PDF of a Gaussian field. When interpreted as a function of the parameters $\boldsymbol{\lambda}$ of the power spectrum the field-level likelihood is

$$-2\ln \mathcal{L}(\{\delta_m(k)\}|\boldsymbol{\lambda}) = \sum_{\mathbf{k}_i,\mathbf{k}_j}\delta(\mathbf{k}_i)\left(\mathrm{Cov}(\boldsymbol{\lambda})\right)^{-1}_{k_i,k_j}\delta^\dagger(\mathbf{k}_j) + \log|\mathrm{Cov}(\boldsymbol{\lambda})| \tag{10.22}$$

where the covariance is given in terms of the power spectrum $P^{theo}(k,\mathbf{p})$ and diagonal in the homogeneous case. We have already discussed this PDF in Sec. 9.5.2 in a different notation. The dagger † is required because the Fourier modes are complex. In this equation we have kept a parameter dependent covariance matrix, and thus the determinant term of the Gaussian. The determinant here is required if we want to make the PDF dependent on cosmological parameters.

10.3.3 Beyond Gaussianity

Gaussianity is often ensured by the **central limit theorem (CLT)**. For example, in the power spectrum estimator, we average over many modes in a single k bin, which have a similar variance and are approximately uncorrelated (at least on roughly linear scales), so the CLT holds to good

approximation. If we can assume Gaussianity then our task in specifying the likelihood is vastly simplified because we "only" need to determine the right covariance matrix (usually from a set of simulations).

However we should stress that likelihoods are not always Gaussian or nearly Gaussian. For small number statistics, Poissonian likelihoods are also common. If the likelihood is more complicated, we need a different approach. Fortunately, if we have enough simulations, we can instead learn $\mathcal{L}(\mathbf{d}|\mathbf{p})$ as a free function from simulations, using machine learning. This approach is called **likelihood-free inference** (LFI) (meaning that we must learn the likelihood). We will get back to these methods in Sec. 28. Of course learning a free PDF is much more difficult than determining just a covariance matrix. In my impression, LFI with ~ 10 variables (such as the cosmological parameters) often works, but it gets difficult in much higher dimension.

10.4 Using the likelihood and the posterior

From the likelihood one can also define an estimator, called the **maximum likelihood estimator** (MLE) $\hat{\boldsymbol{\lambda}}$. It is given by solving

$$\frac{\partial \ln \mathcal{L}(\mathbf{d}|\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}}\Big|_{\boldsymbol{\lambda}=\hat{\boldsymbol{\lambda}}} = 0 \tag{10.23}$$

for $\hat{\boldsymbol{\lambda}}$. Sometimes this can be done analytically. In general the MLE does not have to be the optimal estimator though under common assumptions it is. Further, from the posterior one can define an estimator called the **maximum aposteriori estimator** (MAP) given by solving

$$\frac{\partial \ln \mathcal{P}(\boldsymbol{\lambda}|\mathbf{d})}{\partial \boldsymbol{\lambda}}\Big|_{\boldsymbol{\lambda}=\hat{\boldsymbol{\lambda}}} = 0 \tag{10.24}$$

for $\hat{\boldsymbol{\lambda}}$. If the prior is flat, which is sometimes a good and sometimes a bad choice, the MLE and the MAP are the same. I may add more details on estimation theory later. Using estimators such as MLE under some model is typical for the **frequentist** approach to statistics. In this approach, the error bar is set by calculating the covariance of the estimator analytically, or if that is not possible, by estimating it from simulations (Monte Carlo).

On the other hand, the **Bayesian** approach considers the complete posterior density. A Bayesian would often **sample from the posterior** using MCMC, and summarize the posterior by quantities such as the **posterior mean** (which is not the same as the MAP). Power spectrum analysis is usually done in a Bayesian way using MCMC.

Sometimes the difference between frequentist and Bayesian statistic is also presented in terms of the use of priors and of updating beliefs with new data. In my opinion there is no need to be either a frequentist or a Bayesian and you can consistently use concepts from both sides. A common complaint about frequentist analysis in cosmology is that we have only one universe and cannot repeat the experiment. However one can still run simulations of different initial conditions or analytically integrate over initial conditions. As long as you correctly interpret your math (e.g. you do not claim that a 3-sigma frequentist excess of some estimator is equivalent to a 3 sigma detection of your favorite new physics model) you won't have inconsistencies. The full Bayesian method formally answers the interesting physical questions most directly, but it is not always computationally tractable and not always needed.

10.5 Fisher forecasting

In many situations we want to know the error on our parameters that an experiment can achieve before having taken any data. Theory papers need to estimate whether their effect is observable, and experiments need to be designed to meet specified sensitivity goals. These forecasts are commonly made using the Fisher forecasting formalism (a different option is running MCMC on synthetic data). We first discuss Fisher forecasting for Gaussian likelihoods, but the formalism also generalizes to other likelihoods.

If a given observed variable $O_{\boldsymbol{a}}$ is characterized by Gaussian distributed errors, then its likelihood is

$$\mathcal{L} \propto e^{\chi^2/2}, \tag{10.25}$$

where the χ^2 statistic is defined as:

$$\chi^2 = \sum_{\boldsymbol{a}} \frac{\left[O_{\boldsymbol{a}}(\lambda) - \hat{O}_{\boldsymbol{a}}(\lambda)\right]^2}{\mathrm{Var}\left[O_{\boldsymbol{a}}\right]}, \tag{10.26}$$

where $\hat{O}_{\boldsymbol{a}}$ are the measured values of our observable, for example the power spectrum bins $\hat{P}(k_\alpha)$. To find the best fit parameters $\hat{\lambda}$ we minimize χ^2 (which is equivalent to maximizing the likelihood). We assume here that the variance is not parameter dependent and thus we don't need the determinant term in the likelihood.

If we first work in the 1-dimensional case with only one variable λ we can expand the χ^2 around its minimum

$$\chi^2(\lambda) = \chi^2(\bar{\lambda}) + \frac{1}{2}\left.\frac{\partial^2\chi^2}{\partial\lambda^2}\right|_{\lambda=\bar{\lambda}} (\lambda - \bar{\lambda})^2 \,. \tag{10.27}$$

The linear term vanishes at the minimum. The quadratic term describes the **local curvature of the likelihood**. It tells us how narrow or wide the minimum is, and thus what it's error bar is. If we define

$$\mathcal{F} \equiv \frac{1}{2}\left.\frac{\partial^2\chi^2}{\partial\lambda^2}\right|_{\lambda=\bar{\lambda}} \,, \tag{10.28}$$

then we can estimate the minimum possible error on λ as $1/\sqrt{F}$. Note that the Fisher matrix depends on where we have assumed the minimum to be, i.e. it depends on the **fiducial parameters** $\bar{\lambda}$ of our forecast.

If we compute $\mathcal{F}$ explicitly we get

$$\mathcal{F}_{\lambda\lambda} = \sum_\alpha \frac{1}{\mathrm{Var}\left[O_{\boldsymbol{a}}\right]} \left[\left(\frac{\partial O_{\boldsymbol{a}}}{\partial\lambda}\right)^2 + \left(O_{\boldsymbol{a}} - \hat{O}_{\boldsymbol{a}}\right)\frac{\partial^2 O_{\boldsymbol{a}}}{\partial\lambda^2}\right]. \tag{10.29}$$

To forecast F we will not have observed data. Rather we should be taking the expectation value, which simplifies our expression because $\left\langle O_{\boldsymbol{a}} - \hat{O}_{\boldsymbol{a}}\right\rangle = 0$ at the minimum (because the measurements will fluctuate around the truth). Thus

$$F_{\lambda\lambda} = \langle\mathcal{F}_{\lambda\lambda}\rangle \tag{10.30}$$

$$= \sum_\alpha \frac{1}{\mathrm{Var}\left[O_{\boldsymbol{a}}\right]} \left[\left(\frac{\partial O_{\boldsymbol{a}}}{\partial\lambda}\right)^2\right] \tag{10.31}$$

This quantity is called the **Fisher Information** F. For several variables, this generalizes to the **Fisher information matrix**:

$$F_{\lambda\lambda'} = \langle \mathcal{F}_{\lambda\lambda} \rangle \tag{10.32}$$

$$= \sum_a \frac{1}{\mathrm{Var}\left[O_{\boldsymbol{a}}\right]} \left[\left(\frac{\partial O_{\boldsymbol{a}}}{\partial \lambda} \right) \left(\frac{\partial O_{\boldsymbol{a}}}{\partial \lambda'} \right) \right] \tag{10.33}$$

If the variables are correlated, the Fisher matrix is

$$F_{\lambda\lambda'} = \sum_{a,b} \left(\frac{\partial O_{\boldsymbol{a}}}{\partial \lambda} \right) \mathrm{Cov}^{-1}\left(O_{\boldsymbol{a}}, O_{\boldsymbol{b}}\right) \left(\frac{\partial O_{\boldsymbol{b}}}{\partial \lambda'} \right) \tag{10.34}$$

From the Fisher matrix one can obtain two different errors. If we have several parameters and we assume all parameters except λ are known then

$$\sigma_\lambda = \frac{1}{\sqrt{F_{\lambda\lambda}}} \qquad \text{unmarginalized} \tag{10.35}$$

More commonly, we want to know the error on λ if all other parameters are marginalized over. This is obtained by inverting the Fisher matrix as follows

$$\sigma_\lambda = \sqrt{(F^{-1})_{\lambda\lambda}} \qquad \text{marginalized} \tag{10.36}$$

Often the marginalized errors are significantly larger than the unmarginalized ones. An illustration of this in the 2-parameter case is shown in Fig. 13.

10.5.1 Non-Gaussian likelihoods and the Rao-Cramer bound

The Fisher matrix for **any likelihood** (even non-Gaussian ones) is defined as

$$F_{\lambda\lambda'} = -\left\langle \frac{\partial^2 \ln \mathcal{L}}{\partial\lambda\, \partial\lambda'} \right\rangle \Big|_{\lambda=\hat{\lambda}} \tag{10.37}$$

In general, the Fisher matrix sets a lower bound on the possible error bar, called the **Rao-Cramer bound**. The bound is

$$\sigma_\lambda \geq \frac{1}{\sqrt{F_{\lambda\lambda}}} \qquad \text{unmarginalized} \tag{10.38}$$

$$\sigma_\lambda \geq \sqrt{(F^{-1})_{\lambda\lambda}} \qquad \text{marginalized} \tag{10.39}$$

For maximum likelihood estimators and large enough data sets the Rao Cramer bound is saturated, which is why we wrote an equal sign in the previous section. Some details about the Rao-Cramer bound can be found in Appendix A of 1001.4707. In cosmology we usually assume that the Rao-Cramer bound is saturated in our forecasts.

10.5.2 Priors, subsets, and combining Fisher matrices

If we want to combine the Fisher forecast of two experiments, we can add their Fisher matrices. This has to be done before marginalization over nuisance parameters (unless these are independent for the two experiments). If we want to add a Gaussian prior to a parameter in the Fisher matrix

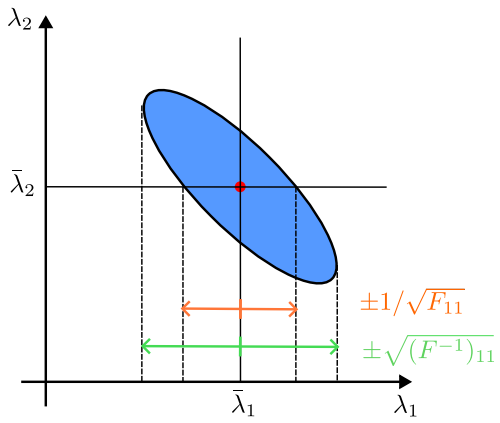


Figure 13. Marginalized and unmarginalized error on parameter λ_1 in a 2 parameter Fisher matrix.

(for example from a different measurements), we add a term to the corresponding diagonal Fisher matrix element

$$F_{\lambda\lambda} \to F_{\lambda\lambda} + \frac{1}{\sigma_\lambda^2}. \tag{10.40}$$

Sometimes we want to marginalize over a subset of the parameters only. This can be done as follows:

- invert F
- remove the rows and columns of parameters we want to marginalize over, to arrive at a smaller matrix which we call G^{-1}
- invert this smaller matrix to get the new Fisher matrix G

A code that helps automatize these operation is pyfisher (https://pyfisher.readthedocs.io/en/latest/). Note that numerical inversion of a Fisher matrix can fail if it is not well conditioned, for example due to numerical inaccuracies.

A common practice is to marginalize all but 2 parameters and then plot their **Fisher ellipses**. An illustration is shown in Fig. 13. A review of Fisher forecasting that explains drawing ellipses is given in 0906.4123.

10.5.3 Fisher matrix for a general Gaussian distribution

Above we have given expressions for the Fisher matrix of a Gaussian distribution with a parameter independent covariance matrix. This is not always a correct treatment (in particular not for the likelihood of a Gaussian field as in Sec. 10.3.2). For a general Gaussian

$$L = \frac{1}{(2\pi)^{n/2}\det \mathbf{C}^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{d}-\boldsymbol{\mu}(\boldsymbol{\lambda}))^T\mathbf{C}^{-1}(\boldsymbol{\lambda})(\mathbf{d}-\boldsymbol{\mu}(\boldsymbol{\lambda}))\right] \tag{10.41}$$

one can show that the Fisher matrix is

$$F_{ij} = \boldsymbol{\mu}_{,i}^T\mathbf{C}^{-1}\boldsymbol{\mu}_{,j} + \frac{1}{2}\mathrm{Tr}[\mathbf{C}^{-1}\mathbf{C}_{,i}\mathbf{C}^{-1}\mathbf{C}_{,j}] \tag{10.42}$$

where $,i$ is the partial derivative with respect to λ_i. Here I have used matrix notation rather than index notation (sum over i, j). This form of the Fisher matrix appears very often in cosmology papers. A derivation of this result can be found for example in arxiv:0906.0664. If our likelihood is a Gaussian random field with mean zero, the first term is zero.

10.6 Sampling the posterior: MCMC

A typical posterior in cosmology might have between 10 and 100 parameters (physical parameters plus nuisance parameters). Above we discussed evaluating the posterior $\mathcal{P}(\boldsymbol{\lambda}|\mathbf{d})$. In the case of a power spectrum analysis, the computational cost to evaluate the posterior is usually dominated by evaluating $P^{theo}(k_i, \mathbf{p})$, for example with CAMB. A single evaluation of the posterior may take seconds to minutes. It is not possible to simply evaluate the posterior on a grid in high dimensions, to find the region of large posterior. For example, if we wanted to sample each axis of the posterior with 10 points and we had 20 parameters, we would need 10^{20} calls of the posterior, which is completely impossible even with supercomputing.

Fortunately we don't need to evaluate the posterior everywhere. In most of the regions it is vanishingly small. Instead we want to find the region where the posterior is large. You may first think to descend the gradient of the (negative) log posterior function to find its minimum, and indeed this is a possible approach when your likelihood is differentiable (either numerically or analytically, or using auto-differentiation). Note however that there could be multiple minima if the posterior is "multimodal". Assuming a single minimum, we can find the MAP by gradient descent. However, finding the MAP is not enough. We also want to determine error bars and covariances of the parameters, and be able to marginalize over parameters that we are not interested in. Again, this is difficult to do at the level of the posterior function in high dimensions (sometimes approximations are possible).

However there is a much better approach, which can deal with with the ~ 10 to ~ 100 parameters of a common analysis in cosmology: **sampling from the posterior**. There are many such sampling algorithms, the most popular being variants of **Markov Chain Monte Carlo**. If we have an algorithms that can sample from the posterior, i.e. give us sets of parameters $\boldsymbol{\lambda}$ that are likeli under the posterior, with the right probability, we can use these samples to obtain the **posterior mean and variance of these parameters**:

$$\bar{\lambda}_a = \frac{1}{m_{\text{sample}}} \sum_{i=1}^{m_{\text{sample}}} \lambda_a^i \tag{10.43}$$

$$\text{Var}[\lambda_a] = \frac{1}{m_{\text{sample}} - 1} \sum_{i=1}^{m_{\text{sample}}} (\lambda_a^i - \bar{\lambda}_a)^2 \tag{10.44}$$

We can also make the famous **corner plots** that are often shown in cosmology papers (see for example the Planck results in 1807.06209 Fig. 5). A key property of MCMC is that it scales approximately linearly with the number of parameters, so we can do quite high-dimensional problems. MCMC methods do not require the target distribution (often the posterior distribution in Bayesian inference) to be normalized. However, they do require the ability to evaluate the unnormalized version of the target distribution up to a constant factor.

How does MCMC sampling work? There is a really nice discussion in Dodelson-Schmidt Sec. 14.6 which I will briefly summarize. A Markov Chain is an algorithm where we draw a new sample λ' from λ, but without considering earlier samples. The algorithm is completely described by the conditional probability $K(\lambda'|\lambda)$ that takes us from a sample λ to the next one, λ'. The fundamental requirement on K, in order for the MCMC sampler to sample from the right posterior, is called **detailed balance**:

$$P(\lambda)K(\lambda'|\lambda) = P(\lambda')K(\lambda|\lambda') \tag{10.45}$$

This means that the rate for the forward reaction $\lambda \to \lambda'$ is the same as for the reverse reaction $\lambda' \to \lambda$, which means we have reached an equilibrium distribution. If we start with a distribution of λ that follows $P(\lambda)$, then an algorithm that obeys detailed balance will stay in this distribution. Further, if you start with an arbitrary sample $\lambda_{\rm init}$, after drawing sufficiently many samples, the algorithm will end up in distribution and have forgotten about its starting point (in the same way as we can reach thermodynamic equilibrium from any initial conditions if we wait long enough). This is called the **burn in phase** of MCMC. MCMC is closely connected to thermodynamics, where an equilibrium distribution loses its memory of the initial conditions.

There are different choices for $K(\lambda'|\lambda)$ that obey detailed balance. A common choice is the **Metropolis Hastings algorithm**. In this algorithm, we draw the next parameter sample from a Gaussian, symmetric around the current parameter sample. This sample is then **accepted** with a probability given by

$$p_{\rm acc}(\lambda', \lambda) = \min\left(\frac{P(\lambda')}{P(\lambda)}, 1\right) \tag{10.46}$$

If the new sample is not accepted, we repeat the previous step in the chain. You can check that this procedure obeys detailed balance. The free parameter here is the width of the Gaussian from which the next parameter is draw. If it is too small, the sampler will take a long time to map out the PDF and may get stuck in local minima. If it is large, the sampler will have a low acceptance rate since most samples will be very unlikely. Many algorithms adjust this value dynamically. A good acceptance rate is about 1/3.

The most popular sampler (currently) in cosmology is called **emcee**, which implements an algorithm called "Affine Invariant Markov Chain Monte Carlo (MCMC) Ensemble sampler". Emcee and other popular algorithms use several so called **walkers** which sample from the PDF in parallel. In practice it is usually not important that you understand your MCMC algorithm at a fundamental level, but it is critical that you use it correctly:

- We need to **discard samples** from the burn-in phase. One can often clearly see the burn-in phase in the chain plots coming from the sampler (see Sec. 11 for an example).

- MCMC Samples are not statistically independent. It takes a while until the "memory" of a sample is forgotten. This is called the **auto-correlation length**. One can pick one sample per auto-correlation length for analysis. This is called **thinning of the chain**. Samplers usually come with some estimator of the auto-correlation length.

- Sometimes one can misjudge the convergence and auto-correlation length of an MCMC chain. Chains may be slowly drifting or even oscillating, without being noticeable at the chain length we probed. There is no absolutely guaranteed method to avoid such problems.
- The many Monte Carlo walkers (typically 20 or more) should give statistically equivalent samples. Comparing the different chains and their "mixing" helps judging the convergence of the MCMC, for example using the **Gelman-Rubin statistic**.
- Chain convergence is slowed by degeneracies in the posterior. In such a case, a change of variables is helpful. Some samplers in cosmology such as CosmoMC have functions to deal with this problem for power spectrum analysis.

A typical length of an MCMC chain could be 100.000 samples. Roughly speaking, for an auto-correlation length of 100 samples this would give us 1000 independent samples (see the emcee documentation for best practices of auto-correlation analysis and thinning). We'll see example MCMC results in Sec. 11.

A common question is what prior we should use. Common choices are

- Flat priors in some window (constant probability per $d\lambda$). This is the most common case.
- Priors that are flat in the log of λ in some window (constant probability per $d\ln\lambda$). This is useful if we are unsure even about the order of magnitude of the parameter.
- Priors that are Gaussian, particularly coming from a previous independent measurement.

Note first that if the data is very informative, then the likelihood will completely dominate the posterior and the prior becomes irrelevant (as long as it is nonzero at the maximum of the likelihood). Conversely if the data is weak, the choice of prior changes the result substantially. In that case no strong measurement can be made. The main reasons to put an informative prior are

- If we have a strong and trustworthy measurement for a parameter from a different uncorrelated experiment and we want to include that information (usually as a Gaussian prior).
- If we have a physical theory that gives a reliable prior, such as that Ω_m cannot be negative or that the primordial curvature perturbations are Gaussian.

10.7 Other algorithms beyond MCMC

While MCMC still dominates astrophysics, there are other inference algorithms that are becoming important. For very high dimensional problems, say more than 100 parameter, MCMC becomes too slow to converge. This is because random jumps become more and more unlikely to result in accepted samples. Instead, a sampling algorithm that has knowledge of the gradient of the function can be much more efficient. Such an algorithm is **Hamiltonian Monte Carlo** (also called Hybrid Monte Carlo). This approach is in particular used when we want to sample over field variables, such as the lensing potential or the initial conditions of the universe. We'll discuss this more in Sec. 28.3. Another form of MCMC which is useful to know about is **Gibbs sampling**, used for example in the Planck low-ℓ likelihood.

A different approach that is starting to be used in astrophysics is **variational inference**. In variational inference, one fits a simpler **variational distribution** to approximate the true posterior. This is useful in cases where it would be too expensive to sample from the true posterior. However we still need to be able to evaluate the unnormalized posterior at some points to fit the variational distribution.

10.8 Goodness of fit

There is one more crucial topic of statistics that we need to discuss: Goodness of fit of the model, and the related topic of model testing.

Let's start with the χ^2 distribution. This is the distribution of the sum of squares of a Gaussian. If X_i are Gaussian random variables with mean zero and variance one, then the sum of n squares

$$Y = X_1^2 + X_2^2 + \cdots + X_n^2 \tag{10.47}$$

has a chi-squared distribution with n degrees of freedom with the probability distribution

$$P(Y) = \frac{1}{2^{n/2}\Gamma(n/2)} Y^{n/2-1} e^{-Y/2} \tag{10.48}$$

The χ^2 distribution has the following properties:

- the mean of $P(Y)$ is equal to the number of degrees of freedom n.
- the variance of $P(Y)$ is equal to $2n$.
- when $n \gg 1$, the chi-squared distribution starts to look like the Gaussian distribution, with mean n and variance $2n$.

For example, the power spectrum estimator uses the sum of squares of Gaussian modes $\delta(\mathbf{k})$ and thus is χ^2 distributed (and approximately Gaussian for enough modes).

The χ^2 distribution arises as the sum of squares of the **residuals** $\mathbf{d} - \mathbf{d}_{\rm model}$ in a **least squares model fitting**:

$$\chi^2_{k_{\rm dof}} = [\mathbf{d} - \mathbf{d}_{\rm model}(\boldsymbol{\lambda})]^T C^{-1} [\mathbf{d} - \mathbf{d}_{\rm model}(\boldsymbol{\lambda})], \tag{10.49}$$

This is also the form of a Gaussian likelihood with parameter independent data covariance C (such as in our power spectrum likelihood).

If our **model is a good fit to the data** we should have

$$\chi^2_{k_{\rm dof}} \approx k_{\rm dof} \tag{10.50}$$

where

$$k_{\rm dof} = N_{\text{data points}} - N_{\text{fitted parameters}} \tag{10.51}$$

As a consistency check, if we have as many model parameters as data parameters we should get a perfect fit without residuals. We can use the properties of the χ^2 distribution such as the variance

$$\mathrm{Var}(\chi^2_{k_{\rm dof}}) = 2k_{\rm dof} \tag{10.52}$$

or the P-value to quantify whether the fit is good. So for a good fit we would have

$$\chi^2 \approx k_{\rm dof} \pm \sqrt{2k_{\rm dof}} \tag{10.53}$$

If the fit is good, this implies for example that 68% of the data points are withing the 1σ error. Otherwise we see how many sigmas we are away from a good fit.

If the χ^2 is higher than expected it can mean either

- that the model does not fit the data
- or that we have underestimated the data error (wrong covariance matrix)
- or that there are systematic errors in our data
- or that the errors in our data are not Gaussian.

It can also happen that χ^2 is smaller than expected if we overestimated our data error.

One sometimes also defines the **reduced χ^2**:

$$\chi^2_{red} = \frac{\chi^2}{N_{\text{data points}}} \tag{10.54}$$

In the common case that $N_{\text{data points}} \gg N_{\text{fitted parameters}}$ the reduced χ^2 should be around 1.

10.9 Model comparison

The simplest way to compare how well models fit is to compare their χ^2. If we have two models A and B we can calculate the difference in their χ^2

$$\Delta\chi^2 = \chi^2_B - \chi^2_A \tag{10.55}$$

Let's assume that A is a subset of B, for example A is ΛCDM and B is ΛCMB extended with a free equation of state parameter for dark energy (so that $w = -1$ in A but free in B). Of course the fit must be better or equal in model B. If the $\Delta\chi^2$ is large (negative), then model B is a much better fit than model A. According to **Wilk's theorem** $\Delta\chi^2$ can be quantified by a χ^2 distribution with degrees of freedom $k_{dof,B} - k_{dof,A}$ (which would be 1 in the dark energy example).

There are also model comparison tests for cases in which B is NOT a subset of A, in particular the **Bayesian Information Criterium (BIC) and the Akaike Information Criterion (AIC)** (see e.g. Huterer's book).

The most consistent, but computationally challenging, way to compare models is using the Bayesian approach. Here we calculate the **Bayes Factor**

$$B_{AB} = \frac{P(d|A)}{P(d|B)} \tag{10.56}$$

from the evidence ratios of the two models. If we have priors on the models, we get the **posterior odds**

$$\frac{P(A|d)}{P(B|d)} = B_{AB}\frac{P(A)}{P(B)} \tag{10.57}$$

The Bayes factor is difficult to evaluate since we need to integrate over the entire model parameter space (Eq.(10.11)). According to the **Jeffreys' scale**, for equal prior models, $B > 3$ is considered weak evidence, $B > 12$ is considered moderate evidence and $B > 150$ is considered strong evidence (less than 1/150 chance probability) for one model over the other.

11 Analyzing an N-body simulation

The section is currently covered in the Colab notebook and problem set. Will be added here later.

We are going to analyze a few Quijote simulations (https://quijote-simulations.readthedocs.io/). These are a large set of 45,000 simulations, covering different cosmological parameters. The side length of the box of each simulation is 1Gpc/h. A large amount of 17,100 simulations is generated for a fiducial Planck cosmology. This large number of simulations is for example useful to determine covariance matrices. There are many other simulations with different cosmological parameters. In total, Quijote contains 700 terabytes of data and required 35 million CPU core hours.

In an N-body simulation, we track a set of individual particles as they interact gravitationally. In Quijote, we have 512^3 particles on a volume of $1(\mathrm{Gpc}/h)^3$. These partciles are initially placed on a regular grid, with slight displacements that account for the primordial density perturbations. Then we solve the equations of Newtonian gravity, with numerical tricks to speed the process up. We are going to talk more about how N-body simulations work in Sec. 23.

Part III
Cosmic Microwave Background

Due to its linearity, the primary CMB is the cleanest probe of cosmology we have. While the **primary CMB** temperature perturbations have been mapped out almost to cosmic variance, upcoming experiments will measure E-mode polarization in more detail, while primordial **B-mode polarization** has not been detected at all and is a major science target. **Secondary CMB anisotropies**, which are induced by the re-scattering of CMB photons on charges, and by gravitational lensing, have been detected but are far form being fully exploited for cosmology and astrophysics. In this section I will focus more on secondary anisotropies and data analysis methods, and be brief on primary CMB physics which is interesting but mostly worked out. We will also, for the first time in this course, discuss "real world" data analysis issues such as detector noise and the mask, which makes even power spectrum estimation rather complicated. Finally I will discuss the topic of foreground cleaning, which is relevant also for many other types of data.

Further reading

The general references of Unit 1 all contain a discussion of the CMB. In addition I recommend

- Anthony Challinor's 2015 lecture notes Part III Advanced Cosmology - Physics of the Cosmic Microwave Background.
- Ruth Durrer's textbook on CMB physics.

Both sources go into far more detail than this course.

We will also use the excellent computational notebooks provided by the CMB data analysis summer school https://sites.google.com/cmb-s4.org/summer-school-2021/notebooks?authuser=0.

12 Random fields on the sphere

12.1 Spherical harmonics

Consider a random real scalar field on the 2-sphere, denoted $f(\hat{n})$, where $\hat{n}$ is a unit vector pointing in the direction. **Spherical harmonics** are a basis to represent any (well-behaved) function on the sphere, in close analogy to the Fourier expansion in Euclidean space. The expansion in spherical harmonics is given by

$$f(\hat{n}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} f_{lm} Y_{lm}(\hat{\mathbf{n}}) \tag{12.1}$$

The Y_{lm} are familiar from quantum mechanics as the position-space representation of the eigenstates of the angular momentum operators $\hat{L}^2 = -\nabla^2$ and $\hat{L}_z = -i\partial_\phi$ (setting $\hbar = 1$):

$$\nabla^2 Y_{lm} = -l(l+1) Y_{lm}, \tag{12.2}$$

$$\partial_\phi Y_{lm} = im Y_{lm} \tag{12.3}$$

with l an integer ≥ 0 and m an integer with $|m| \leq l$.

The spherical harmonics are orthonormal over the sphere,

$$\int d\hat{n} Y_{lm}(\hat{\mathbf{n}}) Y^*_{l'm'}(\hat{\mathbf{n}}) = \delta_{ll'}\delta_{mm'} \tag{12.4}$$

The **spherical multipole coefficients** of $f(\hat{n})$ are

$$f_{lm} = \int d\hat{\mathbf{n}} f(\hat{n}) Y^*_{lm}(\hat{\mathbf{n}}) \tag{12.5}$$

There are various phase conventions for the Y_{lm}; here we adopt $Y^*_{lm} = (-1)^m Y_{l,-m}$ so that $f^*_{lm} = (-1)^m f_{l,-m}$ for a real field.

The spherical harmonics are products of associated Legendre polynomials and an azimuthal phase factor:

$$Y_{lm}(\theta, \phi) = \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}} P^m_l(\cos\theta) \exp^{im\phi}$$

The correspondence between multipoles and angles is $l \sim \pi/\Theta$ where Θ is in radians.

12.2 2-point function

It can be shown that for the 2-point correlation function of the f_{lm} to be rotationally invariant, it must have the form

$$\langle f^*_{lm} f_{l'm'} \rangle = C_l \delta_{ll'} \delta_{mm'} \tag{12.6}$$

The real quantity C_l is the **angular power spectrum** of $f(\hat{n})$. Gaussian random fields on the sphere are again fully determined by their two-point function (i.e., their covariance) and hence their angular power spectrum.

As in the Euclidean case, there is a relation between the power spectrum and the 2-point function in position space. We can calculate it as follows:

$$\langle f(\hat{\mathbf{n}}) f(\hat{\mathbf{n}}') \rangle = \sum_{lml'm'} \langle f^*_{lm} f_{l'm'} \rangle Y_{lm}(\hat{\mathbf{n}}) Y^*_{l'm'}(\hat{\mathbf{n}}') \tag{12.7}$$

$$= \sum_{lm} C_l Y_{lm}(\hat{n}) Y^*_{lm}(\hat{\mathbf{n}}') \tag{12.8}$$

$$= \sum_{l} \frac{2l+1}{4\pi} C_l P_l(\hat{\mathbf{n}} \cdot \hat{\mathbf{n}}') \tag{12.9}$$

$$= C(\theta), \tag{12.10}$$

where $\mu = \hat{\mathbf{n}} \cdot \hat{\mathbf{n}}' = \cos\theta$ and we used the **addition theorem for spherical harmonics**,

$$\sum_m Y_{lm}(\hat{\mathbf{n}}) Y^*_{lm}(\hat{\mathbf{n}}') = \frac{2l+1}{4\pi} P_l(\hat{\mathbf{n}} \cdot \hat{\mathbf{n}}'). \tag{12.11}$$

We see that the 2-point function depends only on the angle, as we require from isotropy. The inverse relation, going from position space to momentum space, is

$$C_l = 2\pi \int_{-1}^{1} d\cos\theta \, C(\theta) P_l(\cos\theta). \tag{12.12}$$

In analogy to what we did in Eq.(9.43), we can calculate the variance of the field

$$C(0) = \sum_l \frac{2l+1}{4\pi} C_l \approx \int \frac{l(l+1)C_l}{2\pi}\, d\ln l. \tag{12.13}$$

The quantity

$$D_l = \frac{l(l+1)C_l}{2\pi} \tag{12.14}$$

is commonly plotted and gives the contribution to the variance per log range in l. For a scale invariant power spectrum we have $D_l = \text{const}$.

12.3 Discretization with HEALPix and Pixell

Unlike in the Euclidean case, pixelizing a sphere is not so straight forward. For example, to put a measured CMB temperature map on a computer, we need to somehow store the field value at fixed positions (or pixels) on the sphere. The industry standard to achieve this has been HEALPIX, although newer alternatives exist.

To quote from the healpix paper (arxiv:0409513): "The simplicity of the spherical form belies the intricacy of global analysis on the sphere. There is no known point set which achieves the analogue of uniform sampling in Euclidean space and allows exact and invertible discrete spherical harmonic decompositions of arbitrary but band-limited functions. Any existing proposition of practical schemes for the discrete treatment of such functions on the sphere introduces some (hopefully small) systematic error dependent on the global properties of the point set. The goal is to minimise these errors and faithfully represent deterministic functions as well as realizations of random variates both in configuration and Fourier space while maintaining computational efficiency."

The approach of the paper is to propose the **Hierarchical Equal Area, iso-Latitude Pixelisation (HEALPix)** of the sphere. This approach can be used conveniently with the **healpy** package in python. Data from cosmological surveys, such as Planck, is often delivered in the healpix format. More recently, a different library called **Pixell** https://github.com/simonsobs/pixell is also being used.

12.4 Projections of 3D random fields to the sphere

In cosmology we often need to project 3d fields onto 2d spheres. For example, we do perturbation theory in Euclidean space, but we observe on the light cone which is sperically symmetric.

For example, the CMB is approximately given by a projection of the 3-dimensional potential fluctuations onto a sphere centered around our current position with comoving radius $\chi_\star \sim 13800$ Mpc^{-1} (the distance that light travelled since recombination). This is not precisely correct, since recombination has a finite width, but we will take this into account later. The spherical projection we discuss now is needed more generally (e.g to calculate the galaxy power spectrum on the light cone).

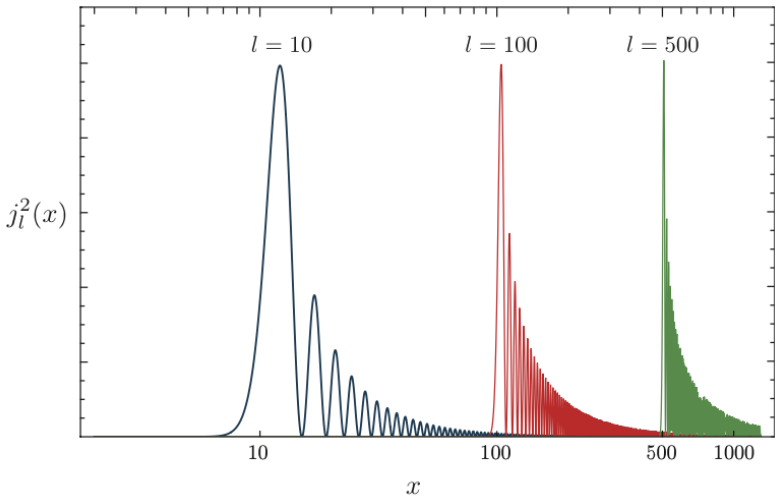


Figure 14. Examples of the spherical Bessel function $j_l(x)$ (from Baumann's Cosmology Lectures).

We project a 3d random field $F(\mathbf{x})$ over a 2-sphere of radius r, centred on the origin, to form the field $f(\hat{\mathbf{n}}) = F(r\hat{\mathbf{n}})$. Expanding $F(\mathbf{x})$ in Fourier modes, we have

$$f(\hat{\mathbf{n}}) = \int \frac{d^3k}{(2\pi)^3} F(\mathbf{k}) e^{ikr\hat{\mathbf{k}}\cdot\hat{\mathbf{n}}} \tag{12.15}$$

$$= 4\pi \sum_{lm} i^l \left(\int \frac{d^3k}{(2\pi)^3} F(\mathbf{k}) j_l(kr) Y^*_{lm}(\hat{\mathbf{k}}) \right) Y_{lm}(\hat{\mathbf{n}}) \tag{12.16}$$

where we have used the **Rayleigh plane-wave expansion**

$$e^{i\mathbf{k}\cdot\mathbf{x}} = \sum_l i^l (2l+1) j_l(kr) P_l(\hat{\mathbf{k}}\cdot\hat{\mathbf{n}}) \tag{12.17}$$

$$= 4\pi \sum_{lm} i^l j_l(kr) Y^*_{lm}(\hat{\mathbf{k}}) Y_{lm}(\hat{\mathbf{n}}) \tag{12.18}$$

Here, $r = |\mathbf{x}|$ and $j_l(kr)$ are the **spherical Bessel functions**. Extracting the spherical multipoles of $f(\hat{\mathbf{n}})$ from above, we have

$$f_{lm} = 4\pi i^l \int \frac{d^3k}{(2\pi)^3} F(\mathbf{k}) j_l(kr) Y^*_{lm}(\hat{\mathbf{k}})$$

The spherical Bessel function peak at $kr = l$. This means that the observed multipoles l mainly probe spatial structure in the 3D field $F(\mathbf{x})$ with wavenumber $k \approx l/r$, but higher k also contribute. The Bessel functions are oscillatory and need to be evaluated precisely. Examples are plotted in Fig.14. Evaluating such Bessel function integrals numerically in cosmology is very common and there are methods developed to speed them up, in particular the **FFTlog algorithm** (e.g. 1705.05022).

Now we relate the power spectrum of the projected field to the power spectrum of the 3d field:

$$\langle f_{lm} f^*_{l'm'} \rangle = (4\pi)^2 i^l (-i)^{l'} \int \frac{d^3k}{(2\pi)^3} \int \frac{d^3k'}{(2\pi)^3} \left\langle F(\mathbf{k}) F^*(\mathbf{k}') \right\rangle j_l(kr) j_{l'}(k'r) Y^*_{lm}(\hat{\mathbf{k}}) Y_{l'm'}(\hat{\mathbf{k}'}) \tag{12.19}$$

$$= 4\pi i^l (-i)^{l'} \int \frac{dk k^2}{2\pi^2} P_F(k) j_l(kr) j_{l'}(kr) \int d\hat{\mathbf{k}} Y^*_{lm}(\hat{\mathbf{k}}) Y_{l'm'}(\hat{\mathbf{k}}) \tag{12.20}$$

$$= 4\pi \delta_{ll'} \delta_{mm'} \int \frac{dk k^2}{2\pi^2} P_F(k) j_l^2(kr) \tag{12.21}$$

where we used $\langle F(\mathbf{k}) F^*(\mathbf{k}') \rangle = (2\pi)^3 P_F(k) \delta^3(\mathbf{k} - \mathbf{k}')$. Thus we get

$$C_l = 4\pi \int \frac{dk k^2}{2\pi^2} P_F(k) j_l^2(kr) \tag{12.22}$$

$$= 4\pi \int d\ln k \, \Delta^2(k) j_l^2(kr) \tag{12.23}$$

where in the last step we defined the dimensionless power spectrum as in Eq.(9.30).

12.5 Power spectrum estimator and covariance

Power spectrum estimation works in close analogy to the Euclidean case we discussed above. The power spectrum estimator is simply

$$\hat{C}_l = \frac{1}{2l+1} \sum_m f^*_{lm} f_{lm} \tag{12.24}$$

Let's first check that the estimator is unbiased:

$$\langle \hat{C}_l \rangle = \frac{1}{2l+1} \sum_m \langle f_{lm} f^*_{lm} \rangle = \frac{1}{2l+1} \sum_m C_l = C_l. \tag{12.25}$$

Now we calculate the variance of the estimator

$$\mathrm{cov}(\hat{C}_l, \hat{C}_{l'}) = \langle \hat{C}_l \hat{C}_{l'} \rangle - \langle \hat{C}_l \rangle \langle \hat{C}_{l'} \rangle \tag{12.26}$$

$$= \frac{1}{(2l+1)(2l'+1)} \sum_{m,m'} \langle f_{lm} f^*_{lm} f_{l'm'} f^*_{l'm'} \rangle - \langle C_l \rangle \langle C_{l'} \rangle \tag{12.27}$$

$$= \frac{1}{(2l+1)(2l'+1)} \sum_{m,m'} 2C_l^2 \delta_{mm'} \delta_{ll'} \tag{12.28}$$

$$= \frac{2C_l^2}{2l+1} \delta_{ll'}. \tag{12.29}$$

where we used Wick's theorem. We see that for a Gaussian field the covariance matrix is diagonal (this does not hold if we have an experimental mask that breaks isotropy as we shall discuss soon). Our precision is limited by the number of available modes, which gives the **cosmic variance** error

$$\frac{\Delta C_l}{C_l} = \sqrt{\frac{2}{2l+1}}$$

which is inversely proportional to the square root of the number of modes $N_{\mathrm{mode}} = 2l + 1$.

12.6 Flatsky coordinates

For an experiment that covers only a small part of the sky, spherical harmonics are not neccessary. Instead, one can use **flat-sky** coordinates. These coordinates are defined at the tangential surface to the sphere at some point in the sky. In flatsky coordinates, we can use an ordinary 2-d Fourier transfrom:

$$f(\hat{\mathbf{n}}) = \int \frac{d^2\mathbf{l}}{(2\pi)^2} f_{\mathbf{l}}\, e^{i\mathbf{l}\cdot\mathbf{x}}. \tag{12.30}$$

and inverse Fourier transform

$$f(\boldsymbol{l}) = \int d^2\hat{\mathbf{n}}\, f(\hat{\mathbf{n}})\, e^{i\mathbf{l}\cdot\mathbf{x}}. \tag{12.31}$$

One can formally relate the spherical harmonics expression to the Fourier modes by taking the large-l limit of the Legendre polynomials (see e.g. Liddle, Lyth book Sec 10.3). The correspondence between the power spectra

$$\langle f_{\ell m} f^*_{\ell' m'}\rangle = \delta_{\ell\ell'}\delta_{mm'}C_\ell, \quad , \quad \langle f(\mathbf{l}) f^*(\mathbf{l}')\rangle \equiv (2\pi)^2\delta^{\mathcal{D}}(\mathbf{l}-\mathbf{l}')C_\ell^{\rm flat} \tag{12.32}$$

is simply

$$C_\ell = C_\ell^{\rm flat} \tag{12.33}$$

To work with the flatsky approximation numerically we need to discretize the fourier transform in the same way as we did in the 3d field.

13 Primary CMB power spectrum

In this course we don't cover relativistic perturbation theory, and instead simply use the results from CAMB or CLASS. Let's have a look at the results and discuss the main features.

The CMB temperature is given by

$$\Theta(\hat{n}) \equiv \frac{T(\hat{n}) - T_0}{T_0} = \sum_{\ell m} a_{\ell m} Y_{\ell m}(\hat{n})\,, \tag{13.1}$$

where $T_0 \sim 2.7$K is the mean temperature and

$$a_{\ell m} = \int \mathbf{d}\Omega\, Y^*_{\ell m}(\hat{n})\Theta(\hat{n})\,. \tag{13.2}$$

The CMB power spectrum is then

$$\langle a^*_{\ell m} a_{\ell' m'}\rangle = C_\ell^{TT}\delta_{\ell\ell'}\delta_{mm'} \tag{13.3}$$

13.1 Transfer functions and line-of-sight solution

The linear evolution which relates $\mathcal{R}$ and ΔT is given by the **transfer function** $\Delta_{T\ell}(k)$ through the k-space integral

$$a_{\ell m} = 4\pi(-i)^\ell \int \frac{d^3k}{(2\pi)^3}\, \Delta_{T\ell}(k)\, \mathcal{R}_{\mathbf{k}}\, Y_{\ell m}(\hat{\mathbf{k}}) \tag{13.4}$$

Using the addition theorem we get

$$C_\ell^{TT} = \frac{2}{\pi}\int k^2 dk \underbrace{P_{\mathcal{R}}(k)}_{\text{inflation}} \underbrace{\Delta^2_{T\ell}(k)}_{\text{evolution,projection}} \quad . \tag{13.5}$$

The transfer funtions are calculated by CAMB or CLASS.

On large scales, modes were still outside of the horizon at recombination (the **Sachs-Wolfe regime**) the transfer function $\Delta_{T\ell}(k)$ is simply the Bessel function generated by the spherical projection we discussed above

$$\Delta_{T\ell}(k) = \frac{1}{3} j_\ell(k[\tau_0 - \tau_{\rm rec}]) \,. \tag{13.6}$$

The angular power spectrum on large scales (small ℓ) therefore is

$$C_\ell^{TT} = \frac{2}{9\pi}\int k^2 dk\, P_{\mathcal{R}}(k)\, j_\ell^2(k[\tau_0 - \tau_{\rm rec}]) \,. \tag{13.7}$$

This is sometimes called the "snapshot approximation" or "instantaneous recombination approximation".

On smaller scales, we need to take into account that recombination does not happen instantaneously, but rather in a finite time window (with a comoving width of about 10 Mpc). Further, some CMB perturbations on large scales are also sourced at later times in the universe (in particular during reionization). To take these effects into account in the transfer functions, one needs to do a **line-of-sight** integral over a "source term" $S(k,\tau)$ as follows

$$\Delta_{T\ell}(k) = \int_0^{\tau_0} d\tau S(k,\tau) j_\ell(k\tau) \,. \tag{13.8}$$

The source term $S(k,\tau)$ comes from solving the Boltzmann equation, and the Bessel function is again due to the spherical projection. CAMB and CLASS are calculating these transfer functions for us. The full details are explained in one of the famous papers of cosmology, astro-ph/9603033, which proposed the method.

13.2 The physics of the CMB Power spectrum

Let's have a look at the CMB power spectrum as measured by Planck, ACT and SPT, in Fig. 15. Consider the different regions:

- On the **largest scales** (Region I), modes re-enter the horizon after recombination and thus they do not evolve. This gives an approximately flat power spectrum in D_l. See Sec. 6.6.2 for a discussion of horizon exit and re-entry.

- Intermediate regions (Region II) are dominated by the **baryon acoustic oscillations (BAO)**. The BAO are oscillations in the primordial plasma of photons and electrons.

- On smaller scales (Region III) the primary perturbations are getting exponentially suppressed due to diffusive damping (also called **Silk damping**). As the photons move from over-dense to under-dense regions, they effectively smooth out the fluctuations in the photon-baryon fluid on their typical scattering length scale. This leads to a suppression

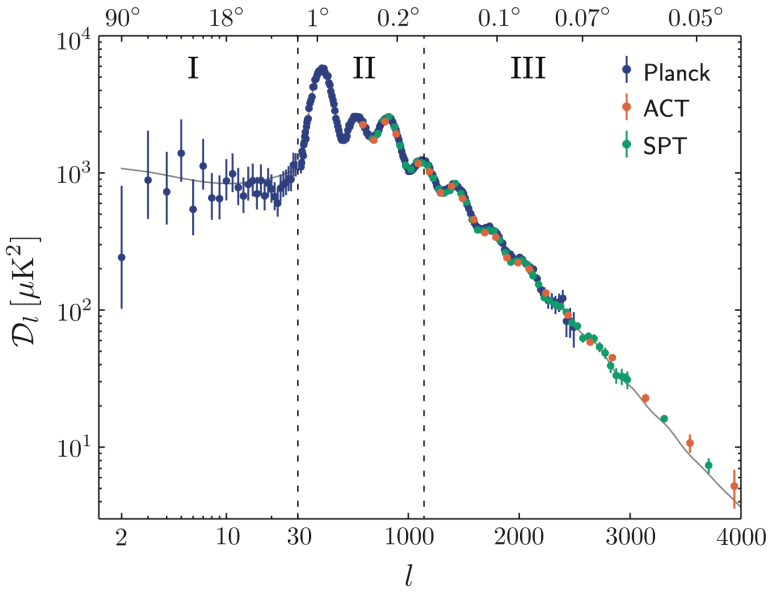


Figure 15. The CMB temperature power spectrum (plot from Baumann's Cosmology Lectures).

of the anisotropy in the CMB at small scales (large multipole moments) On these small scales, secondary anisotropies due to lensing and kSZ start to dominate. We will discuss secondary anisotropies in Sec. 17.

On all of these scales, the anisotropies are generated by three different effects, which appear as terms in the transfer function:

- The **Sachs-Wolfe (SW) effect** is the largest contribution. It combines the temperature inhomogeneity in the primordial plasma (due to the density perturbations) with the redshift of the photons (which have to "climb out of their potential well" when they are in an overdense region). Due to the redshift it turns out that colder CMB spots correspond to higher density regions.
- The **Doppler effect** is the change in photon energy due to scattering off moving electrons.
- The **Integrated Sachs-Wolfe effect (ISW)** describes the additional gravitational redshift due to the evolution of the metric potentials along the line-of-sight. This effect occurs during radiation domination (early ISW) and during dark energy domination (late ISW). The late ISW adds power at very low $l < 10$.

For a plot of the different contributions see for example Baumann's book Fig. 7.7. As was the case for the matter power spectrum, the CMB power spectrum is very sensitive to cosmological parameters. For a nice illustration see Plate 4 in the review astro-ph/0110414v1.

14 Analyzing the CMB power spectrum

We will now discuss how to analyze the CMB power spectrum from a real experiment. There are several experimental complications:

- The experiment has a finite resolution, which is decribed by the **beam**.
- There are sources of **noise** (from the detector and e.g. the atmosphere).
- There is a finite region of the sky that is observed (described by the **mask**), which breaks statistical isotropy.
- There are **foregrounds** such as galactic dust and synchrotron radiation, which obscure the true CMB signal.

To infer cosmological parameters, we need to compare the theory C_l^{theo} prediction (which depends on cosmological parameters in a known way determined by the laws of physics) with the data C_l^{obs}. There are in principle two directions how to approach this problem.

- In **backward modelling**, one first tries to remove the experimental effects from the data, to arrive at a reconstruction of the signal had there been no experimental effects. This reconstructed signal is then compared with the theory. The advantage of this approach is that one can easily compare measurements from different experiments at map level. Most of our discussion below is of this sort.
- In **forward modelling** one models the experimental effects on the theory result. We would model how the theory power spectrum changes due to the experimental effects and compare this $C_l^{theo,forward}$ to the C_l^{obs}. This approach has the advantage that it is easier to propagate errors and one can cleanly separate theory from data.

14.1 Beam and Noise

A real CMB experiment that observes the sky has a finite angular resolution and the detector (and atmosphere) induce noise in the measurement. We will observe a temperature $\theta^{obs}(\hat{\mathbf{n}})$ at a direction $\mathbf{n}$ in the sky which is given by

$$\theta^{obs}(\hat{\mathbf{n}}) = \int d\Omega' \theta(\hat{\mathbf{n}}')\, B(\hat{\mathbf{n}}, \hat{\mathbf{n}}') \, + n(\hat{\mathbf{n}}) \tag{14.1}$$

where $\theta(\hat{\mathbf{n}}')$ is the true CMB temperature signal, $B(\hat{\mathbf{n}}, \hat{\mathbf{n}}')$ is the **beam** or **point-spread function (PSF)** which tells us how the detector reacts to the distribution on the sky, and $n(\hat{\mathbf{n}})$ is **noise** which is uncorrelated with the signal. As we can see, the beam is a **convolution in real space**. The observed a_{lm}^{obs} are then given by

$$a_{\ell m}^{obs} = \int \mathbf{d}\Omega \, Y_{\ell m}^*(\hat{n}) \Theta^{obs}(\hat{n}) \, . \tag{14.2}$$

In harmonic space we can express this as

$$a_{\ell m}^{obs} = \sum_{l'm'} B_{lm,l'm'} a_{l'm'} + n_{lm} \tag{14.3}$$

It is often a good approximation that the beam is constant on the sky and isotropic. In this case one gets

$$a_{\ell m}^{obs} = B_l a_{lm} + n_{lm} \tag{14.4}$$

For a Gaussian beam, the B_l are given by

$$B_l = \exp^{\frac{-l^2}{2}\Theta^2_{beam}} \tag{14.5}$$

where Θ_{beam} is related to the width of the beam. For small l the beam is approximately 1 ($l\Theta_{beam} \ll 1$) while for large l it is approximately zero (i.e. it washes out anoisotropies on these scales). The noise can often be approximated as Gaussian, in which case it is fully determined by the 2-point function

$$\langle n_{lm} n^*_{l'm'} \rangle = N_l \delta_{ll'} \delta_{mm'} \tag{14.6}$$

where N_l is called the **noise power spectrum**. There are various forms of noises as we will discuss. You can often download the B_l and N_l of a CMB experiment such as Planck.

If the noise power spectrum is known (from measurements and modelling of the detector), and the noise is Gaussian, and the beam and noise are isotropic, one can show (e.g. Dodelson 14.4.1) that the unbiased power spectrum estimator is

$$\hat{C}(l) = B_l^{-2} \left(\frac{1}{2l+1} \sum_{m=-l}^{l} \left| a^{obs}_{lm} \right|^2 - N(l) \right) \tag{14.7}$$

and the variance of the estimator is

$$\mathrm{Var}\left[\hat{C}(l)\right] = \frac{2}{2l+1} \left[C(l) + N(l) B_l^{-2} \right]^2 \tag{14.8}$$

Compare this to our results without beam and noise in Eq. 12.26. The variance now consists of the cosmic variance (due to finite mode number) and noise variance (due to noise and beam).

We have been using a continuous function $\theta^{obs}(\hat{\mathbf{n}})$ in this discussion, which assumes that we have discretized the sky with such a high resolution that the pixelization doesn't matter (e.g. large pixel number in HEALPix). If, for computational reasons, we'd have to limit this pixelization we would also have to include the so-called **pixel window function**.

The noise in a CMB experiment is generally a combination of three different types:

- **White noise**, where each pixel has a noise that is drawn from a Gaussian around zero, independent from all other pixels. In Sec. 9.3.3 we have seen that in Fourier space this means that $C_l = \text{const}$. For a satellite like Planck this can be a good approximation.
- **Atmospheric noise**, which grows larger on large angular scales, can be understood in terms of Kolmogorov turbulence. Atmospheric noise is correlated between pixels (but nearly uncorrelated in Fourier space, like the CMB).
- **1/f noise** in the detector, which is also correlated between pixels (but nearly uncorrelated in Fourier space). It leads to a "stripy" noise pattern that depends on the scanning strategy of the experiment. This noise is important on large scales and falls as $1/l$. It turns out that a wide variety of detectors all lead to noise that goes up on large scales on the sky with roughly a $1/f$ spectrum. This noise comes from fluctuations in the instrument and environment over time. One approach to reduce $1/f$ noise is to scan angles in the sky at a faster rate than the time dependence of the detector noise.

We'll illustrate these in a computational notebook from the CMB data analysis summer school linked above.

14.2 Simple power spectrum estimator: Transfer function and bias

The **naive power spectrum estimator**

$$\hat{C}_l^{naive} = \frac{1}{2l+1} \sum_m a_{lm}^* a_{lm} \tag{14.9}$$

will, when applied to a masked field, result in a biased measurement of the true theoretical (unmasked) power spectrum.

As a first step to improve the result, one can **apodize the mask** (and/or use the related technique of **inpainting**), which means that we smooth out the sharp boundaries of the mask. Many possible apodizations have been proposed with different trade-offs of sensitivity loss, coupling of adjacent modes, and ringing. The mask smoothly reduces the signal to zero on the boundary. This also means that by apodization we make our data periodic (since it goes to zero on all sides). Aperiodic maps generate spourious power in the Fourier transform. However, even after apodization our power spectrum estimate remains biased.

Let's first discuss a simple method how to obtain an unbiased measurement from the naive power spectrum of the apodized data. This approach generalizes to a more optimal method we will describe later. The naive $\hat{C}_l^{naive}$ are related to the true Cl by a **transfer function** M (which in addition to the mask includes the beam) and a **noise bias** N_l as follows

$$\langle \hat{C}_l^{naive} \rangle = M_l \langle \hat{C}_l^{unbiased} \rangle + N_l \tag{14.10}$$

The transfer function can be estimated by Monte Carlo as follows:

- First generate a large number of simulations with known power spectrum $C^{unbiased}$ and no noise.
- For these simulations estimate $\hat{C}_l^{naive}$.
- From the pairs of true power spectrum and measured power spectrum calculate the transfer function M_l.

The noise bias can be computed by running noise only simulations through the naive power spectrum estimator and computing the average power spectrum. An example of the whole procedure is given in the CMB summer school notebook on power spectrum estimation (on the flat sky).

A useful approximation is that the measured power spectrum is related to the true power spectrum by the sky area fraction $f_{\rm sky}$ covered by the experiment (a number between 0 and 1):

$$\langle \hat{C}_l^{naive} \rangle \approx f_{\rm sky} \langle \hat{C}_l^{unbiased} \rangle \tag{14.11}$$

This approximation does not take into account mode coupling due to the mask, but it does take into account the reduced survey area, and is especially useful for Fisher forecasting.

14.3 Mask and mode coupling

Now let's analyze the problem in more detail. My discussion follows the review 1909.09375. Our data has a **mask** $W(\mathbf{n})$ (also called **window function** or **weighting function**). In the

simplest case, the mask is a discrete function $W = 1$ for observed pixels, and zero otherwise. More generally, the mask can be apodized and have smooth values between 0 and 1. The window function can be expanded in spherical harmonics as

$$w_{\ell m} = \int d\hat{n} W(\hat{n}) Y^*_{\ell m}(\hat{n}) \tag{14.12}$$

with power spectrum

$$\mathcal{W}_\ell = \frac{1}{2\ell+1} \sum_m |w_{\ell m}|^2 \tag{14.13}$$

Given the mask and the true temperature anisotropy $\Theta(\mathbf{n})$, the spherical harmonic expansion of the temperature anisotropy field can be written as

$$\begin{aligned}
\tilde{a}_{\ell m} &= \int d\hat{n} \Theta(\hat{n}) W(\hat{n}) Y^*_{\ell m}(\hat{n}) && \text{(14.14a)} \\
&= \sum_{\ell' m'} a_{\ell' m'} \int d\hat{n} Y_{\ell' m'}(\hat{n}) W(\hat{n}) Y^*_{\ell m}(\hat{n}) && \text{(14.14b)} \\
&= \sum_{\ell' m'} a_{\ell' m'} K_{\ell m \ell' m'}(W), && \text{(14.14c)}
\end{aligned}$$

where $K_{\ell m \ell' m'}$ is the **coupling kernel** between different modes. The $\tilde{a}_{\ell m}$ are still Gaussian variables, as they are the sum of Gaussian variables (the "true" $a_{\ell m}$ that expand the true Θ). However, the multipole coefficients of the temperature field on the partial sky are not independent anymore, as the sky cut introduces the coupling represented by Eq. 14.14c.

By expanding the mask in spherical harmonics, the coupling kernel can be written as follows:

$$\begin{aligned}
K_{\ell_1 m_1 \ell_2 m_2} &= \int d\hat{n} Y_{\ell_1 m_1}(\hat{n}) W(\hat{n}) Y^*_{\ell_2 m_2}(\hat{n}) && \text{(14.15a)} \\
&= \sum_{\ell_3 m_3} w_{\ell_3 m_3} \int d\hat{n} Y_{\ell_1 m_1}(\hat{n}) Y_{\ell_3 m_3}(\hat{n}) Y^*_{\ell_2 m_2}(\hat{n}) && \text{(14.15b)} \\
&= \sum_{\ell_3 m_3} w_{\ell_3 m_3} (-1)^{m_2} \left[\frac{(2\ell_1+1)(2\ell_2+1)(2\ell_3+1)}{4\pi} \right]^{1/2} && \text{(14.15c)} \\
&\quad \times \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & -m_2 & m_3 \end{pmatrix},
\end{aligned}$$

where we used the Gaunt integral

$$\begin{aligned}
g^{ll'l''}_{mm'm''} &= \int d\Omega\, Y_{lm}(\hat{n}) Y_{l'm'}(\hat{n}) Y_{l''m''}(\hat{n}) && \text{(14.16)} \\
&= \sqrt{\frac{(2l+1)(2l'+1)(2l''+1)}{4\pi}} \begin{pmatrix} l & l' & l'' \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l & l' & l'' \\ m & m' & m'' \end{pmatrix}. && \text{(14.17)}
\end{aligned}$$

which expresses the integral over three spherical harmonics in terms of Wigner 3j symbols. The coupling kernel is singular and therefore Eq. 14.14c cannot be inverted to compute the true $a_{\ell m}$. This makes sense as a small part of the observed sky should not allow us to reconstruct the true entire sky.

14.4 Pseudo-Cl estimator and PyMaster

The standard approach for CMB estimation is the **Pseudo-Cl** approach from astro-ph/0105302. Pseudo-Cl are near optimal in most cases and fast to evaluate. The Pseudo-CL approach is also modestly called the "MASTER" estimator (Monte Carlo Apodised Spherical Transform EstimatoR).

The cut-sky coefficients can be used to define the **pseudo-C_ℓ power spectrum**

$$\tilde{C}_\ell = \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} \tilde{a}_{\ell m} \tilde{a}^*_{\ell m} \, . \tag{14.18}$$

From Eq. 14.22, a relation between the true power spectrum and the pseudo-power spectrum can be derived taking the ensamble average (in the same way as in section 14.2 but now taking into account mode coupling):

$$\begin{aligned}
\langle \tilde{C}_{\ell_1} \rangle &= \frac{1}{2\ell_1+1} \sum_{m_1=-\ell_1}^{\ell_1} \langle \tilde{a}_{\ell_1 m_1} \tilde{a}^*_{\ell_1 m_1} \rangle && \text{(14.19a)} \\
&= \frac{1}{2\ell_1+1} \sum_{m_1=-\ell_1}^{\ell_1} \sum_{\ell_2 m_2} \sum_{\ell_3 m_3} \langle a_{\ell_2 m_2} a^*_{\ell_3 m_3} \rangle K_{\ell_1 m_1 \ell_2 m_2}[W] K^*_{\ell_1 m_1 \ell_3 m_3}[W] && \text{(14.19b)} \\
&= \frac{1}{2\ell_1+1} \sum_{m_1=-\ell_1}^{\ell_1} \sum_{\ell_2} \langle C_{\ell_2} \rangle \sum_{m_2=-\ell_2}^{\ell_2} |K_{\ell_1 m_1 \ell_2 m_2}[W]|^2 && \text{(14.19c)} \\
&= \sum_{\ell_2} M_{\ell_1 \ell_2} \langle C_{\ell_2} \rangle \, . && \text{(14.19d)}
\end{aligned}$$

The last line in Eq. 14.19a can be obtained by expanding the kernel couplings in spherical harmonics and making use of the orthogonality relations of the Wigner-3j symbols. The coupling matrix $M_{\ell_1 \ell_2}$ is thus given by:

$$M_{\ell_1 \ell_2} = \frac{2\ell_2+1}{4\pi} \sum_{\ell_3} (2\ell_3+1) \mathcal{W}_{\ell_3} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ 0 & 0 & 0 \end{pmatrix}^2 . \tag{14.20}$$

which can be evaluated numerically without needing to run simulations. The unbiased power spectrum estimator is then

$$\hat{C}_\ell = \sum_{\ell'} M^{-1}_{\ell\ell'} \tilde{C}_{\ell'} \, . \tag{14.21}$$

If we observe a sufficiently large part of the sky, the coupling matrix $M_{\ell\ell'}$ is invertible. When we see only a smaller part of the sky, the matrix can become singular: some modes end up being in the masked region of the sky. In such a case, it makes sense to bin the ℓ into larger bins, until the matrix becomes invertible again.

The state-of-the art public implementation of the MASTER approach is called **PyMaster** (or NaMaster when not in Python). It is documented here: 1809.09603. PyMaster can do pseudo-Cl on fullsky and flatsky, and also includes polarization (and foreground mode deprojection which we have not yet discussed). The pseudo-Cl formalism also extends in a straight forward way to

the cross-correlation of two fields (as long as we use the same mask for them). The pseudo-Cl are then

$$\tilde{C}_\ell^{ab} = \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} \tilde{a}_{\ell m} \tilde{b}^*_{\ell m} \, . \tag{14.22}$$

for two fields a and b (such as CMB temperature and E-mode polarization).

14.5 Wiener filtering

The pseudo-CL estimator is fast but it is still not fully optimal. A completely optimal power spectrum estimator can be developed by using **inverse covariance filtering** which is also called **Wiener filtering** or $\boldsymbol{C^{-1}}$ **filtering**. Wiener filtering is required for any optimal analysis of survey data (including e.g. non-Gaussianity search), but it is computationally very expensive and thus not always performed. Wiener filtering also replaces (and improves over) mask apodization.

Assuming the data vector $\mathbf{d}$ is the linear sum of signal $\mathbf{s}$ and noise $\mathbf{n}$ with independent covariance matrices S and N, the Wiener filtered data $\mathbf{d}_{WF}$ is defined by

$$\mathbf{d}_{WF} = S(S+N)^{-1}\mathbf{d}. \tag{14.23}$$

For a data set with N pixels, the direct inversion of a dense $N \times N$ covariance matrix is impossible for current CMB maps with millions of pixels. Conjugate gradient solvers are usually employed to perform Wiener filtering of CMB data. But the computational costs are enormous for Planck resolution and Wiener filtering a large ensemble of maps remains very difficult even with large computing resources. An example of a small Wiener filtered CMB map is shown in Fig. 16.

Often it can be assumed that the noise covariance N is diagonal in pixel space, i.e. the noise is assumed uncorrelated between pixels. We can represent the mask as a limiting case of anisotropic noise, by taking the noise level to be infinity in masked pixels. (In a code implementation, we set the corresponding entries of N^{-1} to zero). On the other hand, the signal covariance matrix is diagonal in momentum space. Because the two matrices are not diagonal in any common basis (except in the special case of a fullsky observation without mask, where the noise is also diagonal in momentum spacce), the matrix inversion is computationally hard.

The Wiener filter is the **optimal reconstruction of the signal given the noise**, for a Gaussian field with known power spectrum. That means it is the maximum a posteriori solution of the posterior

$$-\log P(s|d) = \frac{1}{2}(s-d)^T N^{-1}(s-d) + \frac{1}{2}s^T S^{-1} s + \text{const.} \tag{14.24}$$

The Wiener filter also minimizes the mean squared error (MSE) between the true signal and the reconstructed signal. For more discussion of this see 1905.05846.

Based on the Wiener filtererd data, one can then construct the **Quadratic Maximum Likelihood (QML)** which is the provably optimal power spectrum estimator. We refer to appendix B of 1909.09375 for a discussion of this estimator. It involves Wiener filtering the data and then estimating the mode coupling matrix from simulations.

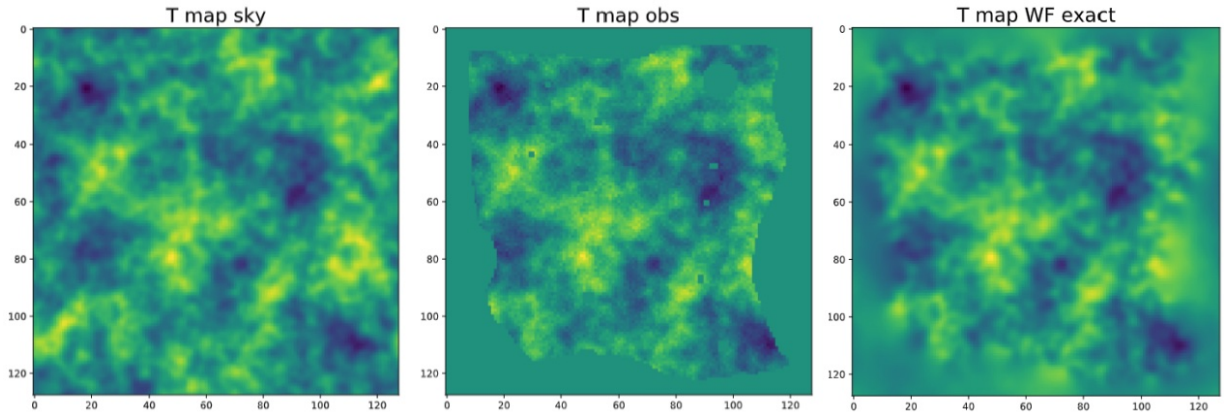


Figure 16. Wiener filtering example (plot from 1905.05846). Left: true signal, Middle: Observed noisy and masked data, Right: Wiener filtered data, which is a reconstruction of the underlying true map.

14.6 Likelihood of the CMB

Once we have a power spectrum estimator (from pseudo-CL or QML or a different estimator), as well as its covariance, we need a likelihood that we can MCMC sample from, to obtain cosmological parameters. As in our analysis of a dark matter simulation in Sec. 11, a good approximation is a Gaussian with a fixed covariance matrix. This choice was made for example by Planck in their "high-ℓ likelihood" used for $\ell > 30$. Planck also had a completely different "low-ℓ likelihood" for $\ell < 30$ (1907.12875). The reason for that is that the power spectrum estimator is based on the square of Gaussian variables (the a_{lm}), which is not Gaussian distributed. For large enough ℓ we can average over enough modes to make the distribution of the $\hat{C}_\ell$ Gaussian by the Central Limit Theorem but for small ℓ that does not happen. On the other hand, for small ℓ there are less modes involved so we can allow for a computationally more expensive approach. In particular, it is possible to make a likelihood at map/pixel level, rather than at power spectrum level. For details of possible small-scale and large-scale likelihoods, we refer to the review 1909.09375.

14.7 Tools to sample the CMB likelihood

Rather than discussing mathematical details of likelihoods I want to introduce a state-of-the-art tool for working with CMB likelihoods (and other cosmological probes): **Cobaya (code for Bayesian analysis in Cosmology)** https://cobaya.readthedocs.io/. Cobaya for example will be used to analyze data from Simons Observatory. It builds a common framework that includes:

- Theory codes (CLASS and CAMB)
- Built-in likelihoods of cosmological experiments (Planck, Bicep-Keck, SDSS etc). Collaborations can release their likelihood as Cobaya modules.
- Various MCMC samplers.
- Tools to analyze the MCMC samples and make common plots.

A similar project, more widely used in the large-scale structure community is **CosmoSIS (COSMOlogical Survey Inference System)** https://cosmosis.readthedocs.io/.

If you want to combine say the the Planck and DES likelihood to sample cosmological parameters, perhaps with some extension of LambdaCDM, a practical approach is to set up this analysis in Cobaya. You should not try to analyze e.g. the Planck data directly from map level for a power spectrum analysis, since you'd have to redo all the hard work of the Planck collaboration to make a reliable likelihood with correct covariance. Collaborations also release their likelihood directly, without going through Cobaya. Sometimes these can be directly imported as Python modules. See for example the ACT CMB likelihood here: https://github.com/ACTCollaboration/pyactlike. We will look at an example script that uses Cobaya.

15 Polarization and primordial B-modes

We now give a brief overview of CMB polarization, which is usually expressed in terms of E-modes and B-modes, so that a complete analysis of CMB perturbations includes T, E, B where T is the temperature. CMB polarization from E-modes roughly doubles the information on primordial physics compared to T alone. B-mode polarization, if of primordial origin, would detect primordial gravitational waves. Details about CMB polarization, B-modes and primordial gravitational waves can be found in the review 1510.06042 and additional visual explanations of polarization are presented in astro-ph/9706147. For details on the Stoke parameters see here: astro-ph/0409734v2. The books by Dodelson and Baumann also have detailed sections on polarization with useful illustrations.

Polarization is generated by the scattering between photons and free electrons. A quadrupolar anisotropy (in the rest frame of the electron) of incoming unpolarized light, which Thompson scatters on the electron, leads to outgoing polarized light as shown in Fig. 17. Towards the end of inflation, when photons decouple from the electrons and protons, density perturbations in the primordial plasma lead to such a quadrupolar anisotropy. Therefore, there should be some correlation between temperature and polarization anisotropies.

The mathematical characterization of CMB polarization anisotropies is complicated by the fact that polarization is not a scalar field. To define the polarization we need the Stokes parameters. Recall that a monochromatic plane electromagnetic wave can be represented as

$$\mathbf{E}(t) = E_x \cos(\omega t)\hat{\mathbf{x}} + E_y \cos(\omega t - \varphi)\hat{\mathbf{y}}, \tag{15.1}$$

where we put the phase by convention in the second term. Depending on φ the wave can be linearly, elliptically or circular polarized. For any electronagnetic wave (not just a monochromatic plane wave) the **Stokes parameters** are defined by the expectation values (time averages) of the transverse components as

$$\begin{aligned} I &= |E_x|^2 + |E_y|^2, \\ Q &= |E_x|^2 - |E_y|^2, \\ U &= 2|E_x||E_y| \cos\varphi, \\ V &= 2|E_x||E_y| \sin\varphi \end{aligned}$$

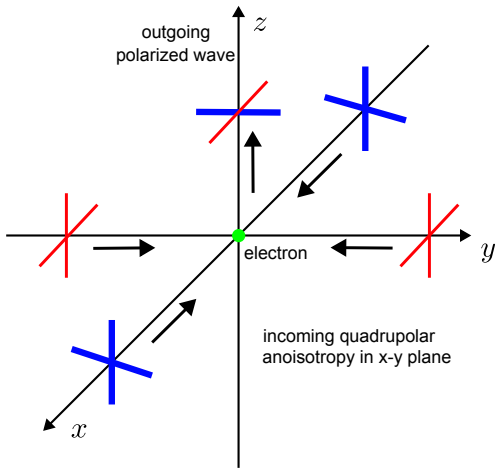


Figure 17. Generation of CMB polarization by scattering of a quadrupole anisotropy. Bold blue lines are hotter, thin red lines are colder. Figure adapted from Dodelson-Schmidt.

I is the intensity of the light, which is proportional to the temperature. The CMB is a sum of unpolarized light for which ($U = U = V = 0$) and linearly polarized light ($\varphi = 0$ and $Q \neq 0, U \neq 0, V = 0$). Therefore a CMB experiment measures an intensity map I and a Q and U map. Only exotic theories of the early universe can produce circular polarization ($V \neq 0$). The polarization fraction of the CMB is about 10%.

While T is a scalar and does not change under rotation, the quantities Q and U transform under rotation by an angle ψ as a spin-2 field $(Q \pm iU)(\hat{n}) \to e^{\mp 2i\psi}(Q \pm iU)(\hat{n})$. This is because they are "headless vectors" so that a 180° rotation brings them back to themselves. The harmonic analysis of $Q \pm iU$ therefore requires expansion on the sphere in terms of tensor (spin-2) spherical harmonics

$$(Q \pm iU)(\hat{n}) = \sum_{\ell,m} a_{\pm 2,\ell m}\, {}_{\pm 2}Y_{\ell m}(\hat{n})\,. \tag{15.2}$$

While Q and U maps come naturally out of experiments, for theoretical analysis it is more convenient to work with scalar quantities. These can be obtained as follows:

$$a_{E,\ell m} \equiv -\frac{1}{2}\left(a_{2,\ell m} + a_{-2,\ell m}\right) \tag{15.3}$$

$$a_{B,\ell m} \equiv -\frac{1}{2i}\left(a_{2,\ell m} - a_{-2,\ell m}\right) \tag{15.4}$$

which are the multipole coefficients of the scalar **E-mode** and **B-mode** fields:

$$E(\hat{n}) = \sum_{\ell,m} a_{E,\ell m}\, Y_{\ell m}(\hat{n}) \tag{15.5}$$

$$B(\hat{n}) = \sum_{\ell,m} a_{B,\ell m}\, Y_{\ell m}(\hat{n})\,. \tag{15.6}$$

Pure E-mode fields are curl-free and pure B-mode fields are divergence free, in close analogy with electrodynamics.

The angular power spectra are defined as before

$$C_\ell^{XY} \equiv \frac{1}{2\ell+1}\sum_m \langle a^*_{X,\ell m} a_{Y,\ell m}\rangle\,, \qquad X,Y = T,E,B\,. \tag{15.7}$$

The auto power spectra are TT, EE and BB. Some of the cross power spectra are zero. Although E and B are both invariant under rotations, they behave differently under parity transformations. E-modes are parity even (like temperature) an B-modes are parity odd. For this reason, in a parity invariant early universe, the cross power spectrum TE is non-zero while TB or EB are zero. Note however that secondary (non-primordial) anisotropies and foregrounds can generate non-zero TB and EB-correlations.

A crucial physical insight found in the late nineties (astro-ph/9609169) is that **scalar (density) perturbations create only *E*-modes and no *B*-modes**. On the other hand, tensor (gravitational wave) perturbations create both E-modes and B-modes. For this reason, current and upcoming experiments try to detect primordial B-modes to detect gravitational waves. Note however that foregrounds and gravitational lensing do generate B-modes, and these have to be cleaned out in order to not confuse them with a primordial signal.

Once we have calculated the E-mode and B-mode power spectra, which are scalars, cosmological analysis works in much the same way as for temperature T. For example, CAMB and CLASS can calculate polarization transfer functions $\Delta_{E\ell}(k)$ and $\Delta_{B\ell}(k)$ so that the power spectrum of EE is

$$C_\ell^{EE} = \frac{2}{\pi}\int k^2 dk P_{\mathcal{R}}(k)\Delta^2_{E\ell}(k) \tag{15.8}$$

and similar for TE and BB.

16 Primordial non-Gaussianity

The cosmic microwave background is an the ideal probe of primordial non-Gaussianity, i.e. of interactions (and thus correlations) between the primordial modes. This is because of the linearity of the CMB. In the future, it may be possible to beat the CMB constraints with large-scale structure, but this is probably at least a decade away (with the exception of so called "local non-Gaussianity"). Good reviews on primordial non-Gaussianity are 1001.4707 (which this section is based on) and 1003.6097. The formalism we are discussing here also generalizes to other **bispectra (i.e. 3 point correlators)**, including non-primordial ones and bispectra of galaxy surveys.

16.1 Primordial bispectra

Recall that for the primordial potential we found (from statistical homogeneity and isotropy):

$$\langle \Phi(\mathbf{k})\Phi^*(\mathbf{k}')\rangle = (2\pi)^3 \mathbf{d}_D(\mathbf{k}-\mathbf{k}')P_\Phi(k)\,, \tag{16.1}$$

The equivalent statement for the 3-point correlator is

$$\langle \Phi(\mathbf{k}_1)\Phi(\mathbf{k}_2)\Phi(\mathbf{k}_3)\rangle = (2\pi)^3\,\mathbf{d}_D(\mathbf{k}_{123})\,B_\Phi(k_1,k_2,k_3)\,. \tag{16.2}$$

Here, the delta function enforces the triangle condition, that is, the constraint that the wavevectors in Fourier space must close to form a triangle, $\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 = 0$.

A well studied model is the the **local model** in which contributions from 'squeezed' triangles are dominant, that is, with e.g. $k_3 \ll k_1,\, k_2$. In this model, non-Gaussianity is created as follows:

$$\Phi(\mathbf{x}) = \Phi_{\rm L}(\mathbf{x}) + \Phi_{\rm NL}(\mathbf{x}) \tag{16.3}$$

$$= \Phi_{\rm L}(\mathbf{x}) + f_{\rm NL}\big[\Phi^2_{\rm L}(\mathbf{x}) - \langle\Phi^2_{\rm L}(\mathbf{x})\rangle\big] \tag{16.4}$$

where $f_{\rm NL}$ is called the nonlinearity parameter. The bound on $f_{\rm NL}$ from Planck is about $f_{\rm NL} < 5$. For this model one can show that

$$B_\Phi(k_1, k_2, k_3) = \; 2f_{\rm NL}\left[P_\Phi(k_1)P_\Phi(k_2) + P_\Phi(k_2)P_\Phi(k_3) + P_\Phi(k_3)P_\Phi(k_1)\right] \tag{16.5}$$

The bispectrum is often written in terms of the dimensionless **shape function**

$$S(k_1, k_2, k_3) \equiv (k_1k_2k_3)^2 B_\Phi(k_1, k_2, k_3)\,, \tag{16.6}$$

A different primordial bispectrum that is often considered is the **equilateral model** with shape function

$$S^{\rm equil}(k_1, k_2, k_3) = \frac{(k_1 + k_2 - k_3)(k_2 + k_3 - k_1)(k_3 + k_1 - k_2)}{k_1k_2k_3}\,. \tag{16.7}$$

Unlike the local model, this one peaks for equilateral triangles, so the local and equilateral models probe different kinds of correlations.

16.2 CMB bispectrum

The CMB bispectrum is the three point correlator of the $a_{\ell m}$. Using Eq. 13.4 we obtain

$$B^{\ell_1\ell_2\ell_3}_{m_1m_2m_3} = \langle a_{\ell_1m_1}a_{\ell_2m_2}a_{\ell_3m_3}\rangle \tag{16.8}$$

$$= (4\pi)^3(-i)^{l_1+l_2+l_3}\int \frac{d^3k_1}{(2\pi)^3}\frac{d^3k_2}{(2\pi)^3}\frac{d^3k_3}{(2\pi)^3}\Delta_{l_1}(k_1)\Delta_{l_2}(k_2)\Delta_{l_3}(k_3)\times \tag{16.9}$$

$$\langle\Phi(\mathbf{k}_1)\Phi(\mathbf{k}_2)\Phi(\mathbf{k}_3)\rangle\, Y_{\ell_1m_1}(\hat{\mathbf{k}}_1)\, Y_{\ell_2m_2}(\hat{\mathbf{k}}_2)\, Y_{\ell_3m_3}(\hat{\mathbf{k}}_3) \tag{16.10}$$

$$= \left(\frac{2}{\pi}\right)^3\int x^2dx\int dk_1dk_2dk_3(k_1k_2k_3)^2 B_\Phi(k_1, k_2, k_3)\,\Delta_{\ell_1}(k_1)\Delta_{\ell_2}(k_2)\Delta_{\ell_3}(k_3) \tag{16.11}$$

$$\times\, j_{\ell_1}(k_1x)j_{\ell_2}(k_2x)j_{\ell_3}(k_3x)\int d\Omega_{\hat{\mathbf{x}}}\, Y_{\ell_1m_1}(\hat{\mathbf{x}})Y_{\ell_2m_2}(\hat{\mathbf{x}})Y_{\ell_3m_3}(\hat{\mathbf{x}}) \tag{16.12}$$

where we have inserted the exponential integral form for the delta function in the bispectrum definition. The last integral over the angular part of $\mathbf{x}$ is the Gaunt integral, while x is the radial conformal distance. The full bispectrum $B^{\ell_1\ell_2\ell_3}_{m_1m_2m_3}$ can be expressed in terms of the **reduced bispectrum** $b_{\ell_1\ell_2\ell_3}$ as

$$B^{\ell_1\ell_2\ell_3}_{m_1m_2m_3} = \mathcal{G}^{\ell_1\ell_2\ell_3}_{m_1m_2m_3} b_{\ell_1\ell_2\ell_3}\,. \tag{16.13}$$

The reduced bispectrum is given by

$$b_{\ell_1\ell_2\ell_3} = \left(\frac{2}{\pi}\right)^3 \int x^2 dx \int dk_1 dk_2 dk_3 \; (k_1 k_2 k_3)^2 \; B_\Phi(k_1, k_2, k_3) \tag{16.14}$$

$$\times \, \Delta_{\ell_1}(k_1) \, \Delta_{\ell_2}(k_2) \, \Delta_{\ell_3}(k_3) \, j_{\ell_1}(k_1 x) \, j_{\ell_2}(k_2 x) \, j_{\ell_3}(k_3 x) \, . \tag{16.15}$$

which relates the primordial bispectrum, predicted by inflationary theories, to the reduced bispectrum observed in the cosmic microwave sky. This formula is the equivalent of the power spectrum relation

$$C_\ell = \frac{2}{\pi} \int dk k^2 P_\Phi(k) \Delta_\ell^2(k). \tag{16.16}$$

16.3 Optimal estimator for bispectra

For a fullsky observation it can be shown that the optimal estimator for $f_{\rm NL}$ is

$$\hat{f}_{\rm NL} = \frac{1}{\mathcal{N}} \sum_{\{\ell_i, m_i\}} \frac{\mathcal{G}^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3} b^{f_{\rm NL}=1}_{\ell_1 \ell_2 \ell_3}}{C_{\ell_1} C_{\ell_2} C_{\ell_3}} a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \tag{16.17}$$

$$\mathcal{N} = \sum_{\{\ell_i, m_i\}} \frac{\left(\mathcal{G}^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3} b^{f_{\rm NL}=1}_{\ell_1 \ell_2 \ell_3}\right)^2}{C_{\ell_1} C_{\ell_2} C_{\ell_3}} \, , \tag{16.18}$$

where $b_{\ell_1\ell_2\ell_3}$ is the reduced bispectrum and $\mathcal{G}^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3}$ is the Gaunt integral and $\mathcal{N}$ is the normalization factor. This estimator can be interpreted as summing up all mode triplets weighted by their expected signal-to-noise. See 1001.4707 and 1003.6097 for two different derivations of this result. This kind of estimator is called a **cubic estimator**, because it uses three copies of the a_{lm} (the power spectrum estimator on the other hand is a quadratic estimator).

The noise and beam of the experiment can be included with the following replacements:

$$C_\ell \to C_\ell \mathcal{B}_\ell^2 + N_\ell \, , \qquad b_{\ell_1\ell_2\ell_3} \to b_{\ell_1\ell_2\ell_3} \mathcal{B}_{\ell_1} \mathcal{B}_{\ell_2} \mathcal{B}_{\ell_3} \, ; \tag{16.19}$$

$\mathcal{B}$ is the beam and N_ℓ is the noise power spectrum (constant for uncorrelated white noise). The noise is assumed to be Gaussian (which is a very good approximation because the bispectrum, if non-zero, is much smaller than the power spectrum). Including the effect of the mask is a little harder and we won't review it here. It involves adding a **linear term** to the cubic estimator above. Details can be found in the same reviews.

The way how primordial bispectrum analysis is performed is that theorists have come up with a large collection of theoretically motivated bispectrum templates $b_{\ell_1\ell_2\ell_3}$ (such as local and equilateral), and we have run the bispectrum estimator on all of these templates (Planck 1905.05697). While no statistically significant detection has been made, many models (or at least part of their parameter space) have been excluded in this way. Instead of running bispectrum estimators, one can also measure the bispectrum in bins (as we do in the power spectrum), but all measurements are consistent with zero.

16.4 The separability trick

I want to mention one more important aspect of bispectrum estimation, which also often occurs in cosmology. As it is written above, the bispectrum estimator is computationally intractable, as it is a sum over six variables l, m all of which go to about 2500 for Planck resolution. Fortunately the estimator can be rewritten in a much better form. If the primordial **shape function is separable**, i.e. it can be written in the form

$$S(k_1, k_2, k_3) = X(k_1)\, Y(k_2)\, Z(k_3) + 5 \text{ perms.}\,, \tag{16.20}$$

then the reduced bispectrum can be written as

$$b_{\ell_1\ell_2\ell_3} = \int dx x^2 X_{\ell_1}(x)\, Y_{\ell_2}(x)\, Z_{\ell_3}(x) + \text{ 5 perms}\,, \tag{16.21}$$

where we have defined the quantities:

$$\begin{aligned} X_\ell(x) &\equiv \int dk k^2\, X(k)\, j_\ell(kx)\, \Delta_\ell(k)\,, \\ Y_\ell(x) &\equiv \int dk k^2\, Y(k)\, j_\ell(kx)\, \Delta_\ell(k)\,, \\ Z_\ell(x) &\equiv \int dk k^2\, Z(k)\, j_\ell(kx)\, \Delta_\ell(k)\,. \end{aligned} \tag{16.22}$$

In that case, using the definition of the Gaunt integral, the estimator can be rewritten as

$$\mathcal{E}(a) = \frac{1}{\mathcal{N}} \int dx\, x^2 \int d\Omega_{\hat{\mathbf{n}}} M_X(r, \hat{\mathbf{n}}) M_Y(x, \hat{\mathbf{n}}) M_Z(x, \hat{\mathbf{n}}) + \text{perms.}\,, \tag{16.23}$$

where

$$\begin{aligned} M_X(x, \hat{\mathbf{n}}) &\equiv \sum_{\ell m} \frac{a_{\ell m} X_\ell(x)}{C_\ell} Y_{\ell m}(\hat{n})\,, \\ M_Y(x, \hat{\mathbf{n}}) &\equiv \sum_{\ell m} \frac{a_{\ell m} Y_\ell(x)}{C_\ell} Y_{\ell m}(\hat{n})\,, \\ M_Z(x, \hat{\mathbf{n}}) &\equiv \sum_{\ell m} \frac{a_{\ell m} Z_\ell(x)}{C_\ell} Y_{\ell m}(\hat{n})\,, \end{aligned} \tag{16.24}$$

By a detailed examination of the operations, one finds that this reduces the computational cost from $\mathcal{O}(\ell_{max}^5)$ to $\mathcal{O}(\ell_{max}^3)$ operations, which is can be easily calculated in practice. This re-writing of the estimator is sometimes called a **fast position space estimator** (since we work with the maps M in position space rather than Fourier space). Not all theoretical shapes are separable. However it is often possible to expand unseparable shapes into separable shapes (see 0912.5516).

17 Secondary anisotropies: CMB lensing

Lensing is the leading secondary effect on the CMB anisotropies below $\ell \simeq 4000$ (Fig. 18). It

- smooths acoustic peaks
- transfers power to small scales
- introduces non-Gaussianity
- makes B-mode polarization by lensing E-modes. Thus de-lensing is important for B-mode searches.

The lensing effect can be used to reconstruct the lensing potential, a map of the integrated mass density of the universe on large scales. The lensing potential can be used as a probe of cosmological parameters including neutrino masses and dark energy. An important feature of lensing is that it **probes the entire mass density** (since any mass and energy gravitate in General Relativity), while e.g. a galaxy survey only probes luminous matter. This is why lensing is critical to probe dark matter.

My brief discussion of CMB lensing is based on the review astro-ph/0601594v4. Another good review is 0911.0612. We will only discuss temperature, but polarization is lensed in the same way.

17.1 CMB lensing potential

Weak lensing remaps the unlensed CMB map as follows. The lensed CMB temperature in a direction $\hat{\mathbf{n}}$, $\tilde{T}(\hat{\mathbf{n}})$, is given by the unlensed temperature in the deflected direction

$$\tilde{T}(\hat{\mathbf{n}}) = T(\hat{\mathbf{n}}') = T(\hat{\mathbf{n}} + \boldsymbol{\alpha}) \tag{17.1}$$

where $\boldsymbol{\alpha}$ is a deflection angle. This result follows from General Relativity. At lowest order the deflection angle is a pure gradient, $\boldsymbol{\alpha} = \nabla\psi$. The **lensing potential** is defined by

$$\psi(\hat{\mathbf{n}}) \equiv -2\int_0^{\chi_*} \mathrm{d}\chi\, \frac{(\chi_* - \chi)}{\chi_*\chi}\Psi(\chi\hat{\mathbf{n}}; \eta_0 - \chi), \tag{17.2}$$

where χ is conformal radial distance along the line of sight, χ_* is the conformal distance to recombination, and η_0 is the conformal time today, and Ψ is the Newtonian gravitational potential. From this one can calculate the power spectrum of the lensing potential C_l^Ψ. It is defined by

$$\langle \psi_{lm}\psi^*_{l'm'}\rangle = \delta_{ll'}\delta_{mm'}C_l^\psi. \tag{17.3}$$

where

$$C_l^\psi = 16\pi \int \frac{\mathrm{d}k}{k}\int_0^{\chi_*}\mathrm{d}\chi\int_0^{\chi_*}\mathrm{d}\chi'\,\mathcal{P}_\Psi(k;\eta_0-\chi,\eta_0-\chi')j_l(k\chi)j_l(k\chi')\left(\frac{\chi_*-\chi}{\chi_*\chi}\right)\left(\frac{\chi_*-\chi'}{\chi_*\chi'}\right). \tag{17.4}$$

which in the linear regime can be calculated by a Botzmann solver like CAMB, depending on cosmological parameters. $\mathcal{P}_\Psi(k;\eta_0-\chi,\eta_0-\chi')$ is the power spectrum between unequal times.

It is also often useful to work with the CMB **convergence** given by

$$\kappa(\mathbf{n}) = -\frac{1}{2}\nabla^2\Psi(\mathbf{n}) \tag{17.5}$$

and thus

$$\kappa(\mathbf{l}) = -\frac{l^2}{2}\Psi(\mathbf{l}) \tag{17.6}$$

The convergence probes the integrated matter density between us and the CMB (since from the Poisson equation $\nabla^2\Psi(\mathbf{n}) \propto \rho$ with density ρ). A visual example of the quantities involved in CMB lensing is uploaded to the lecture files.

17.2 Lensed CMB map

We will now use flatsky coordinates, which are often used for CMB lensing, and simplify the expressions. Of course, everything can also be expressed in spherical harmonics. Our flat-sky 2D Fourier transform convention for the temperature field here is:

$$\Theta(\mathbf{x}) = \int \frac{\mathrm{d}^2\mathbf{l}}{2\pi}\,\Theta(\mathbf{l})e^{i\mathbf{l}\cdot\mathbf{x}}, \qquad \Theta(\mathbf{l}) = \int \frac{\mathrm{d}^2\mathbf{x}}{2\pi}\,\Theta(\mathbf{x})e^{-i\mathbf{l}\cdot\mathbf{x}}. \tag{17.7}$$

The power spectrum for our statistically isotropic temperature field is diagonal in $\mathbf{l}$, and is given by

$$\langle\Theta(\mathbf{l})\Theta^*(\mathbf{l}')\rangle = C_l^\Theta\delta(\mathbf{l}-\mathbf{l}'). \tag{17.8}$$

For weak lensing, to good approximation, the lensing effect can be described by Taylor expansion. To first order we have

$$\begin{aligned}\tilde{\Theta}(\mathbf{x}) &= \Theta(\mathbf{x}') = \Theta(\mathbf{x}+\boldsymbol{\nabla}\psi)\\ &\approx \Theta(\mathbf{x}) + \nabla^a\psi(\mathbf{x})\nabla_a\Theta(\mathbf{x})\end{aligned} \tag{17.9}$$

Going to Fourier space, one can show that to first order the lensed CMB field is given by

$$\tilde{\Theta}(\mathbf{l}) \approx \Theta(\mathbf{l}) - \int \frac{\mathrm{d}^2\mathbf{l}'}{2\pi}\,\mathbf{l}'\cdot(\mathbf{l}-\mathbf{l}')\psi(\mathbf{l}-\mathbf{l}')\Theta(\mathbf{l}') \tag{17.10}$$

This shows that there will now be some **mode coupling** between modes $\Theta(\mathbf{l})$ and $\Theta(\mathbf{l}')$, assuming a fixed lensing potential Ψ. That means that there will be off-diagonal components in the covariance matrix of the observed CMB. We could now also derive the powers spectrum of the lensed CMB field $\tilde{C}_l^\Theta$ by calculating $\left\langle\tilde{\Theta}(\mathbf{l})\tilde{\Theta}^*(\mathbf{l})\right\rangle$.

17.3 Quadratic estimator for lensing

We now outline how the CMB lensing potential can be measured. The standard approach to do this is the **quadratic estimator formalism**. The quadratic estimator can be used in may situation in cosmology, where a large scale "background" field (here the lensing potential) modulates the statistics of small scale observables (here the CMB temperature perturbations). As we have seen, for a fixed lensing potential, the distribution of the observed temperature will not be isotropic. This suggests that we may be able to use the quadratic off-diagonal terms of the ψ-fixed correlation $\langle\tilde{\Theta}(\mathbf{l})\tilde{\Theta}(\mathbf{l}')\rangle_\Theta$ to constrain the lensing potential in our sky realization. The subscript in the expectation value means that Θ here is a random variable while Ψ is a fixed realization.

Averaging over realizations of the unlensed temperature field Θ to first oder in the lensing potential gives

$$\langle\tilde{\Theta}(\mathbf{l})\tilde{\Theta}^*(\mathbf{l}-\mathbf{L})\rangle_\Theta = \delta(\mathbf{L})\,C_l^\Theta - \int \frac{\mathrm{d}^2\mathbf{l}'}{2\pi}\left[\mathbf{l}'\cdot(\mathbf{l}-\mathbf{l}')\psi(\mathbf{l}-\mathbf{l}')\langle\Theta(\mathbf{l}')\Theta^*(\mathbf{l}-\mathbf{L})\rangle\right.$$
$$\left. + \mathbf{l}'\cdot(\mathbf{l}-\mathbf{L}-\mathbf{l}')\psi^*(\mathbf{l}-\mathbf{L}-\mathbf{l}')\langle\Theta(\mathbf{l})\Theta^*(\mathbf{l}')\rangle\right] \quad (17.11)$$
$$= \delta(\mathbf{L})\,C_l^\Theta + \frac{1}{2\pi}\left[(\mathbf{L}-\mathbf{l})\cdot\mathbf{L}\,C^\Theta_{|\mathbf{l}-\mathbf{L}|} + \mathbf{l}\cdot\mathbf{L}\,C_l^\Theta\right]\psi(\mathbf{L}) \quad (17.12)$$

To estimate the lensing potential we thus want to sum over all quadratic combinations $\tilde{\Theta}(\mathbf{l})\tilde{\Theta}^*(\mathbf{l}-\mathbf{L})$ with some weighting factor g that needs to be determined:

$$\hat{\psi}(\mathbf{L}) \equiv N(\mathbf{L})\int \frac{\mathrm{d}^2\mathbf{l}}{2\pi}\,\tilde{\Theta}(\mathbf{l})\tilde{\Theta}^*(\mathbf{l}-\mathbf{L})g(\mathbf{l},\mathbf{L}), \quad (17.13)$$

where $g(\mathbf{l},\mathbf{L})$ is the weighting function and $N(\mathbf{L})$ is a normalization. This strategy is originally from astro-ph/0301031 and has been re-used for many different applications.

To find the weighting function and normalization, we impose two conditions:

- The estimator should be unbiased, i.e $\langle\hat{\psi}(\mathbf{L})\rangle_\Theta = \psi(\mathbf{L})$
- The variance (error) of our estimator should be as small as possible.

It can be shown that this gives the weights

$$g(\mathbf{l},\mathbf{L}) = \frac{(\mathbf{L}-\mathbf{l})\cdot\mathbf{L}\,C^\Theta_{|\mathbf{l}-\mathbf{L}|} + \mathbf{l}\cdot\mathbf{L}\,C_l^\Theta}{2\tilde{C}_l^{\mathrm{tot}}\tilde{C}^{\mathrm{tot}}_{|\mathbf{l}-\mathbf{L}|}}. \quad (17.14)$$

and the normalization is

$$N(\mathbf{L})^{-1} = \int \frac{\mathrm{d}^2\mathbf{l}}{(2\pi)^2}\left[(\mathbf{L}-\mathbf{l})\cdot\mathbf{L}\,C^\Theta_{|\mathbf{l}-\mathbf{L}|} + \mathbf{l}\cdot\mathbf{L}\,C_l^\Theta\right]g(\mathbf{l},\mathbf{L}). \quad (17.15)$$

As was the case for the bispectrum, this estimator can be re-written as a fast position space estimator.

This estimator (with minor modifications to take into account the mask and noise), is for example used in the recent ACT CMB lensing analysis (2004.01139), which is also in flatsky coordinates. The sperhical hamonics version of this estimator was used in the Planck analysis (1807.06210). However, because lensing is a non-linear operation, the quadratic estimator is not optimal in general. For existing experiments (Planck, ACT), it is optimal, but for Simons Observatory it will already be slightly sub-optimal and for future very high resolution experiments it can be very suboptimal. A completely optimal lensing analysis can be made with a field-level likelihood, but is computationally extremely expensive. References on this topic include 1708.06753,1704.08230,astro-ph/0209489.

17.4 Physics with CMB lensing

Once the lensing potential is reconstructed, one can estimate its power spectrum in the usual way and use the lensing power spectrum in a cosmological analysis to constrain parameters. Lensing is in particular a great probe of the size of matter perturbations at later times. By comparing the amplitude of primordial (primary) CMB perturbations with the amplitude of late time perturbations from lensing, one can study the **growth of structure** in the universe. This for example can be used to contstrain neutrino masses, as the free streaming of neutrinos suppresses growth. The measured growth of structure at late times is currently an exciting topic in cosmology, with evidence for a disagreement with Lambda-CDM (see eg. 2203.06142,2304.05203) called the S_8 **tension** or σ_8 **tension**.

Another important application of the lensing potential is to cross-correlate it with a different tracer of matter, such as a galaxy survey. Such cross-power spectra can also be very sensitive to various cosmological parameters, for example local primordial non-Gaussianity (e.g. 1710.09465).

18 Secondary anisotropies: Sunyaev-Zeldovich effect

Apart from being gravitationally lensed, the second thing that happens to photons on the way from the CMB to us is being re-scattered by charges (mostly free electrons) in **inverse Compton scattering**. This re-scattering is called the **Sunyaev-Zeldovich effect**. This effect comes in several variants:

- The **thermal Sunyaev-Zeldovich (tSZ) effect** is the scattering of photons on **hot electrons**, i.e. on their thermal velocities.
- The **kinetic Sunyaev-Zeldovich (kSZ) effect** is the scattering of photons on eletrons due to the electron's bulk movement (i.e. all the electrons in a galaxy move on average with the velocity of the galaxy).

These effects are by far the most important SZ effects. There are however smaller effects including the **polarized SZ effect** and the **rotational SZ effect**. The total probability of a CMB photon to be re-scattered between recombination and Earth is about 5%. This probability is related to the **optical depth** and the **visibility function**. To my knowledge there is no comprehensive review on SZ anisotropies.

In passing I want to mention that apart from lensing and electron scattering there is a class of secondary anisotropies which come from the evolution of gravitational potentials over time. These cause the **(late time) ISW effect, the Rees-Sciama effect (also called non-linear ISW effect) and the moving lens effect**. Finally of course all these secondary effects are combined, for example SZ anisotropies are lensed, and there can be multiple scatterings etc. These higher order effects are not yet detectable.

18.1 Thermal SZ effect

TSZ is generated by any hot ionized gas. In CMB maps, the tSZ is visible in particular from galaxy clusters, which are the largest massive structures in the Universe, formed by gravitational

collapse. Their comoving size is a few Mpc and their angular sizes range from about one arcminute to about one degree (depending on size and distance). Clusters can be detected in various ways, e.g. by galaxy surveys in the optical, by tSZ emission, or by X-ray astronomy (Bremsstrahlung emission of the electrons on the nuclei). The temperature, measured from X-ray, is typically a few keV.

The thermal SZ effect generated by a gas of electrons at temperature T_e leads to a **spectral distortion of the CMB emission law**. The difference between the distorted CMB photon distribution I_ν and the original CMB blackbody spectrum $B_\nu(T_{\rm CMB})$

$$\Delta I_\nu = I_\nu - B_\nu(T_{\rm CMB}) \tag{18.1}$$

can be calculated to be:

$$\Delta I_\nu = y \frac{x e^x}{(e^x - 1)} \left[\frac{x(e^x + 1)}{(e^x - 1)} - 4 \right] B_\nu(T_{\rm CMB}) \tag{18.2}$$

where and $x = h\nu / kT_{\rm CMB}$. The dimensionless parameter y, called **Compton-y parameter**, is proportional to the integral of the electron pressure along the line of sight:

$$y = \int_{\rm los} \frac{kT_e}{m_e c^2} n_e \sigma_{\rm thomson} dl$$

where T_e is the electron temperature, m_e the electron mass, c the speed of light, n_e the electron density, and $\sigma_{\rm thomson}$ the Thomson cross section. A multi-frequency CMB detector can measure the **Compton-y map**, which is caused by the tSZ effect.

18.2 Matched filter and tSZ stacking

Roughly 80% of the baryons in a cluster are not contained within galaxies, but rather exist as a cloud of gas bound within the gravitational potential well created by a dark matter halo that caries the vast majority of the mass of the cluster. Within this well, the dilute gas becomes ionized and heated to temperatures of millions of Kelvin. Detailed calculations show that the tSZ effect leads to decrement of power at frequencies below the 220 GHz and extra power at higher frequencies. This result is redshift independent. A nice illustration of the tSZ sources can be found in the CMB S4 summer school notebooks.

Finding tSZ sources is a typical application of another generally important method in cosmology, the **matched filter method**. The matched filter is the optimal way to detect a localized object (e.g. a theoretical template of the cluster profile) that is (linearly) added to noisy data. It is given by a convolution between the signal profile and the CMB map. In harmonic space, for a spherically symmetric profile, it can be written as

$$\psi(\hat{\mathbf{n}}) = \sum_{\ell m} \frac{\Theta_{\ell m} S_\ell}{C_\ell + N_\ell} Y_{\ell m}(\hat{\mathbf{n}}). \tag{18.3}$$

where S_ℓ is the spherical harmonics transform of the radial profile of the signal $S(r)$ (e.g. the tSZ profile), $\Theta_{\ell m}$ is the CMB map, and $C_\ell + N_\ell$ is the CMB power spectrum plus the instrumental noise power spectrum. The output of the matched filter is a "heat map" of detection probabilities,

which has its maxima where a tSZ source exists. A matched filter usually comes with some parameter to scan over, e.g. the radius of the profile.

Some details on the matched filter method can be found e.g here 2106.03718. An application of the matched filter for a completely different problem (finding primordial particle production), and a discussion of why it is optimal, can be found e.g. here 1910.00596 (Sec. 3B).

For tSZ sources, one often wants to understand signals at the low mass and therefore low signal to noise end, where the matched filter may not be able to pick up the signal. With an external catalogue of galaxy clusters, one can co-add the signals from objects in the external catalogue to boost the signal to noise. This is called **tSZ stacking**. From the stack, one can then infer parameters of cluster physics, such as the radial profile of the gas temperature. **Stacking local sources** with an external catalogue to enhance SNR is also a generally important technique. For example, it was recently used to detect a 21cm intensity signal with CHIME (2202.01242).

18.3 Kinetic SZ effect

The kSZ effect is not temperature dependent, and leads to a blackbody contribution to the CMB in the same way as lensing. At high $\ell \gtrsim 4000$ the kSZ is the dominant contribution to CMB temperature, as shown in Fig. 18. The kSZ can be used both to probe the gas distribution of clusters and galaxies, as well as for cosmology. The kSZ temperature is given by the line-of-sight integral

$$\Theta^{kSZ}(\hat{\mathbf{n}}) \propto \sigma_T \int dr\, n_e(r, \hat{\mathbf{n}}) v_r(r, \hat{\mathbf{n}}) \tag{18.4}$$

where n_e is the electron density and v_r is the radial velocity of the structure that contains the electron (not the velocity caused by temperature). A nice application, which I developed with my collaborators, is to use this signal to **reconstruct the velocity field**, by making a template for n_e using a survey of the galaxy density δ_g. One can then write a quadratic estimator (as in the case of lensing, but here I chose to work in spherical harmonics) for the velocities which is schematically

$$\hat{v}_r(L, M) = N \sum_{\ell, m, \ell', m'} g(L, M, \ell, m, \ell', m') \Theta_{\ell, m} \delta_g(\ell', m', z) \tag{18.5}$$

where again we can find the weights $g(L, M, \ell, m, \ell', m')$ that deliver an unbiased minimum variance estimator. Here z is the redshift of the galaxy bin. The reconstructed velocity map has similar cosmological applications as the lensing potential map. This method will be promising for Simons Observatory. More details can be found in (1707.08129, 1810.13423). The quadratic estimator is not the only way to do cosmology with the kSZ, a review of methods can be found in 1810.13423.

19 Foregrounds and foreground cleaning

Reviews of CMB foregrounds, focused on the physics rather than algorithms, are 1606.03606 (on which my foreground discussion is based on) and "CMB foreground: A concise review (by Kiyotomo Ichiki)".

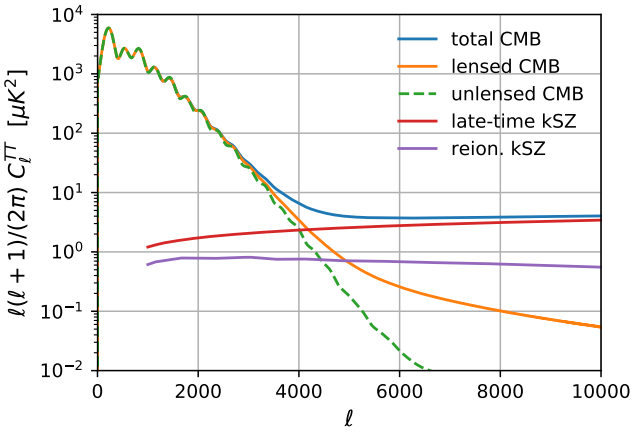


Figure 18. The CMB power spectrum C_l^{TT} from primary CMB, gravitational lensing, late-time kSZ ($z < 6$) and reionization kSZ. We have only shown contributions with blackbody frequency dependence. Non-blackbody contributions (CIB, tSZ) can be mostly removed using multifrequency analysis. Note that the kSZ from both late times and reionization is not known very precisely, the curves come from different theoretical models or simulations. Plot from 1810.13423.

A good review of foreground cleaning and component separation methods is astro-ph/0702198v2. The Planck papers 1303.5072 (appendices), 1502.01588 and 1807.06208 also review these methods and show nice illustration of the raw data and the component separated maps. I can only give you a glimpse of these methods here.

19.1 Galactic foregrounds of the CMB

Before discussing algorithms, I will briefly list the **galactic foregrounds** of the CMB, which are the dominant contribution. There are also extragalactic effects that one can consider a foreground, in particular the SZ effect, the **Cosmic Infrared Background (CIB)** (from unresolved infrared sources) and generally any **point sources** from astrophysical objects such as radio galaxies, infrared galaxies, quasars, which are not re-solved by the instruments used in CMB observations.

The foreground situation is different for temperature and polarization. While in temperature, foregrounds are under good control, in polarization (especially for B-modes at low ℓ) they are still a formidable obstacle, as evidenced by the incorrect claim for a primordial gravitational wave detection by BICEP2.

The most important foregrounds are:

- **Synchrotron radiation** which is emitted by relativistic cosmic ray (CR) electrons, which are accelerated by the Galactic magnetic field.
- **Free-free radiation**, or thermal bremsstrahlung, is emitted by free electrons interacting with ions, in ionised gas.
- **Thermal dust** radiation is blackbody emission from interstellar dust grains with typical temperatures $T \sim 20K$.

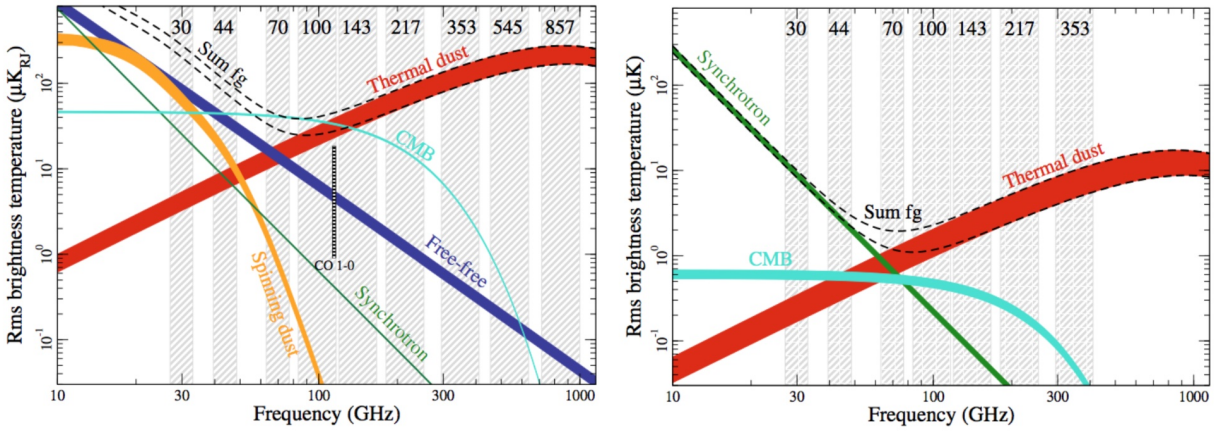


Figure 19. CMB foreground components in temperature (left) and polarization (right). Plot from Planck: 1502.01588, see there for more details.

- **Spinning dust** radiation is emitted by the smallest interstellar dust grains and molecules, which can rotate at GHz frequencies.

All of these have different spectral characteristics which is essential for foreground cleaning. While in temperature, the CMB is of similar amplitude as the foregrounds depending on frequency, in polarization the foregrounds dominate at all frequencies. A plot of the various components compared to the primary CMB is shown in Fig.19.

19.2 The ILC algorithm

The most basic foreground cleaning algorithm for the CMB, which is used in several variants in practice, is the **internal linear components (ILC)** method. Assume that the measured temperature anisotropy T_i in a frequency channel i is a sum

$$T_i = a_i s + f_i + n_i \tag{19.1}$$

where s is the common signal that we want to estimate (such as the CMB), f_i are foregrounds in channel i and n_i is the noise in this channel. The coefficient a_i is the frequency dependence or **spectral energy distribution (SED)** of the signal. This is the **only physical input** required for the ILC. In the case of the CMB this is the known black body spectrum. We also need to assume that the signal is statistically independent from the noise and foreground. Note that the signal does not have to be the CMB, it could also be e.g. the tSZ temperature.

This equation above is basis independent. In the **real space ILC** we work in pixel space so that

$$T_i(\hat{n}) = a_i s(\hat{n}) + f_i(\hat{n}) + n_i(\hat{n}) \tag{19.2}$$

and for the **harmonic space ILC** we use spherical harmonics

$$T^i_{\ell m} = a_i s_{\ell m} + f^i_{\ell m} + n^i_{\ell m} \tag{19.3}$$

The harmonic space version is optimal if the fields are statistically isotropic, however galactic foregrounds are not isotropic. The real space ILC on the other hand can deal with statistical

anisotropy but is not suited for scale-dependent behavior. Both advantages can be combined in the wavelet basis, which is local in position space and harmonic space at the same time, which results in the **Needlet ILC (NILC)**. NILC is one of Planck's four component separation methods.

The ILC is a linear combination of the input maps

$$\hat{s} = \sum_i w_i T_i \tag{19.4}$$

weighted with weights w_i, so that $\hat{s}$ is **unbiased** and **minimum variance**. This can be done with a constrained optimization using a Lagrangian multiplier. The result is

$$\mathbf{w} = \frac{\mathbf{A}^T \mathbf{C}^{-1}}{\mathbf{A}^T \mathbf{C}^{-1} \mathbf{A}} \tag{19.5}$$

where $\mathbf{A}$ is the vector of the SED coefficients a_i. The covariance matrix is estimated from the data, for example in harmonic space we have

$$C_{ij} = \frac{1}{2m+1} \sum_m T^i_{\ell m} T^{j*}_{\ell m}. \tag{19.6}$$

The CMB S4 summer school notebooks have an example of the ILC.

In the above we have assumed almost nothing about the foregrounds. However, **if the SED of one or more foreground signals are known**, we can **deproject them from our final map**. This is called **constrained ILC**. The constrained ILC can sometimes significantly reduce foreground biases, with only a small increase in variance. Such advanced versions of the ILC algorithm are discussed e.g. in 2307.01043 (sec. III).

19.3 Component separation

Above we have discussed how to foreground clean the maps to obtain a single "signal" $\hat{s}$. More generally, one wants to split the total temperature anisotropy into several components. Consider a **linear mixture model** in pixel space

$$\mathbf{T}(\hat{n}) = \mathbf{A}\mathbf{s}(\hat{n}) + \mathbf{n}(\hat{n}) \tag{19.7}$$

Here $\mathbf{T}$ is the vector of observed frequency channels, $\mathbf{A}$ is the **mixing matrix** (describing how each signal, such as the tSZ, projects into each frequency) and $\mathbf{s}$ is the vector of components that we want to determine. In principle we simply want to invert this equation to obtain the components $\mathbf{s}$. Because of the noise and the fact that the matrix is in general non-invertible (not even a square matrix), there is quite a range of possible solutions, reviewed in astro-ph/0702198v2, depending on various prior assumptions that one can make. Again, there are different possible bases to work in, such as real space and harmonic space. Solving for $\mathbf{s}(\hat{n})$ can also be done with an optimizer or using MCMC sampling (e.g. Planck's **Commander pipeline**). One can also include external data for the various signal components, to make useful templates.

Interestingly, it is even possible to determine the components if the mixing matrix is not known, if the components of the linear mixture can be assumed to be statistically independent. This is possible because statistical independence is a strong mathematical property and often

a physically plausible one. This direction is called **blind separation or independent component analysis (ICA)**. ICA ideas are used in Planck's **SMICA** pipeline (Spectral Matching Independent Component Analysis).

Part IV
Large-Scale Structure

We now move on to 3-dimensional probes of the large-scale structure (LSS) of the universe such as galaxy surveys. We have access to such 3-dimensional data at a much later time (redshift $z \lesssim 10$) than the CMB (redshift $z \simeq 1100$). A major complication compared to the CMB is that matter evolves non-linearly at these later times, both due to gravitation and due to "baryonic" physics. Further, most of the matter density δ_m in the universe is not directly observable. A large part is contained in dark matter, and even most baryonic matter is contained in dilute gas rather than luminous stars. To probe most cosmological parameters, ideally we'd like to measure the matter power spectrum $P_m(k)$, but we can only measure the power spectrum of **tracers** of large-scale structures, such as different galaxy populations. We thus need to learn **how these tracers relate to the matter density**, which can be done on large enough scales with the **bias expansion**. Further, we need to take into account that in cosmology we can only measure the **redshift** of galaxies, but not their absolute distance (unless they contain a standard candle). We thus need to study the topic of **red shift space distortions**.

In this unit we primarily learn to analyze galaxy survey data (but other 3-dimensional probes of the universe work almost the same). Galaxy survey data comes in two broad classes: **Spectroscopic galaxy surveys** take a spectrum of each galaxy, to obtain a precise redshift. **Photometric galaxy surveys** take pictures of the sky in several wavelengths, which allows a rough determination of the redshift. The latter is much easier to do experimentally, so the galaxy sample sizes are much larger, but on the other hand the lack of precise distances loses a lot of information. Both surveys types have different strengths. The geometry of space (dark energy) can best be probed with spectroscopic surveys which deliver precise BAO measurements.

Further reading

The general references of Unit 1 all contain a discussion of galaxy surveys. In addition I recommend

- The classic review "Large-Scale Structure of the Universe and Cosmological Perturbation Theory" astro-ph/0112551.
- For the connection between matter and galaxies, the galaxy bias, there is the review "Large-Scale Galaxy Bias" 1611.09787
- Hannu Kurki-Suonio's lecture notes "Galaxy Survey Cosmology " https://www.mv.helsinki.fi/home/hkurkisu/
- Specifically on EFT of LSS there are lecture notes from Senatore, Baldauf, Ivanov, and Philcox.

20 The galaxy power spectrum at linear scales

We first start with a discussion of the linear galaxy power spectrum. We will introduce galaxy bias, shot noise and red shift space distortions, but defer a more detailed discussion to later. This section follows Dodelson-Schmidt chapter 11. Later we will extend our discussion to non-linear scales.

20.1 Linear galaxy bias

On large scales, it turns out that the density perturbations of the galaxy density δ_g is related by a constant **linear galaxy bias** to the matter density δ_m:

$$\delta_g(\mathbf{x}, \tau) = b_1(\tau)\delta_m(\mathbf{x}, \tau) \tag{20.1}$$

Here b_1 means that this is the first order bias. The bias depends on conformal time τ or equivalently redshift z. We will briefly discuss the derivation of this result, as well as higher order biases, later. The bias depends sensitively on the galaxy sample considered and is in general red-shift dependent. A typical galaxy bias for a survey like DESI could be $b_1 \sim 2$, i.e. the overdensities of galaxies are twice as large as those of matter.

20.2 Shot noise

In addition to having a bias, a further difference between the matter field and the galaxy field is that the latter is a **point cloud** rather than a continuous field. An approximate way to think about this is that the galaxy field is a **Poisson sampling** where the mean in each volume element of space is modulated by the underlying biased matter density. This leads on large scales to a galaxy field

$$\delta_g(\mathbf{k}) = b\delta_m(\mathbf{k}) + n(\mathbf{k}) \tag{20.2}$$

where n is white noise (i.e. pixels have uncorrelated noise). In terms of the power spectrum we get

$$P_g(k) = b^2 P_m(k) + N(k) \tag{20.3}$$

where the (shot-) noise is approximately inverse to the comoving galaxy density

$$N(k) = \frac{1}{\bar{n}_g} \tag{20.4}$$

Note in particular that **shot noise** is flat in k. While the $\frac{1}{\bar{n}_g}$ approximation is not very precise, especially at high halo density (where halos may not form independently of eachother), the fact that the noise is flat on large scales holds to good approximation.

20.3 Velocity field on large scales

We also need to know the velocity perturbations on large scales, which are the source of red shift distortions. On linear scales the matter velocity and matter density perturbations are related by

$$\boldsymbol{u}_m(\mathbf{k},\tau) = faH\frac{i\mathbf{k}}{k^2}\delta(\mathbf{k},\tau) \tag{20.5}$$

We will derive this result in Sec. 21.2.2. The factor f is called the **linear growth rate**. The growth rate is close to unity for a ΛCDM universe and exactly 1 for a flat matter-dominated cosmology. Notice that the velocity in Fourier space is proportional to the wavevector k.

20.4 Red shift space

A galaxy survey measures the sky angles (θ,ϕ) and the redshift z of galaxies. The position of a galaxy in configuration space (position space) is

$$\mathbf{x}(z,\theta,\phi) = \chi_{true}\hat{\mathbf{n}}(\theta,\phi) \tag{20.6}$$

where χ_{true} is the true (not measurable) comoving distance of the galaxy. We define the 3-dimensional position of the galaxy in **red shift space** as

$$\mathbf{x}_{obs}(z,\theta,\phi) = \chi(z_{obs})\hat{\mathbf{n}}(\theta,\phi) \tag{20.7}$$

The distance $\chi(z)$ is the comoving distance at red shift z (if z was only due to the Hubble expansion). The function $\chi(z)$ depends on cosmological parameters, and we evaluate it at some **fiducial cosmological parameters**. The fact that these parameters are not exactly known is also important, and leads to the Alcock-Paczynski effect that we will discuss below. For now assume that the cosmological parameters are known.

In reality, galaxies do move with respect to the background frame and their redshift is given by the Hubble flow and their peculiar velocity $\boldsymbol{u}$ as

$$1+z = \frac{1}{a_{em}}\left(1+\boldsymbol{u}_g\cdot\hat{\mathbf{n}}\right) = \frac{1}{a_{em}}\left(1+u_{||}\right) \tag{20.8}$$

where a_{em} is the scale factor at which the light from the galaxy was emitted (the above is a non-relativistic approximation, galaxies don't move faster than $\sim 1\%$ of the speed of light). The observed position of the galaxy in red shift space $\mathbf{x}_{obs}$ is thus given by a correction $\Delta\mathbf{x}_{RSD}$ to the true position $\mathbf{x}$ of the galaxy as

$$\mathbf{x}_{obs} = \mathbf{x} + \Delta\mathbf{x}_{RSD} \tag{20.9}$$

$$= \mathbf{x} + \frac{u_{||}(\mathbf{x})}{aH}\hat{\mathbf{n}} \tag{20.10}$$

where RSD means **red shift space distortion**. Of course, we don't know $u_{||}$ of a given galaxy.

Red shift space distortions are not only a problem. Because they are sensitive to velocities, they can also be used to constain cosmology, in particular through the growth factor.

20.5 Redshift space distortions of the density field

To measure cosmological parameters, we need to be able to calculate the observed galaxy power spectrum with RSD included. On linear scales, this effect was derived by Kaiser in 1987 and leads to the **Kaiser red shift term**. We will only summarize the calculation, see Dodelson for more details. The starting point is the observation that, since RSD neither creates nor destroys galaxies, the densities in red shift space and configuration space must be related by

$$n_{g,obs}(\mathbf{x}_{obs})d^3x_{obs} = n_g(\mathbf{x})d^3x \tag{20.11}$$

We can write the volume element in spherical coordinates as $d^3x = x^2 dx d\Omega$ and $d^3x_{obs} = x_{obs}^2 dx_{obs} d\Omega$ where $d\Omega$ is the same in both coordinates. Therefore the densities are related by a Jocobian J as

$$n_{g,obs}(\mathbf{x}_{obs}) = J\, n_g(\mathbf{x}) \tag{20.12}$$

with

$$J \equiv \left|\frac{d^3x}{d^3x_{\rm obs}}\right| = \left|\frac{dx}{dx_{\rm obs}}\right| \frac{x^2}{x_{\rm obs}^2} \tag{20.13}$$

The Jacobian can be calculated and simplified (on large scales) to

$$J \approx 1 - \frac{1}{aH}\frac{\partial}{\partial x} u_{\parallel} \tag{20.14}$$

For density perturbations $\delta = \bar{n}(1+\delta)$ it follows (to first order in perturbations) that

$$1 + \delta_{g,\rm obs}(\mathbf{x}_{\rm obs}) = \left[1 + \delta_g(\mathbf{x}[\mathbf{x}_{\rm obs}]) - \frac{1}{aH}\frac{\partial}{\partial x} u_{\parallel}(\mathbf{x}[\mathbf{x}_{\rm obs}])\right] \tag{20.15}$$

We now have the building blocks that we need to calculate the galaxy power spectrum. We first note that in the above equation we can set $\mathbf{x}_{\rm obs} = \mathbf{x}$ at lowest order in the perturbations. This is because expanding the arguments of δ_g and $\boldsymbol{u}$ would lead to higher order terms that would be small. We also use linear galaxy bias to express δ_g in terms of δ_m. We can also equal the galaxy velocity $\boldsymbol{u}_g$ to the matter velocity $\boldsymbol{u}_m$. Physically this is because the velocities are sourced by the attraction of all the matter in the universe, not just that of galaxies. With these approximations we get

$$\delta_{g,\rm RSD}(\mathbf{x}) = b_1\delta_m(\mathbf{x}) - \frac{\partial}{\partial \mathbf{x}}\left[\frac{\boldsymbol{u}_m(\mathbf{x})\cdot\hat{\mathbf{x}}}{aH}\right] \tag{20.16}$$

Next we introduce the **distant observer approximation**, also called the **plane parallel approximation**. The idea is to take the line of sight $\hat{\mathbf{x}}$ to agree with the z-axis and treat it as fixed, neglecting changes from galaxy to galaxy. This is justified for galaxies that are relatively nearby on the sky. We can then replace $\boldsymbol{u}_m(\mathbf{x})\cdot\hat{\mathbf{x}} \to \boldsymbol{u}_m(\mathbf{x})\cdot\hat{\mathbf{e}}_z$. Using the distant observer approximation we can evaluate the Fourier transform $\delta_{g,\rm RSD}(\mathbf{k})$ as follows:

$$\delta_{g,\rm RSD}(\mathbf{k}) = \int d^3x\, e^{-i\mathbf{k}\cdot\mathbf{x}}\left[b_1\delta_m(\mathbf{x}) - \frac{\partial}{\partial x}\left(\frac{\boldsymbol{u}_m(\mathbf{x})\cdot\hat{\mathbf{e}}_z}{aH}\right)\right] \tag{20.17}$$

which can be evaluated, using Eq. 20.5 for the velocities, to give

$$\delta_{g,\mathrm{RSD}}(\mathbf{k}) = [b_1 + f\mu_k^2]\delta_m(\mathbf{k}) \tag{20.18}$$

where $\mu_k = \hat{\mathbf{e}}_z \cdot \hat{\mathbf{k}}$ is the vector between the line of sight and the perturbation. This is called the **Kaiser redshift space distortion**. The apparent overdensity in redshift space is thus larger than in configuration space (except for transverse perturbations where $\mu = 0$).

20.6 Redshift space distortions of the galaxy power spectrum

By squaring the RSD galaxy density contrast, and reintroducing shot noise, we find that the linear power spectrum is given by

$$P_{g,\mathrm{RSD}}(k, \mu_k, z) = P_L(k, z)\left[b_1 + f\mu_k^2\right]^2 + P_N \tag{20.19}$$

The redshift dependent power spectrum is usually expanded as

$$P_{g,\mathrm{RSD}}^{(l)}(k) = \frac{2l+1}{2}\int_{-1}^{1} d\mu_k P_l(\mu_k) P_{g,\mathrm{RSD}}(k, \mu_k) \tag{20.20}$$

using Legendre polynomials (as appropriate for an azimuthally symmetric function). The power spectrum is then

$$P_{g,\mathrm{obs}}(k, \mu_k) = \sum_l P_l(\mu_k) P_{g,\mathrm{obs}}^{(l)}(k) \tag{20.21}$$

By plotting the monopole $l = 0$, quadrupole ($l = 2$) and hexadecapole $l = 4$ (the other ones are negligibly small), one can avoid plotting the μ dependence. Most of the signal-to-noise is in the monopole.

20.7 Alcock–Paczynski effect

An additional distortion to the observed power spectrum comes from the fact that the cosmological parameters are not precisely known (in fact, we want to measure them from the power spectrum). Therefore our fiducial relation between χ and z has some error

$$\chi_{\mathrm{fid}}(z) = \chi(z) + \delta\chi(z) \tag{20.22}$$

One can again propagate this error through the Jacobian as we did in our derivation of RSD. This leads to an additional anisotropy of the measured power spectrum. The derivation can be found in Dodelson-Schmidt 11.1.3.

20.8 Red shift binned angular correlation functions

For photometric surveys, we don't have precise individual redshifts for galaxies. Instead, the photometry can only be used to split galaxies roughly into **redshift bins**. For each red shift bin, the redshift density of sources is given by a window function of form

$$W(\chi) = \frac{1}{N_g}\frac{dN_g}{d\chi} \tag{20.23}$$

where W is normalized to unity and drops to zero outside of the interval.

The angular galaxy density in the bin is then given by

$$\Delta_g(\hat{n}) = \int_0^\infty d\chi W(\chi)\delta_{g,\rm obs}(x = \hat{n}\chi, \tau = \tau_0 - \chi) \tag{20.24}$$

Going to multipole space we get

$$\Delta_{g,lm} = 4\pi i^l \int \frac{d^3k}{(2\pi)^3} Y^*_{lm}(\hat{\mathbf{k}}) \int_0^\infty d\chi W(\chi) j_l(k\chi)\delta_{g,\rm obs}(\mathbf{k}, \tau(\chi)) \tag{20.25}$$

The power spectrum

$$\langle \Delta_{g,lm}\Delta^*_{g,l'm'}\rangle = \delta_{ll'}\delta_{mm'}C_g(l) \tag{20.26}$$

is given by

$$C_g(l) = \frac{2}{\pi}\int k^2 dk \int_0^\infty d\chi W(\chi) j_l(k\chi) \int_0^\infty d\chi' W(\chi') j_l(k\chi') P_{g,\rm obs}(k, \tau(\chi), \tau(\chi')). \tag{20.27}$$

Note that this includes a non-equal time power spectrum, which takes into account the different times probed due to the light-cone. In the so-called **Limber approximation** this becomes

$$C_g(l) = \int \frac{d\chi}{\chi^2} W^2(\chi) P_{g,\rm obs}\left(k = \frac{l+1/2}{\chi}, \tau(\chi)\right) \tag{20.28}$$

This approximation avoids evaluating the Bessel function integrals and is used in many cosmology papers. The Limber approximation is valid if the radial extent of the bin is much larger than the scale of the angular scale of the multipole l under consideration. More about the accuracy of the Limber approximation can be found here: 0809.5112.

21 Overview of LSS Perturbation Theory

In the next sections we study structure formation, i.e. how the universe grows from small initial inhomogeneities to the cosmic web we observe today. This process can be treated to good approximation using Newtonian physics on a flat expanding spacetime. Newtonian gravity is indeed used in the vast majority of research about structure formation, with "relativistic corrections" being an active topic of research. Both perturbative calculations in large-scale structure and N-body simulations can be done to high precision while neglecting relativistic effects (but taking into account the expansion of spacetime of course). We start with a discussion of analytic perturbation theory. Perturbation theory, in the modern EFT version, is still the state of the art in extracting cosmological parameters from large-scale structure surveys.

21.1 Fluid approximation

In Newtonian perturbation theory our goal is to calculate how the matter density $\rho(\mathbf{x})$ evolves in time. We will make the approximation that matter is a **collisionless fluid**, i.e that it consists of a continuous matter density that interacts only gravitationally. Clearly, the observed matter in the universe looks very different, it contains different forms of matter and clumps into galaxies which

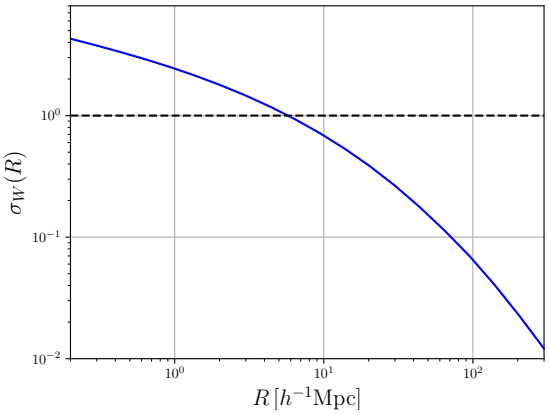


Figure 20. The standard deviation of the density field, Eq. (21.3), when smoothed over different scales R, where R is the width of the smoothing filter in position space, at redshift $z = 0$. The value at $R = 8h^{-1}$Mpc is the definition of the common cosmological parameter σ_8.

have complicated non-gravitational physics. Nevertheless, the collisionless fluid approximation works on large enough scales and can be systematically improved on intermediate scales to include complicated small-scale physics in the **effective field theory of inflation** which we will outline below.

Let's first see why perturbation theory is possible and on what scales. To do so, we **define** the filtered density field $\delta_W(\mathbf{x})$,

$$\delta_W(\mathbf{x}) = \int d^3y\, W(|\mathbf{x}-\mathbf{y}|)\delta_m(\mathbf{y}), \tag{21.1}$$

where $W(\mathbf{x})$ is the filtering kernel that we can take to be isotropic. This filtering corresponds to a multiplication in Fourier space:

$$\delta_W(\mathbf{k}) = W(\mathbf{k})\delta_m(\mathbf{k}), \tag{21.2}$$

where $W(\mathbf{k})$ is the Fourier transform of the isotropic filtering kernel, such as a real-space tophat. The variance of the filtered field is

$$\sigma_W^2 \equiv \langle(\delta_W)^2(\mathbf{x})\rangle = \int \frac{d^3k}{(2\pi)^3}\int \frac{d^3k'}{(2\pi)^3}\langle\delta_W(\mathbf{k})\delta_W^*(\mathbf{k}')\rangle e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}} \tag{21.3}$$

$$= \int \frac{d^3k}{(2\pi)^3} P_L(k)|W(\mathbf{k})|^2 \tag{21.4}$$

$$= \frac{1}{2\pi^2}\int d\ln k\, k^3 P_L(k)|W(\mathbf{k})|^2. \tag{21.5}$$

which is plotted in Fig. 20 as a function of the smoothing scale. We see when perturbations get smaller than one, indicating that a perturbative expansion in δ_k is possible.

21.2 Standard (Eulerian) Perturbation Theory

We now briefly review cosmological perturbation theory. My discussion follows the introduction by Philcox. The standard review is astro-ph/0112551.

21.2.1 Equations of motion

To describe the universe as a fluid we need the following variables

- $\delta(\mathbf{x}, \tau)$: Overdensity of matter related to the the density $\rho(\mathbf{x}, \tau)$ by $\delta(\mathbf{x}, \tau) = \rho(\mathbf{x}, \tau)/\bar{\rho}(\tau) - 1$
- $\mathbf{v}(\mathbf{x}, \tau)$: Fluid velocity. Note that in the fluid approximation we cannot describe a situation where matter clumps of different velocity pass through each other.
- $\phi(\mathbf{x}, \tau)$: Peculiar gravitational potential (corrected for the background expansion)
- $\sigma_{ij}(\mathbf{x}, \tau)$: Viscous stress tensor. $\sigma_{ij} = 0$ for a perfect fluid, which we consider here, but it becomes important in the EFTofLSS.

In the collisionless fluid approximation we consider the total matter distribution, dark matter and baryons, together.

The equations of motion, which can be derived from the **collisionless Boltzmann equation**, in the Newtonian limit, are

- The **Continuity equation**

$$\dot{\delta}(\mathbf{x}, \tau) + \nabla \cdot [(1 + \delta(\mathbf{x}, \tau))\mathbf{v}(\mathbf{x}, \tau)] = 0 \tag{21.6}$$

  Here and below dots indicate derivatives with respect to conformal time.

- The **Euler equation**

$$\dot{\mathbf{v}}(\mathbf{x}, \tau) + [\mathbf{v}(\mathbf{x}, \tau) \cdot \nabla]\mathbf{v}(\mathbf{x}, \tau) = -\mathcal{H}(\tau)\mathbf{v}(\mathbf{x}, \tau) - \nabla\phi(\mathbf{x}, \tau) \tag{21.7}$$

  where $\mathcal{H} = aH$ is the comoving Hubble parameter. The Euler equation is the equivalent to $\mathbf{F} = m\mathbf{a}$ for a fluid element. The left hand side is the "convective time derivative" and the right hand side has a force term due to the gravitational potential and a term due to the Hubble expansion.

- The **Poisson equation**

$$\nabla^2\phi(\mathbf{x}, \tau) = 4\pi G a^2(\tau)\bar{\rho}(\tau)\delta(\mathbf{x}, \tau) = \frac{3}{2}\mathcal{H}^2(\tau)\Omega_m(\tau)\delta(\mathbf{x}, \tau) \tag{21.8}$$

One can solve these equations perturbatively on scales where these perturbations are small, so that the pertubative expansion converges.

21.2.2 Linear solutions

To find the linear solution that describes the universe on large scales, we linearize the fluid equations, i.e., dropping any terms of second or higher order in $\{\delta, \mathbf{v}, \phi\}$. Introducing the velocity potential

$$\theta(\mathbf{x}, \tau) = \nabla \cdot \mathbf{v}(\mathbf{x}, \tau) \tag{21.9}$$

this yields the following equations for the first-order fields, $\delta_1, \mathbf{v}_1$:

$$\theta_1(\mathbf{x}, \tau) = -\dot{\delta}_1(\mathbf{x}, \tau) \tag{21.10}$$

$$\ddot{\delta}_1(\mathbf{x}, \tau) + \mathcal{H}(\tau)\dot{\delta}_1(\mathbf{x}, \tau) - \frac{3}{2}\mathcal{H}^2(\tau)\Omega_m(\tau)\delta_1(\mathbf{x}, \tau) = 0. \tag{21.11}$$

where we eliminated the peculiar potential. These are solved by a separable solution, such that

$$\delta_1(\mathbf{x}, \tau) = D(\tau)\delta_L(\mathbf{x}) \tag{21.12}$$

$$\theta_1(\mathbf{x}, \tau) = -\mathcal{H}(\tau)f(\tau)D(\tau)\delta_L(\mathbf{x}), \tag{21.13}$$

where $\delta_L(\mathbf{x})$ is the **linear density field** set by inflation (and k-dependent transfer functions that take into account mode evolution in the early universe, see Sec. 9.4.1). This drops a "decaying mode". The **growth factor** is given by the integral solution

$$D(\tau) = D_0\mathcal{H}(\tau)\int_0^{a(\tau)} \frac{da'}{\mathcal{H}^3(a')}, \tag{21.14}$$

where D_0 ensures the normalization condition $D(a = 1) = 1$ today. For an Einstein-de-Sitter Universe (with $\Omega_m = 1$), $D(\tau)$ is simply the scale factor $a(\tau)$. For the velocity, we introduced the **(velocity) growth rate**

$$f(\tau) \equiv \frac{d\log D(\tau)}{d\log a} \tag{21.15}$$

We see that densities evolve according to $D(\tau)$ while velocities are enhanced by a factor of $\mathcal{H}(\tau)f(\tau)$.

Switching to Fourier-space, we obtain:

$$\delta_1(\mathbf{k}, \tau) = D(\tau)\delta_L(\mathbf{k}) \tag{21.16}$$

$$\theta_1(\mathbf{k}, \tau) = -\mathcal{H}(\tau)f(\tau)D(\tau)\delta_L(\mathbf{k}), \tag{21.17}$$

and

$$\mathbf{v}(\mathbf{k}, \tau) = i(\mathbf{k}/k^2)\theta(\mathbf{k}, \tau). \tag{21.18}$$

It follows that the linear-order matter power spectrum is given by

$$P^{\rm SPT}_{\rm linear}(\mathbf{k}, \tau) = \langle\delta_1(\mathbf{k}, \tau)\delta_1(-\mathbf{k}, \tau')\rangle' = D^2(\tau)P_L(\mathbf{k}), \tag{21.19}$$

where $P_L(\mathbf{k})$ is the power spectrum of the initial conditions. We have dropped a momentum-conserving Dirac delta function (indicated by the prime in the expectation value $\langle\rangle'$, as is often done).

21.2.3 General perturbative solution

The general perturbative solution works by first rewriting the fluid equations in Fourier space, and then making a series solution, expanding the equations order-by-order in the (assumed small)

parameters δ and θ. Explicitly, we begin with the series solutions

$$\delta(\mathbf{k},\tau) = \sum_{n=1}^{\infty} D^n(\tau)\delta^{(n)}(\mathbf{k}) \tag{21.20}$$

$$\theta(\mathbf{k},\tau) = -\mathcal{H}(\tau)f(\tau)\sum_{n=1}^{\infty} D^n(\tau)\theta^{(n)}(\mathbf{k}), \tag{21.21}$$

where the n-th order solution contains n copies of the linear solution, $\delta^{(1)}(\mathbf{k}) = \delta_L(\mathbf{k})$. We have assumed separability in time and space which is an excellent approximation (and exact for Einstein de-Sitter universes), though deviations can occur at high order. The n-th order contribution takes the form:

$$\delta^{(n)}(\mathbf{k}) = \int_{\mathbf{p}_1\cdots\mathbf{p}_n} F_n(\mathbf{p}_1,\ldots,\mathbf{p}_n)\delta^{(1)}(\mathbf{p}_1)\ldots\delta^{(1)}(\mathbf{p}_n)(2\pi)^3\delta_D(\mathbf{p}_1+\ldots+\mathbf{p}_n-\mathbf{k}), \tag{21.22}$$

$$\theta^{(n)}(\mathbf{k}) = \int_{\mathbf{p}_1\cdots\mathbf{p}_n} G_n(\mathbf{p}_1,\ldots,\mathbf{p}_n)\delta^{(1)}(\mathbf{p}_1)\ldots\delta^{(1)}(\mathbf{p}_n)(2\pi)^3\delta_D(\mathbf{p}_1+\ldots+\mathbf{p}_n-\mathbf{k}). \tag{21.23}$$

This is the convolution of n linear density fields with a kernel, F_n or G_n. The kernels up to second order are given by:

$$F_1(\mathbf{p}) = 1, \quad F_2(\mathbf{p}_1,\mathbf{p}_2) = \frac{5}{7}\alpha(\mathbf{p}_1,\mathbf{p}_2) + \frac{2}{7}\beta(\mathbf{p}_1,\mathbf{p}_2), \tag{21.24}$$

$$G_1(\mathbf{p}) = 1, \quad G_2(\mathbf{p}_1,\mathbf{p}_2) = \frac{3}{7}\alpha(\mathbf{p}_1,\mathbf{p}_2) + \frac{4}{7}\beta(\mathbf{p}_1,\mathbf{p}_2) \tag{21.25}$$

where

$$\alpha(\mathbf{p}_1,\mathbf{p}_2) = \frac{\mathbf{p}_1\cdot\mathbf{k}}{p_1^2}; \quad \beta(\mathbf{p}_1,\mathbf{p}_2) = \frac{k^2\mathbf{p}_1\cdot\mathbf{p}_2}{2p_1^2p_2^2}; \quad \mathbf{k} = \mathbf{p}_1+\mathbf{p}_2. \tag{21.26}$$

These integrals in general have to be evaluated numerically. The integrals in principle go up to infinite momenta, where perturbations are not small and physics is not perturbative. This is an inconsistency in SPT, that is fixed in the **effective field theory of large-scale structure**, which starts from a **smoothing** of the underlying fields. We'll get back to this issue in Sec. 22.

The above equations in principle allow us to compute density (and velocity) field statistics at arbitrary order. The most important and basic one is the equal-time power spectrum, $P(k,\tau) = \langle\delta(\mathbf{k},\tau)\delta(-\mathbf{k},\tau)\rangle$ which can be written in terms of δ^n correlators as

$$P^{\rm SPT}(k,\tau) = D^2(\tau)P^{(11)}(k) + D^4(\tau)\left[P^{(13)}(k) + P^{(22)}(k)\right] + \ldots \tag{21.27}$$

where $P^{(ij)}(k) = \langle\delta^{(i)}(\mathbf{k})\delta^{(j)}(-\mathbf{k})\rangle$, and we have assumed Gaussian initial conditions, such that any correlator involving an odd number of linear density fields vanishes. In the same way we can compute higher-order correlators. The next most important one is the three-point function, or bispectrum, which at lowest order is given by

$$B(k_1,k_2,k_3,\tau) = \langle\delta(\mathbf{k}_1,\tau)\delta(\mathbf{k}_2,\tau)\delta(\mathbf{k}_3,\tau)\rangle' = D^4(\tau)B^{(211)}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3) + \ldots \tag{21.28}$$

with higher-order contributions containing loop integrals over the linear power spectrum.

21.3 Lagrangian Perturbation theory (LPT)

There is a second important way to perform perturbation theory, in a different set of variables. This is the Lagrangian formulation. Let's briefly outline this approach. One can describe a fluid in two ways:

- In the **Eulerian picture** described above, we describe the matter density $\rho(\mathbf{x}, t)$ and the velocity field $\mathbf{v}(\mathbf{x}, t)$ as a function of a fixed spatial coordinate $\mathbf{x}$.

- In the **Lagrangian picture**, instead of working with densities, we describe the movement of particles (or fluid elements) from their initial comoving coordinate $\mathbf{q}$ to their later comoving Eulerian coordinate $\mathbf{x}$ by defining the **displacement field** $\Psi(\mathbf{q}, \tau)$ so that

$$\mathbf{x}(\tau) = \mathbf{q} + \Psi(\mathbf{q}, \tau).$$

  All coordinates are comoving, so the expansion of the Universe does not change them. Note that $\Psi = 0$ initially so that $\mathbf{q}$ is the same as the usual comoving coordinate at initial time, $\tau = 0$. Once we have calculated the displacement field, using Lagrangian perturbation theory, we can estimate the observable density field $\rho(\mathbf{x}, t)$ from it.

Lagrangian perturbation theory looks similar to SPT, i.e. we can calculate a series solution of form

$$\Psi(\mathbf{q}, \tau) = \sum_{n=0}^{\infty} D^n(\tau) \Psi^{(n)}(\mathbf{q}), \tag{21.29}$$

As in the Eulerian case, the n-th order solution can be written as a convolution over n copies of the linear density field δ_L:

$$\Psi^{(n)}(\mathbf{k}, \tau) = \frac{i}{n!} \int_{\mathbf{p}_1 \dots \mathbf{p}_n} L_n(\mathbf{p}_1, \dots, \mathbf{p}_n) \delta_L(\mathbf{p}_1) \dots \delta_L(\mathbf{p}_n) (2\pi)^3 \delta_D(\mathbf{p}_1 + \dots + \mathbf{p}_n - \mathbf{k}), \tag{21.30}$$

However, the integrals over the kernels are in general harder to evaluate than those of SPT.

Some comments on the relation of Eulerian and Lagrangian PT:

- The first order LPT solution called the **Zeldovich approximation** and its second order extension called **2-LPT** are remarkably good at reproducing the full non-linear density field at intermediate scales. They outperform the first and second order SPT solutions substantially. However, by including so called IR resummation, one can improve SPT and ultimately both Eulerian and Lagrangian perturbation theory give equivalent results (see e.g. Senatore's EFTofLSS lecture notes).

- The Zeldovich approximation (1-LPT) and 2-LPT are used to set up initial conditions for N-body simulations. N-body simulations track particles, so the use of Lagrangian particle displacements makes intuitive sense. The reason why N-body simulations need perturbation theory is to set up initial particle displacements (of equal mass particles) that incorporate the initial inhomogeneities from inflation, as well as to speed up computation time by treating small density fluctuations analytically until they grow sufficiently to require N-body simulation.

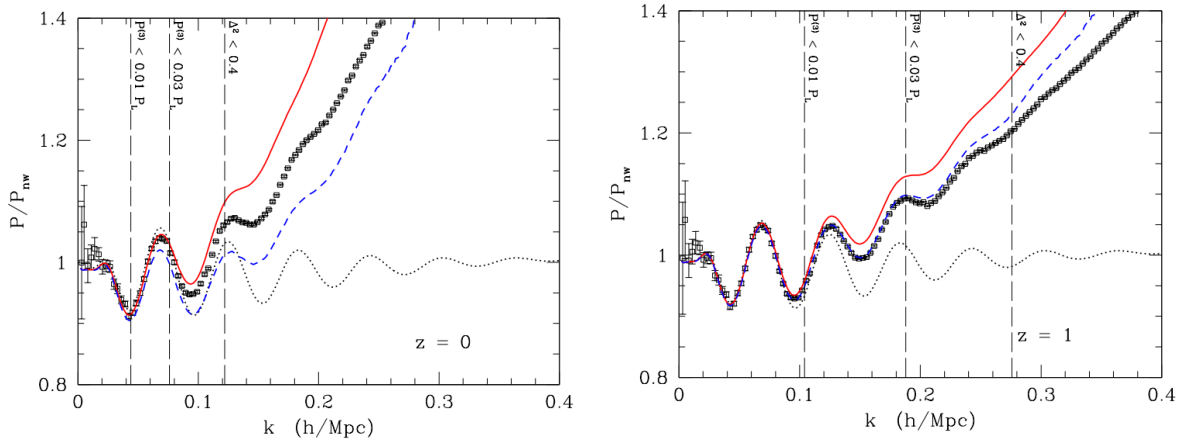


Figure 21. Plots showing a comparison of N-Body data (black boxes) with theoretical SPT power spectra at tree level (dotted), one loop (solid red), and two loop (dashed blue) orders. The left and right plots show the comparison at redshifts 0 and 1 respectively. Each curve has been divided by the no-wiggle (broadband) power spectrum for clarity of range. The plots are taken from 0905.0479.

- Observations only provide us with Eulerian densities, since we cannot look back in time to observe the movement of a chunk of matter to its initial position. Observations are thus closer to Eulerian theory. However N-body simulations readily provide both displacement fields and densities.

The dual description of structure formation in the Eulerian and Lagrangian picture continues to be important even for machine learning based methods. For example, a neural network structure formation emulator can either be trained to output Eulerian density fields $\rho(\mathbf{x})$ or to output the displacement field ψ of particles, and indeed both have been tried.

22 Effective Field Theory of Large-Scale Structure*

***This section was developed and taught by Sai Chaitanya Tadepalli.**

22.1 Problems with SPT

In previous sections, we discussed the Standard Perturbation Theory (SPT) of the matter overdensity in an expanding universe during the matter-domination era. The derivation of SPT inherently assumes that the distribution of matter on large-scales can be treated as pressureless and collision-less fluid. Clearly this assumption has certain drawbacks and fails to accurately predict the matter power spectrum even on large scales where it is supposed to perform very well (i.e. on scales where the matter overdensity variance is much less than unity and hence certainly perturbative).

To visualize the performance of SPT, consider the plot shown in Fig. 21 where we show the SPT matter power spectrum fitting at linear, one, and two-loop order to the data obtained from numerical N-Body simulations. At the outset, we observe that the SPT performs well at very large scales ($k \sim O(10)H_0/c \approx 0.003$h/Mpc) where the residual is sub-percent. On these scales,

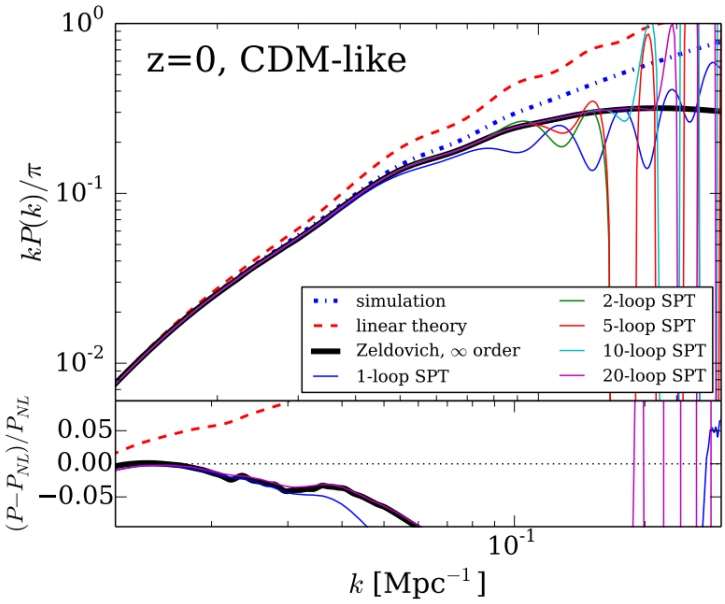


Figure 22. Top panel: $z = 0$ matter overdensity power spectrum in a 1D CDM-like model calculated analytically using linear theory, the Zeldovich approximation (LPT at any order), and SPT to the specified order in the overdensity. Note that even a 20 loop order SPT does not perform any better than a two loop SPT. Figure is taken from 1502.07389

the universe is close to a perfect fluid and traces the initial conditions very well (which implies minimal mode mixing). However, linear SPT begins to fail at scales $\sim O(0.01)$ h/Mpc at $z = 0$. One might consider adding the next order terms in PT, the one-loop terms $P^{(13)}$ and $P^{(22)}$, to our theoretical fit to improve the range of SPT. This is shown by the solid red curve in Fig. 21. Clearly, the one-loop SPT does not improve our fit better than the linear theory. Here, we may be tempted to add higher loop order terms such as second and third to improve the fit. This is shown in the dashed blue curve where we show the performance of SPT up to two loops. Interestingly, adding higher-order loops does not improve our fit. The fitted curve appears to oscillate around the true data points. This exercise when carried up to as large as 10 loop orders reveals a similar pattern, as illustrated in Fig. 22. Hence, we deduce that SPT fails to fit the nonlinear matter power spectrum on scales $k \ll 0.3\text{h/Mpc} \equiv \text{k}_{\text{NL}}(\text{z} = 0)$. Therefore, SPT needs to be improved.

When deriving SPT, the solution to the nonlinear coupled equations (Euler, Poisson and continuity) of the matter overdensity contrast $\delta(k)$ in Fourier space was given in terms of corrections to the linear solution $\delta^{(1)}$:

$$\delta(k) = \delta^{(1)}(k) + \delta^{(2)}(k) + \delta^{(3)}(k) + ... \tag{22.1}$$

where each nonlinear correction is given by

$$\delta^{(n)}(k) = \int d^3q_1...d^3q_n \; \delta^n \left(\vec{k} - \sum_i \vec{q}_i \right) F_n \left(q_1, ..., q_n \right) \delta^{(1)}(q_1).....\delta^{(1)}(q_n) \tag{22.2}$$

with $F_n(..)$ the symmetrized kernel of the nth-order solution.

The above expansion hinges on the assumption of perturbativity which requires that each nth order correction must be smaller than the $(n-1)$th order term. This is needed for the perturbative solution to exist in the first place. However, as we will show below the loop terms inherently contain contributions from the internal momenta (or modes) where our perturbation theory is bound to break down. This lack of a clear small expansion parameter in the SPT is the prime reason for its failure. To this end, consider the one-loop term $P^{(13)}$ as given below

$$P^{(13)}(k) = \left\langle \delta_k^{(1)} \delta_p^{(3)} \right\rangle' + \left\langle \delta_k^{(3)} \delta_p^{(1)} \right\rangle' \tag{22.3}$$

$$= 6P^{(11)}(k) \int \frac{d^3q}{(2\pi)^3} F_3(\vec{k}, \vec{q}, -\vec{q}) P^{(11)}(q) \tag{22.4}$$

where $P^{(11)}(k) = \left\langle \delta_k^{(1)} \delta_p^{(1)} \right\rangle'$ is the linear matter power spectrum and $'$ denotes that we have absorbed the Dirac delta function and the factor of $(2\pi)^3$. On very large scales, i.e. in the limit $k \to 0$, the kernel $F_3 \to k^2/q^2$. Hence, we find the UV ($k/q \to 0$) limiting behavior of $P^{(13)}$ is

$$\lim_{k\to 0} P^{(13)}(k) \approx -\frac{61}{630\pi^2} k^2 P^{(11)}(k) \int_0^\infty dq P^{(11)}(q). \tag{22.5}$$

The integral in the above expression goes over all internal momenta q and hence the integrand $P^{(11)}(q)$ is evaluated on very small scales. This raises serious concerns as the linear power spectrum $P^{(11)}$ is not valid on scales beyond $\sim k_{NL}$ and yet we are summing over all scales down to those of individual galaxies, stars, planets and even dust!

The concerns over the summation of internal momenta over very small scales leads to a related issue with the SPT. Let us consider that the linear power spectrum can be approximated by a power-law form

$$P^{(11)}(k) \propto k^n. \tag{22.6}$$

This is an excellent approximation on very large and quasi-linear scales where $n \approx 1$ and ≈ -1.5 respectively. These values are reflective of our universe where the initial conditions are governed dominantly by adiabatic initial (primordial) conditions. Upon substituting the power-law linear power spectrum into the UV limiting integral for $P^{(13)}$, we find that the integral diverges if $n \geq -1$. Fortunately, for adiabatic initial conditions, the spectral index $n \to -3$ as $k \to \infty$ and hence the integral converges. However, in models extending beyond a pure adiabatic assumption, such as those featuring a small fraction of primordial large blue-tilted ($n > -1$) isocurvature fluctuations, the integral becomes divergent. Consequently, the SPT fails to meet the general requirement of applicability to arbitrary initial conditions (see 1301.7182 for details)

In the next subsection, we will show how these problems can be ameliorated by the EFTofLSS formalism.

22.2 Coarse graining and effective fluid

The primary goal of the EFTofLSS program is straightforward: to develop a consistent perturbation theory for the expanding Universe which is convergent, accurate, and can be applied in the presence of arbitrary initial conditions.

The key problem in the SPT was the evaluation of loop integrals over scales where the internal propagator (linear power spectrum) is known to break down. Hence, a straightforward solution is to regulate these integrals by evaluating them up to a finite cutoff scale Λ. Similar to the 'cutoff regularization' in QFT, we can evaluate the UV limit of the one loop term $P^{(13)}$ up to a scale Λ as

$$\lim_{k\to 0} P^{(13)}(k,\Lambda) \sim k^2 P^{(11)}(k) \int_0^{\Lambda} dq P^{(11)}(q). \tag{22.7}$$

By choosing $\Lambda \ll k_{\rm NL}$ we are guaranteed that the integrand $P^{(11)}(q)$ is evaluated on perturbative scales. This seemingly simple solution has an inherent problem. By choosing an arbitrary cutoff scale Λ, we have made our final SPT evaluations Λ-dependent. This is easy to observe for choices of cutoff scales $\Lambda_1 < \Lambda_2 \ll k_{\rm NL}$ such that

$$P^{(13)}(k,\Lambda_2) = P^{(13)}(k,\Lambda_1) + 6P^{(11)}(k) \int_{\Lambda_1}^{\Lambda_2} \frac{d^3q}{(2\pi)^3} F_3(\vec{k},\vec{q},-\vec{q}) P^{(11)}(q). \tag{22.8}$$

In the limit $k \ll \Lambda_1$, we can approximate the integral in above expression using the UV limit derived earlier. Hence we obtain

$$P^{(13)}(k,\Lambda_2) = P^{(13)}(k,\Lambda_1) - \text{constant} \times k^2 P^{(11)}(k) \int_{\Lambda_1}^{\Lambda_2} dq P^{(11)}(q), \tag{22.9}$$

$$= P^{(13)}(k,\Lambda_1) - k^2 P^{(11)}(k) \left[f(\Lambda_2) - f(\Lambda_1) \right]. \tag{22.10}$$

However, our true data points either from an N-body simulation or observed samples are Λ-independent. Hence, even though we have made a positive step in finding a resolution for the failure of SPT, we have introduced an arbitrary scale in our theory which may be physically motivated but is not an accurate description of the data.

The cutoff regularization procedure suggests that the SPT can be explicitly restricted to scales $k < \Lambda$ where Λ is some coarse-graining scale. Hence, we must look for a new 'effective' theory that applies to perturbative long-wavelength modes. This reminds us of the effective field theory (EFT) approach in QFT. In the EFT picture, we define the partition function Z of our theory as

$$Z = \int D\phi e^{iS[\phi]} \tag{22.11}$$

where $S[\phi]$ is an action and is a functional of the field ϕ. EFT hinges on the argument that to describe a low energy regime of field configurations at $k \ll \Lambda$, we do not need high energy momentum field modes. To visualize this, consider a complete UV theory where we can factorize the underlying ϕ field in terms of long and short scales $\phi = \phi_l + \phi_s$. Thus,

$$Z = \int D\phi e^{iS[\phi]} = \int D\phi_l D\phi_s e^{iS[\phi_l,\phi_s]}. \tag{22.12}$$

Next, we integrate over all short-scale modes and obtain

$$Z = \int D\phi_l e^{iS[\phi_l]}. \tag{22.13}$$

Since the partition function remains consistent in both descriptions, the actions, denoted as $S[\phi_l]$ and $S[\phi]$, differ. The new action $S[\phi_l]$ yields a low-energy effective theory, capturing the evolution of the field theory by integrating out ultraviolet (UV) modes and applying rescaling. It is important to note that the new low-energy effective theory action, $S[\phi_l]$ may incorporate residual effects from small-scale modes. The feedback of these small-scale modes on the large scale forms the essence of the success of the Effective Field Theory of Large-Scale Structure (EFTofLSS) formalism.

Similar to the above discussion, we propose that the matter overdensity field δ can be broken down into long and short scale (wavelength) modes where the long wavelength modes are chosen such that they are perturbative. This is achieved in principle by smoothing the matter overdensity field $\delta(x)$ within a smoothing radius $R \sim \Lambda^{-1}$. The smoothing procedure integrates over all short scale ($x < R$ or $k > \Lambda$) information. Hence, we define a new EFT of LSS that consists of smoothed field variables obtained by smoothing the δ, v and ϕ :

$$[\delta]_\Lambda \to \delta_l \qquad [\pi]_\Lambda \to \pi_l \qquad [\phi]_\Lambda \to \phi_l \tag{22.14}$$

where $[O]_\Lambda$ is the operation of smoothing over an operator O and $\pi = \rho v$ is the momentum density operator.

22.2.1 Brief derivation of EFTofLSS fluid equations

What is the EFT procedure? The starting point is the fluid EOMs. Let us begin with the Eulerian PT which aims at solving the system of three fluid equations: Poisson, Continuity, and Euler. Starting from an EdS cosmology, the equations can be written as

$$\nabla^2\phi - \frac{3}{2}\mathcal{H}^2\rho_0\delta = 0, \tag{22.15}$$

$$\partial_\tau\delta + \nabla\cdot\left[(1+\delta)\,\bar{v}\right] = 0, \tag{22.16}$$

$$\partial_\tau v + \mathcal{H}v + (v\cdot\nabla)\,v + \nabla\phi = 0. \tag{22.17}$$

Here, δ and v are the DM number-density fluctuation and peculiar velocity field respectively.

We can construct the equations of motion for an effective fluid by coarse-graining the fluid equations using a smoothing window function. The smoothing guarantees that the Boltzmann hierarchy can be truncated, leaving us with an effective fluid. We define our isotropic smoothing window function $W_\Lambda(\bar{x},\bar{x}')$ as a function of the radial separation r and smoothing radius Λ^{-1}:

$$W_\Lambda(\bar{x},\bar{x}') \equiv \mathcal{F}\left(r,\Lambda^{-1}\right) \tag{22.18}$$

where $r^2 = (x-x')^i\,(x-x')_i$. The isotropy of the window function implies

$$\int d^3x' W_\Lambda\left(\bar{x},\bar{x}'\right)\left(x-x'\right)^i = 0. \tag{22.19}$$

It is convenient to choose a normalized Gaussian function as our smoothing kernel:

$$W_\Lambda(\bar{x}-\bar{x}') = \left(\frac{\Lambda}{\sqrt{2\pi}}\right)^3 e^{-\frac{1}{2}\Lambda^2|\bar{x}-\bar{x}'|^2} \equiv \left(\frac{\Lambda}{\sqrt{2\pi}}\right)^3 e^{-\frac{1}{2}\Lambda^2(x-x')^i(x-x')_i}, \tag{22.20}$$

with its Fourier transform

$$W_\Lambda(k) = e^{-\frac{k^2}{2\Lambda^2}} \tag{22.21}$$

where Λ now represents a k-space, comoving cutoff scale. We regularize our observable quantities by smoothing them which is equivalent to taking convolution in real space with the filter (window function), defining the effective long wavelength quantity as

$$A_l(\bar{x}) = \int d^3x' W_\Lambda\left(\bar{x}, \bar{x}'\right) A(\bar{x}'), \tag{22.22}$$

and split the fields into short and long wavelength fluctuations by defining the short wavelength quantity as

$$A_s(\bar{x}) = A(\bar{x}) - A_l(\bar{x}). \tag{22.23}$$

In Fourier space, this is represented as

$$A_l(k) \equiv W_\Lambda(k) A(k), \tag{22.24}$$

$$A_s(k) \equiv (1 - W_\Lambda(k))\, A(k) \tag{22.25}$$

Specifically, for fields δ, v and ϕ the effective long-wavelength fluctuations are defined as

$$\delta_l(\bar{x}) = \int d^3x' W_\Lambda\left(\bar{x}, \bar{x}'\right) \delta(\bar{x}'), \tag{22.26}$$

$$\phi_l(\bar{x}) = \int d^3x' W_\Lambda\left(\bar{x}, \bar{x}'\right) \phi(\bar{x}'), \tag{22.27}$$

$$(1 + \delta_l(\bar{x}))\, \bar{v}_l(\bar{x}) = \int d^3x' W_\Lambda\left(\bar{x}, \bar{x}'\right) \left(1 + \delta(\bar{x}')\right) \bar{v}(\bar{x}'). \tag{22.28}$$

By applying the smoothing operation to the Euler, Poisson, and Continuity equations, and after numerous simplifications, we obtain the following set of fluid equations (see 1206.2926):

$$\nabla^2 \phi_l - \frac{3}{2}\mathcal{H}^2 \rho_0 \delta_l = 0, \tag{22.29}$$

$$\partial_\tau \delta_l + \nabla \cdot \left[(1 + \delta_l)\, \bar{v}_l\right] = 0, \tag{22.30}$$

$$\partial_\tau \vec{v}_l + \mathcal{H}\vec{v}_l + (\vec{v} \cdot \nabla)\, \vec{v}_l + \nabla\phi_l = -\frac{1}{\rho_l}\left(\partial_j \left[\tau_i^j\right]_\Lambda + \partial_j \left[\tau_i^j\right]_{\partial^2}\right). \tag{22.31}$$

where

$$\rho_l(x) \equiv \rho_0 d_l(x) = \rho_0 \left(1 + \delta_l(x)\right), \tag{22.32}$$

and

$$\left[\tau_i^j\right]_\Lambda = \left[\rho(\bar{x}') v_s(\bar{x}') v_s^j(\bar{x}') + \frac{2\partial^{j'}\phi_s(\bar{x}')\partial_{i'}\phi_s(x') - \delta_i^j \partial^{k'}\phi_s(\bar{x}')\partial_{k'}\phi_s(x')}{8\pi G}\right]_\Lambda \tag{22.33}$$

$$\left[\tau_i^j\right]_{\partial^2} = \left(\rho_l(\bar{x}) \frac{\partial_m v_l(\bar{x}) \partial^m v_l^j(\bar{x})}{\Lambda^2} + \frac{2\partial_k\partial_i\phi_l(x)\partial^k\partial^j\phi_l - \delta_i^j \partial_k\partial^m\phi_l(x)\partial^k\partial_m\phi_l}{8\pi G\Lambda^2}\right). \tag{22.34}$$

We see that the long-wavelength fluctuations obey an Euler equation in which the stress tensor τ^{ij} receives contributions from two terms that are induced by the short wavelength ($\left[\tau_i^j\right]_\Lambda$) and long-wavelength ($\left[\tau_i^j\right]_{\partial^2}$) fluctuations respectively. The long wavelength fluctuations are suppressed by $1/\Lambda^2$ factor and can be neglected in the limit $\Lambda \to \infty$. In the large Λ limit, the leading stress tensor is sourced by the short-wavelengths. These residual stress terms arise since multiplication and smoothing do not commute, i.e. $[AB]_\Lambda \neq [A]_\Lambda[B]_\Lambda$. Physically speaking, the intuition for the above stress tensor is that small scale modes appearing in the fluid equations non-linearly modify the dynamics of large scale modes.

Although we started with the EoM of a pressure-less fluid, the effective pressure of the 'imperfect' matter fluid after smoothing (in the limit $\Lambda \to \infty$) is given as

$$p_{\rm eff} = \frac{1}{3}\left[\tau_k^k\right]_\Lambda \tag{22.35}$$

$$= \frac{1}{3}\left(\left[\rho(\bar{x}')v_{s;k}(\bar{x}')v_s^k(\bar{x}')\right]_\Lambda + \left[\frac{\partial^{k'}\phi_s(\bar{x}')\partial_{k'}\phi_s(x')}{8\pi G}\right]_\Lambda\right). \tag{22.36}$$

Hence, we see that the small scale fluctuations induce an effective pressure perturbation on the long-wavelength fluid. One can also see the effect of the small scale velocity fluctuations by taking the first term in Eq. (22.33) and writing it as

$$\frac{1}{\rho_l}\left[\tau^{ij}\right]_\Lambda \sim \frac{1}{\rho_l}\left[\rho(\bar{x}')v_s^i(\bar{x}')v_s^j(\bar{x}')\right]_\Lambda \tag{22.37}$$

$$\sim \delta_l\left[v_s^i(\bar{x}')v_s^j(\bar{x}')\right]_\Lambda \tag{22.38}$$

$$\sim \delta_l c_s^2\delta^{ij} + O\left(\partial_k v_s\right). \tag{22.39}$$

The parameter c_s^2 is the sound speed squared due to the residual pressure of the small scales. The effective stress tensor that we have identified is thus explicitly dependent on the short wavelength fluctuations. These are very large, strongly coupled, and therefore impossible to treat within the effective theory. The next key step in the EFT description is the expansion of this stress tensor in terms of powers of derivatives and δ_l with the expansion coefficients (such as c_s^2) parameterized instead of being computed.

Since we treat the matter as a collisionless and pressureless fluid, it is convenient to introduce the notation of fluid dynamics to understand the various terms that arise from the smoothing procedure, such as the induced stress-tensor as given in Eqs. (22.33) and (22.34). To this end, consider the Naiver-Stokes equation for a fluid velocity $\bar{u}$

$$\rho\left(\frac{\partial\bar{u}}{\partial t} + \bar{u}.\nabla\bar{u}\right) + \rho\nabla\phi = -\nabla p + \nabla\cdot\left\{\eta\left[\nabla\bar{u} + (\nabla\bar{u})^{\rm T} - \frac{2}{3}(\nabla\cdot\bar{u})\,{\rm I}\right] + \zeta\,(\nabla\cdot\bar{u})\,{\rm I}\right\} \tag{22.40}$$

where the coefficients ζ and η are the bulk and shear viscosity. Similarly, we can re-frame the smoothed stress tensor τ by expanding the small-scale modes around their expectation value with a perturbation that is modulated by long-wavelength modes. Hence we write

$$\left[\tau^{ij}\right]_\Lambda = \delta^{ij}\left(p_b + c_s^2\delta\rho_l - c_{bv}^2\frac{\rho_b}{aH}\partial_k v_l^k\right) - \frac{3\rho_b c_{sv}^2}{4aH}\left(\partial^j v_l^i + \partial^i v_l^j - \frac{2}{3}\delta^{ij}\partial_k v_l^k\right) + \Delta\tau + \cdots \tag{22.41}$$

where the parameters c_{bv} and c_{sv} are the coefficients related to the bulk and shear viscosity respectively of the effective fluid. $\Delta\tau$ is the stochastic term (due to small scale fluctuations) uncorrelated with the smoothed field and $\cdots$ represents terms higher order in derivative and power counting in δ_l. **The various coefficients $c_s^2, c_{sv}^2, c_{bv}^2$ encapsulate the backreaction of 'ultraviolet (UV) physics' of the Universe, i.e. that operating on scales beyond our cutoff Λ, on large scale effective fluid**. This seemingly simple addition from the EFTofLSS over SPT is the most significant difference between the two PT formalisms. The free parameters within our new theory aka EFTofLSS are obtained by fitting to the observed data or simulations. This way, EFTofLSS captures the backreaction of small scales on large scales without making any assumptions about small-scale physics. Some of the most complex 'baryonic' effects can also be treated in this way, while remaining completely agnostic about their intricate physics (see 1412.5049 and 2010.02929).

22.3 EFTofLSS solution and renormalization

Having derived the relevant 'smoothed' EoMs for the effective fluid, we solve these using the same perturbative approach implemented in the SPT formalism. Note that the only difference between the SPT and EFT equations is the additional induced stress tensor term in the EFT description. Hence, we write the final solution for the nonlinear matter overdensity field δ_l as

$$\delta_l = \delta_l^{(1)} + \delta_l^{(2)} + \delta_l^{(3)} + \delta_l^{(c)} + ... + \Delta\tau \tag{22.42}$$

where at the one-loop order the only relevant new term is $\delta_l^{(c)}$ which is explicitly given as

$$\delta_l^{(c)} = c^2\nabla^2\delta_l \tag{22.43}$$

with $c^2(\Lambda) = c_s^2(\Lambda) + f\left(c_{sv}^2(\Lambda) + c_{bv}^2(\Lambda)\right)$ as given in Eq. (22.41) and where we have made the Λ dependence of these free parameters explicit. Here, f is the logarithmic growth rate given as $f = d\ln D/d\ln a$.

Finally, the Fourier space matter power spectrum up to one-loop[2] is given as

$$P_\Lambda^{\rm EFT}(k,z) = D^2 P_\Lambda^{(11)}(k) + D^4\left[P_\Lambda^{(13)}(k) + P_\Lambda^{(22)}(k)\right] + D^2 P_\Lambda^{\rm ctr}(k,z) \tag{22.44}$$

where $P_\Lambda^{\rm ctr}$ is referred as the 'counterterm' contribution and $D \equiv D(z)$ is the normalized growth function. The counterterm contribution is expressed as

$$P_\Lambda^{\rm ctr}(k,z) = -c_\Lambda^2(z)k^2 P_\Lambda^{(11)}(k). \tag{22.45}$$

At this order, there are two key differences between the SPT and EFTofLSS predictions: (1) the loop integrals extend only to Λ, since we have smoothed the fields, and (2) the appearance of the final term involving the effective sound-speed c_Λ^2.

[2]We have neglected the contribution from the stochastic term $\Delta\tau$ which will remain sub-dominant for the cosmologies of our interest.

22.3.1 Renormalization

The EFT power spectrum as given above appears to be Λ-dependent due to the inherent dependence of the long-wavelength field δ_l on the smoothing scale Λ. However, we will show that the additional Λ-dependent term $P_\Lambda^{\rm ctr}$ is precisely what we need to make the entire one-loop spectrum approximately Λ-independent. To this end, consider the linear power spectrum $P_\Lambda^{(11)}(k)$:

$$P_\Lambda^{(11)}(k \ll \Lambda) = \left\langle \delta^{(1)}_{k,\Lambda} \delta^{(1)}_{p,\Lambda} \right\rangle' \tag{22.46}$$

$$= \left\langle W_\Lambda(k) \delta^{(1)}_k W_\Lambda(p) \delta^{(1)}_p \right\rangle' \tag{22.47}$$

$$= W_\Lambda^2(k) P^{(11)}(k) \tag{22.48}$$

$$\approx P^{(11)}(k) \tag{22.49}$$

where we used $\delta_{k,\Lambda} \equiv W_\Lambda(k)\delta_k$ and in the last line we approximated the smoothing kernel $W_\Lambda(k) \approx 1$ for $k \ll \Lambda$. Hence, the linear power spectrum is Λ-independent for all scales of interest that are much larger than the smoothing scale. Now, let us consider the $P_\Lambda^{(13)}(k)$ term:

$$P_\Lambda^{(13)}(k) = 6 P_\Lambda^{(11)}(k) \int \frac{d^3q}{(2\pi)^3} F_3(\vec{k}, \vec{q}, -\vec{q}) P_\Lambda^{(11)}(q) \tag{22.50}$$

whose UV limit is given as

$$\lim_{k\to 0} P_\Lambda^{(13)}(k) = -\frac{61}{630\pi^2} k^2 P_\Lambda^{(11)}(k) \int_0^\infty dq P_\Lambda^{(11)}(q) \tag{22.51}$$

$$= -\frac{61}{630\pi^2} k^2 W_\Lambda^2(k) P^{(11)}(k) \int_0^\infty dq W_\Lambda^2(q) P^{(11)}(q) \tag{22.52}$$

$$\approx -\frac{61}{630\pi^2} k^2 P^{(11)}(k) \int_0^\Lambda dq P^{(11)}(q) \tag{22.53}$$

$$= k^2 P_\Lambda^{(11)}(k) f(\Lambda). \tag{22.54}$$

Hence, we find that the $P_\Lambda^{(13)}$ term has an explicit Λ-dependence due to the smoothing procedure. This Λ-dependence is similar to the one we derived for a corresponding $P^{(13)}$ term in SPT and hence leads to similar problems since the complete one loop power spectrum must be inherently Λ-independent. However, unlike SPT, EFTofLSS contains an additional term at one loop order, the counterterm contribution. This contribution has the exact spectral shape $P_\Lambda^{\rm ctr}(k,z) = -c_\Lambda^2(z) k^2 P_\Lambda^{(11)}(k)$ to cancel the apparent Λ-dependence of $P_\Lambda^{(13)}$. To see this, consider the sum of $P_\Lambda^{(13)}$and $P_\Lambda^{\rm ctr}$ for all scales $k \ll \Lambda$ and we will use the approximation that $P_\Lambda^{(11)}(k) \approx P^{(11)}(k)$ for all Λ such that $k \ll \Lambda$. Hence,

$$D^2 P_{\Lambda_2}^{(13)}(k) + P_{\Lambda_2}^{\rm ctr}(k,z) = D^2 P_{\Lambda_1}^{(13)}(k) + D^2 k^2 P_{\Lambda_1}^{(11)}(k) \left[f(\Lambda_2) - f(\Lambda_1) \right] - c_{\Lambda_2}^2(z) k^2 P_{\Lambda_2}^{(11)}(k) \tag{22.55}$$

$$= D^2 P_{\Lambda_1}^{(13)}(k) - k^2 P^{(11)}(k) \left[c_{\Lambda_2}^2(z) - D^2 f(\Lambda_2) + D^2 f(\Lambda_1) \right] \tag{22.56}$$

$$= D^2 P_{\Lambda_1}^{(13)}(k) - c_{\Lambda_1}^2(z) k^2 P_{\Lambda_1}^{(11)}(k) \tag{22.57}$$

Therefore, we find that the counterterm in EFTofLSS 'renormalizes' the $P^{(13)}$ one-loop term such that the apparent Λ-dependence vanishes. Hence, we observe that the microphysical $c^2(z)$ changes as we vary Λ and the variation of c^2 occurs in precisely the manner to cancel any change in $P^{(13)}$ term. For this reason, c^2 is also known as 'ultraviolet counterterm'. In other words, although the individual loop integrals and counterterms are Λ-dependent, their sum isn't: therefore as desired the overall theory is independent of any cutoff scale Λ.

So far we have only considered the $P^{(13)}$ loop term. However, the above argument can be applied to any loop term. Specifically, we note that the apparent Λ-dependence of $P^{(22)}$ term scales as k^4. This k^4 dependence is exactly canceled or absorbed by the lowest order stochastic term $\Delta\tau$ in our EFT expansion. However, since $k \ll \Lambda$, the k^4 dependence is sub-dominant compared to $k^2 P^{(11)}(k)$ for our scales of interest. Therefore, the Λ-dependence of $P^{(22)}$ term is usually neglected along with any contribution from $\langle \Delta\tau_k \Delta\tau_p \rangle'$.

Based on the above discussion, we write the full EFT power spectra at one loop order as first derived in 1206.2926:

$$P^{\rm EFT}(k,z) = D^2(z)P^{(11)}(k) + D^4(z)P^{(22)}(k) + D^4(z)P^{(13)}_\Lambda(k) - D^2(z)c^2_\Lambda(z)k^2 P^{(11)}_\Lambda(k). \quad (22.58)$$

where we remind the reader that the LHS is Λ-independent even though the individual terms $P^{(13)}_\Lambda(k)$ and $P^{\rm ctr}_\Lambda$ can vary with Λ. Note that $c^2 > 0$ implies a positive residual pressure and hence the power reduces on quasi-linear scales. However, note that c^2 is a coefficient of an EFT operator consistent with symmetries and power counting, and we did not make assumptions of the positivity of this coefficient

22.3.2 Physical implication of counterterm c^2

Until now, our exploration of the counterterm within the Effective Field Theory of Large-Scale Structure (EFTofLSS) has predominantly centered on its effectiveness in removing Λ dependence and guaranteeing the renormalizability of loop terms. The parameter c^2 associated with the counterterm serves as a free parameter representing the effective sound speed squared of the effective fluid, acquired through a smoothing process. Although our initial set of fluid equations involved a pressureless fluid, the presence of residual pressure on large scales, stemming from gravitational clustering at smaller scales, was encapsulated in the residual stress tensor term. This residual pressure, parameterized by c^2, becomes measurable through N-Body simulation data. Consequently, the inferred value of the counterterm parameter c^2 obtained from fitting N-Body data holds pertinent insights into the impact of small-scale feedback on large scales. Analogously, the viscosity of a fluid cannot be deduced solely from an effective low-energy fluid description but is experimentally measured before being incorporated into the fluid equations, such as the Navier-Stokes equation, for predictive purposes. The success of EFT over other Perturbation Theory (PT) formalisms largely stems from the incorporation of such free parameters within the theory. These parameters not only serve to renormalize loop integrals but also furnish additional information about small-scale clustering and its influence on large scales. Hence, one can factorize c^2_Λ as

$$c^2_\Lambda(z) = \tilde{c}^2_\Lambda(z) + c^2_{\rm phy}(z) \quad (22.59)$$

where $\tilde{c}^2_\Lambda$ is the Λ-dependent term that absorbs the cutoff dependence of $P^{(13)}$ term, and $c^2_{\rm phy}$ is a 'physical' cutoff independent term that contains nontrivial information regarding the UV effects

of small scales on large scale modes. Hence, we can rewrite the full one-loop EFT matter power spectrum as

$$P^{\rm EFT}(k,z) = D^2(z)P^{(11)}(k) + D^4(z)P^{(22)}(k) + \left(D^4(z)P^{(13)}_\Lambda(k) - 2D^2(z)\tilde{c}^2_\Lambda(z)k^2P^{(11)}\right) - 2D^2(z)c^2_{\rm phy}(z)k^2P^{(11)}. \tag{22.60}$$

Since the $\tilde{c}^2_\Lambda(z)$ term must cancel the Λ-dependence of $P^{(13)}_\Lambda$ at all redshifts, it should vary with redshift exactly like $D^2(z)$. Hence,

$$P^{\rm EFT}(k,z) = D^2(z)P^{(11)}(k) + D^4(z)P^{(22)}(k) + D^4(z)\left(P^{(13)}_\Lambda(k) - 2\tilde{c}^2_\Lambda(0)k^2P^{(11)}\right) - 2D^2(z)c^2_{\rm phy}(z)k^2P^{(11)} \tag{22.61}$$

where $\tilde{c}^2_\Lambda(0)$ is the value of the counterterm at $z = 0$. Note that $c^2_{\rm phy}(z)$ can have an arbitrary redshift dependence, contingent upon the evolution of the residual pressure induced by gravitational clustering. More importantly, we note that the only IR-surviving quantity inherited from the UV effects is the renormalized parameter $c^2_{\rm phy}(z)$. When analyzing cosmologies with different initial conditions and cosmological parameters, a comparison between $c^2_{\rm phy}(z)$ can act as an additional distinguishing feature. For instance, refer to Fig. 3 in 2306.09456 where the authors show the variation of $c^2_{\rm phy}(z)$ as a function of a cosmological parameter that alters the small-scale power. From the structure of the counterterm contribution, we expect the $c^2_{\rm phy}(z) \sim 1/k^2_{\rm NL}(z)$. This gives a value of $c^2_{\rm phy}(z = 0) \approx O(10)$ for $k_{\rm NL}(z = 0) \approx 0.3$ using the renormalization scheme mentioned in 2306.09456. Note that this value differs from the usual $O(1)$ value typically quoted for the bare $c^2(z)$.

22.4 EFTofLSS matter power spectrum result

In Fig. 23 we compare the matter power spectra obtained from SPT and EFT formalism against N-Body data. We observe that the EFT curves perform far better than SPT in matching the N-Body data points. Here, we note that at the one-loop order in EFT, we have only one counterterm, c^2, whereas at the two-loop order we require 3 counterterms. Despite the increase in the number of free parameters (counterterms), the EFT curve at the two-loop order performs better than one loop, and matches with the simulation data point up to $k \approx 0.2$ h/Mpc. This is also shown in Fig. 24 where we show yet another comparison of the EFT curves with their SPT counterparts along with relevant cosmic and theory errors. The shaded blue region enveloping the two loop EFT curve in Fig. 24 indicates our estimate of the theoretical error due to the higher order loop terms. The dashed curve represents an estimate of the cosmic variance error $\propto k^{-3}$ which diminishes at small scales.

22.5 Application to iso-curvature perturbations

Through the above two plots, we have highlighted that the EFT formalism remedies the issues in SPT namely integration over non-perturbative small-scales in loop terms and apparent Λ-dependence. These are ameliorated by introducing a free counterterm parameter. The SPT also suffers from not being applicable to any arbitrary or general set of cosmological initial conditions. As discussed previously, if we consider a power law form for the linear power spectrum

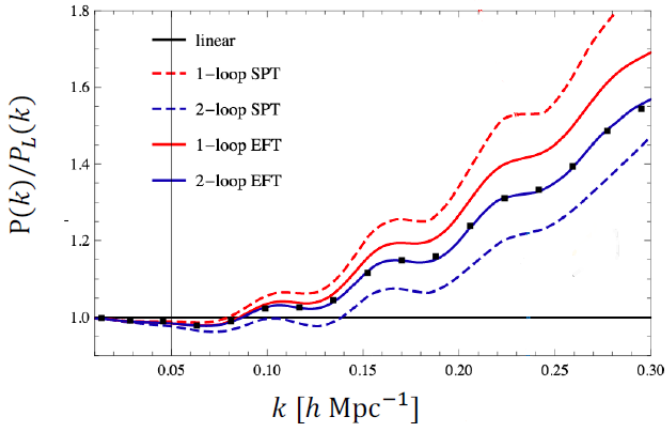


Figure 23. Plot showing comparison of matter power spectrum as obtained from SPT and EFTofLSS against data from N-Body simulations. Each curve has been divided by the linear power spectrum. The fully nonlinear N-Body power spectrum is plotted in black boxes. The red and blue dashed curves show one and two-loop results from SPT whereas similar order curves from EFTofLSS are shown in solid colors. The above figure is taken from O. Philcox's presentation, 2020.

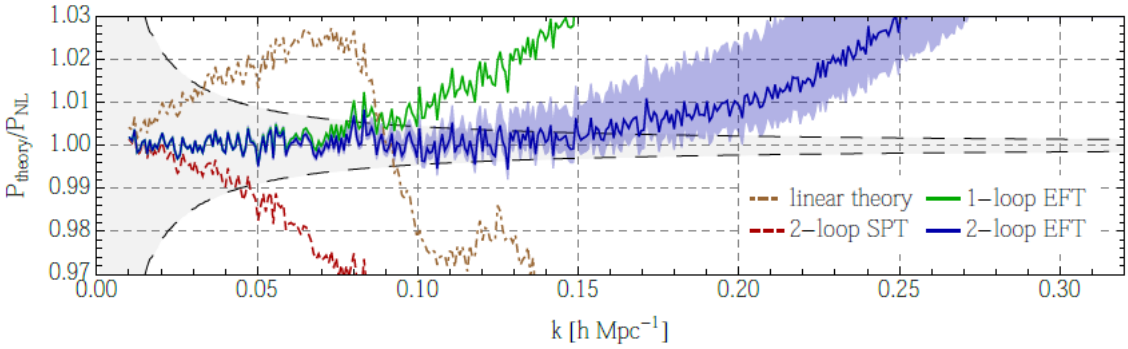


Figure 24. Similar to Fig. 23. Here, the curves are divided by the nonlinear (NL) matter power spectrum as obtained from NBody simulations. Taken from 1507.05326.

$P^{(11)}(k) \propto k^n$, then some of the loop terms diverge for $n > -1$. Such conditions can arise naturally if we consider mixed primordial initial conditions consisting of adiabatic fluctuations with a small fraction of CDM blue-tilted isocurvature power as shown in Fig. 25. This is an example from our own research work published in 2306.09456. We briefly mentioned isocurvature in Sec. 6.6.3.

For such cosmologies, one is often forced to choose a particularly small value of Λ to avoid large spurious contributions from small scales. In the EFTofLSS, however, there are no such divergences. This occurs since the domain of integration is bounded; since the internal momenta q are limited by Λ and the integrand is analytic, the loop integrals are guaranteed to be finite. If there are divergences lurking in the high-q regime, they are themselves absorbed within the counterterms. In Fig. 26 we plot EFT curves for the pure adiabatic and mixed cosmologies. For both of the cases we choose an arbitrarily large value of the cutoff scale $\Lambda \approx 100$ h/Mpc. For

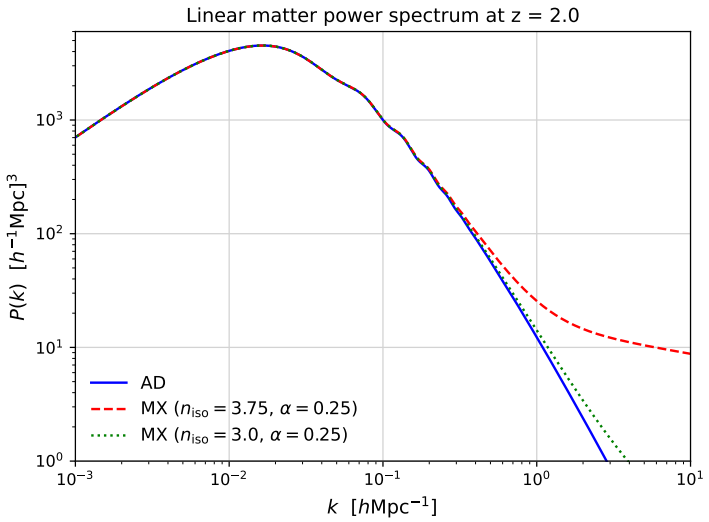


Figure 25. Plot showing a comparison between the linear matter power spectra at a redshift of $z = 2$ for the pure adiabatic and mixed initial conditions. For the mixed case, we show two examples in which the power deviates from the adiabatic scenario on small scales with the spectral indices $n = -0.25$ (dashed) and $n = -1$ (dotted) respectively. Taken from 2306.09456.

the pure adiabatic case, the counterterm tends to an asymptotic value as $\Lambda \to \infty$ and is an $O(1)$ number as shown in the figure. However, due to the diverging structure of the $P_\Lambda^{(13)}$ term for the mixed scenario with $\lim_{k\to\infty} P^{(11)}(k) \propto k^{-0.25}$, the counterterm runs with the variation in Λ. For our choice of $\Lambda = 100$ h/Mpc, we find that $c_\Lambda^2(z = 1) \approx -6.23$ (Mpc/h)^2. While a negative value of the c^2 compared to adiabatic may seem alarming, note that the only IR-surviving quantity inherited from the UV effects is the physically relevant parameter $c^2_{\rm phy}$. For the pure adiabatic and mixed case, the physical parameter $c^2_{\rm phy}$ are nearly identical in magnitude. A small difference in their magnitude is due to a larger power on smaller scales within the mixed scenario. On the other hand, given that the bare c_Λ^2 can become negative for $\Lambda \gtrsim O(3)$, there is an unclear interpretation of this parameter, which leaves room for more intricate UV dynamics being at play here such as for the mixed (isocurvature) scenario. A nonlinear UV model exploration of this issue may be useful to further elucidate the difference.

22.6 From dark matter to galaxies: The bias expansion

Up to this point, we have considered only the statistics of dark matter fluctuations in the Universe. In practice, most observational probes measure either the integrated mass distribution (weak lensing) or the galaxy distribution (photometric or spectroscopic surveys). For this reason, we will now briefly discuss biasing, and the associated calculation of galaxy power spectra and discuss EFTofLSS approach to biasing.

In most nonlinear biasing models (see 1611.09787 for a review of galaxy bias), we use the one-loop galaxy power spectrum obtained through a bias expansion for the galaxy density contrast by including all operators allowed by Galilean symmetry up to cubic order in magnitude of the

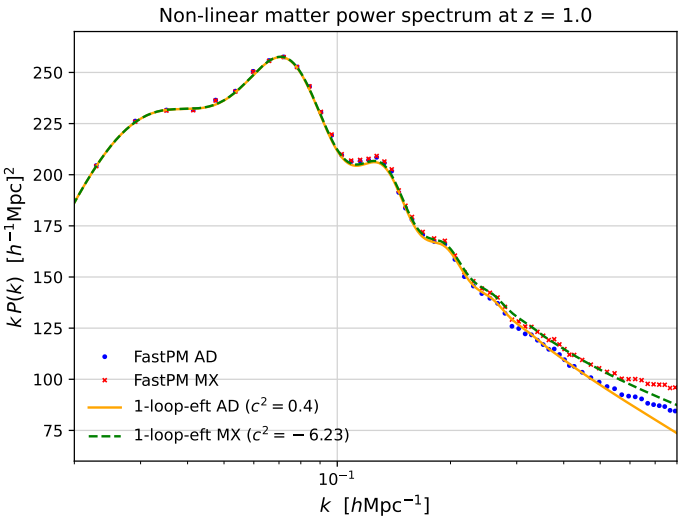


Figure 26. In this figure we highlight the fitting of one-loop EFT power spectrum to the N-body data. Note that we plot scaled power spectrum, $k \times P(k)$, on the y-axis for clarity. For the mixed case, we use fiducial value as $n = -0.25$. The value of the bare c^2_Λ (at cutoff $\Lambda = 100$ h/Mpc) one-loop EFT parameter is given in the label for the EFT curves. Note that the value of c^2_Λ for the mixed case is negative. The one-loop EFT curves is accurate up to ≈ 0.5 h/Mpc at redshift $z = 1$. We also plot the approximate theoretical error band expected from two-loop contributions.

coarse-grained linear overdensity $\delta^{(1)}_l$,

$$\delta_g(x) = \sum_{\mathcal{O}} \left(b_{\mathcal{O}} + \epsilon_{\mathcal{O}}(x)\right) \mathcal{O}(x) + b_\epsilon \epsilon(x) \tag{22.62}$$

$$\begin{aligned} &= b_1\delta(x) + b_\epsilon\epsilon(x) \\ &+ \frac{b_2}{2}\delta^2(x) + b_{\mathcal{G}_2}\mathcal{G}_2(x) + \epsilon_\delta(x)\delta(x) \\ &+ b_{\delta\mathcal{G}_2}\delta(x)\mathcal{G}_2(x) + \frac{b_3}{6}\delta^3(x) + b_{\mathcal{G}_3}\mathcal{G}_3(x) + b_{\Gamma_3}\Gamma_3(x) + \epsilon_{\delta^2}(x)\delta^2(x) + \epsilon_{\mathcal{G}_2}(x)\mathcal{G}_2(x) \\ &+ b_{\nabla^2\delta}\nabla^2\delta(x) + b_{\nabla^2\epsilon}\nabla^2\epsilon(x) \end{aligned} \tag{22.63}$$

where all the operators $\mathcal{O}$ in the above expression are considered to be coarse-grained and the subscripts l or Λ have been dropped for brevity. In Fourier space the Laplacian takes the form $\nabla^2 \to (k/k_*)^2$ where k_* is some characteristic scale of clustering for biased tracers and we restrict to scales $k/k_* \ll 1$. Hence every insertion of a Laplacian is equivalent to a second order correction to an operator $\mathcal{O}$ and the derivative operators in the last-line of Eq. (22.63) are counted approximately as cubic order in bias expansion. Therefore, Eq. (22.63) is a double expansion in density fluctuations and their derivatives. The remaining operator set $\{\delta^2, \mathcal{G}_2, \epsilon_\delta\delta\}$ and $\{\mathcal{G}_2\delta, \delta^3, \mathcal{G}_3, \Gamma_3, \epsilon_{\delta^2}\delta^2, \epsilon_{\mathcal{G}_2}\mathcal{G}_2, \nabla^2\delta(x), \nabla^2\epsilon(x)\}$ are second and third order respectively and we refer the readers 1611.09787 for definition and details regarding these operators. Notably, the operators non-local in δ such as $\mathcal{G}_2 = (\nabla_i\nabla_j\Phi)^2 - \left(\nabla^2\Phi\right)^2$ arise naturally due to gravitational evolution and renormalization requirements respectively. This was first shown in 1402.5916.

22.7 Application of the EFTofLSS to simulations and real data

Finally, within the EFTofLSS, we model the perturbative galaxy-galaxy power spectrum P_{gg} at one loop level as sum of the deterministic, stochastic and counterterm parts:

$$P_{gg} = P_{gg}^{\rm det} + P_{gg}^{\rm sto} + P_{gg}^{\rm ctr}. \tag{22.64}$$

During cosmological parameter inference from simulation or observational data, we fit the aforementioned theoretical power spectrum with the relevant number of bias and counterterm parameters. The theoretical spectrum can be obtained from a few existing codes such as CLASS-PT (2004.10607) for EPT, and Velocileptor (2012.04636) for a more complicated LPT implementation. CLASS-PT is an adaptation of the CLASS code designed to compute the non-linear power spectra of dark matter and biased tracers using one-loop cosmological perturbation theory in Eulerian coordinates. It handles both Gaussian and non-Gaussian initial conditions. It's an easy-to-use and convenient code when performing LSS analysis. Now consider the simplest case where we fit the one loop spectrum to a data in real space. In this case there exists only one free parameter c^2 which is often absorbed into the Laplacian bias coefficient. In redshift space, discussed in Sec. 20.6, we often consider only the first 3 multipoles $\ell = 0, 2, 4$ and attach independent counterterms to each multipole spectra. In Fig. 27 we show the results of a blinded challenge that was performed and reported in 2003.08277 using EFTofLSS in redshift space for the first two multipoles.

In Fig. 28 we show the results from a recent cosmological parameter inference performed using four independent Baryonic Oscillation Spectroscopic Survey (BOSS) datasets across two redshift bins ($z_{\rm eff} = 0.38, 0.61$) in flat ΛCDM, marginalizing over 7 nuisance parameters for each dataset (28 in total) and varying 5 cosmological parameters (ω_b, ω_{cdm}, H_0, A_s, $\sum m_\nu$). The theory model includes a complete perturbation theory description that properly takes into account the non-linear effects of dark matter clustering, short-scale physics, galaxy bias, redshift-space distortions, and large-scale bulk flows. The constraints on H_0 and Ω_m as obtained from the EFT analysis of BOSS data are already competitive with the CMB measurements of Planck for the same cosmological model with varied neutrino masses. This highlights the success of EFTofLSS and setting the stage for precision cosmology from future surveys.

23 N-body simulations

The fluid approximation breaks down on small scales. For example, the velocity field is no longer single valued at a point in space, once **shell crossing** happens (i.e. clouds of mass pass through eachother).

Next to perturbation theory of fluids, the second main way to evaluate the dynamics of the universe are **N-body simulations** of (dark) matter. N-body simulations are not intrinsically perturbative, and can thus in principle extend our reach to non-perturbative scales to extract cosmological parameters with more sensitivity. On the other hand, N-body simulations are computationally costly and it is difficult to simulate the survey volume of a galaxy survey with the required resolution. In addition, dark matter N-body simulations are only valid on scales where baryonic feedback is unimportant. To go to smaller scales, one needs even more computationally

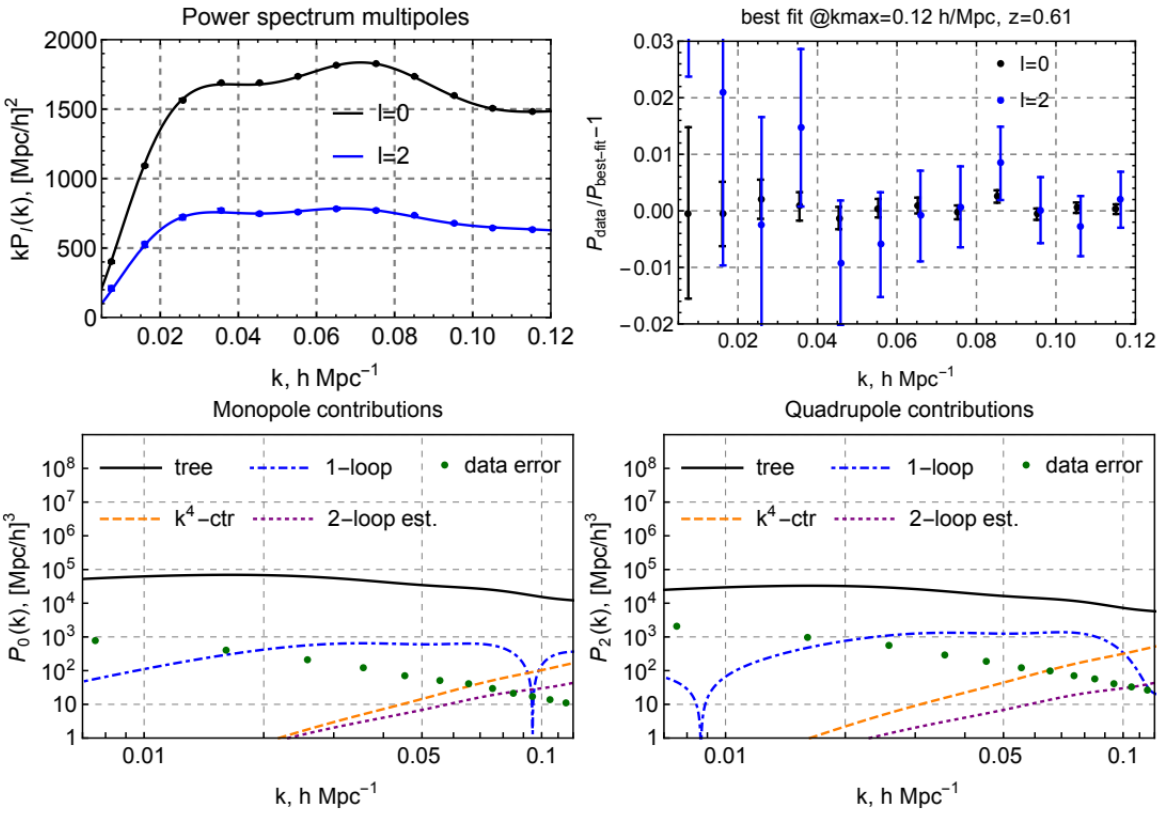


Figure 27. (Taken from 2003.08277) The upper panel shows comparison of the data for the monopole and the quadrupole with the best-fit EFT model. The residuals for the monopole and the quadrupole for the best-fit model (right panel). Note that the quadrupole data points are slightly shifted for better visibility. In the lower panel we show different contributions to the monopole (left panel) and quadrupole (right panel) power spectra. The data errors and the two-loop estimate are also displayed. We plot the absolute values, some terms are negative. Here, k^4-ctr is the contribution due to the Finger-of-God effect.

expensive **(magneto-)hydrodynamic simulations**. Improving simulations is a very active field of research. Using modern autodifferntiation techniques, imported from machine learning, there are even **differentiable simulations**, which we will briefly get back to in Sec. 28.3.

A good review on N-body simulations, though not very recent, is https://ned.ipac.caltech.edu/level5/March03/Bertschinger/paper.pdf. A nice and more recent review is https://arxiv.org/abs/2112.05165. This section is based in part on Dodelson-Schmidt. There are many large sets of simulations that you can download, such as Quijote (1909.05273), CAMELS (2010.00619), IllustrisTNG (1812.05609) and Abacus (2110.11398). These require millions of CPU hours, and sometimes are used in hundreds of publications. For most research projects you will not need to run your own simulations.

23.1 Equations for particles

N-body simulations are typically performed in a cubic volume with periodic boundary conditions, so that particles exiting the volume on one side re-enter on the other side. We discretize the

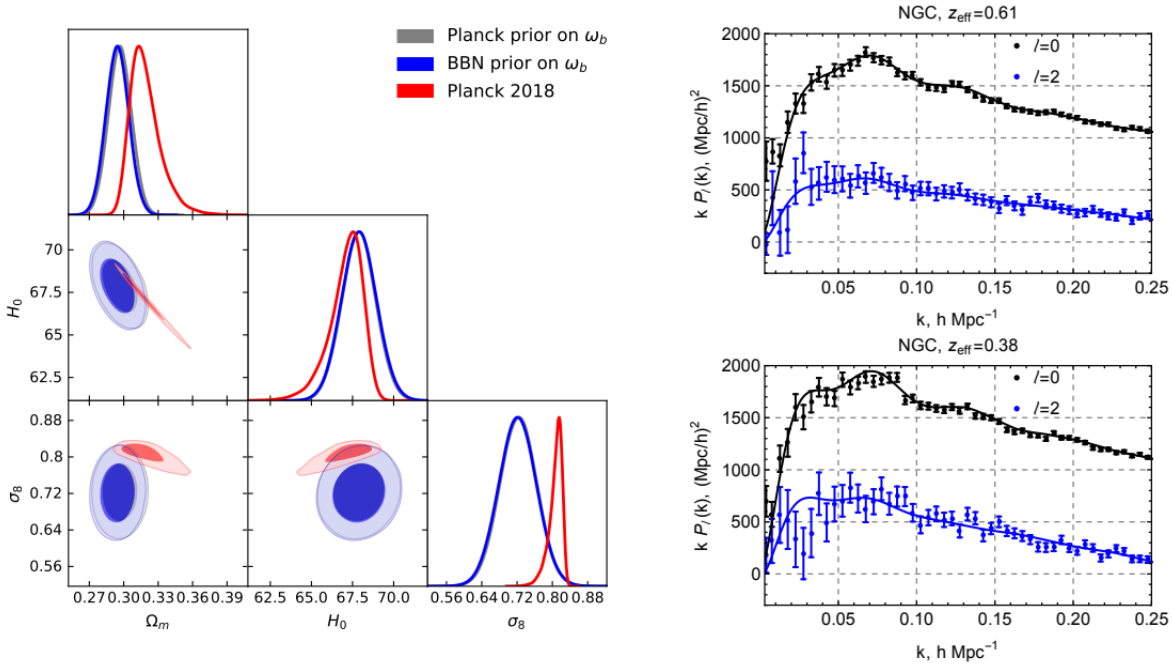


Figure 28. (Taken from 1909.05277) Left panel: The posterior distribution for the late-Universe parameters H_0, Ω_m and σ_8 obtained with priors on ω_b from Planck (gray contours) and BBN (blue contours). For comparison we also show the Planck 2018 posterior (red contours) for the same model (flat ΛCDM with massive neutrinos). Right panel: The monopole (black dots) and quadrupole (blue dots) power spectra moments of the BOSS data for high-z (upper panel) and low-z (lower panel) north galactic cap (NGC) samples, along with the best-fit theoretical model curves. The corresponding best-fit theoretical spectra are plotted in solid black and blue.

dynamics by tracking $N = N_{side}^3$ particles from their initial (almost uniform) position to their late time positions as a function of time. The equations of motion are simply Newtonian gravity in an expanding space time:

$$\frac{dx^i}{dt} = \frac{p^i}{ma} \tag{23.1}$$

$$\frac{dp^i}{dt} = -Hp^i - \frac{m}{a}\frac{\partial\phi}{\partial x^i} \tag{23.2}$$

Introducing the **superconformal momentum** $\mathbf{p}_c = a\mathbf{p}$, which is conserved in the absence of perturbations, this can be re-written as

$$\frac{dx^i}{dt} = \frac{p_c^i}{ma^2} \tag{23.3}$$

$$\frac{dp_c^i}{dt} = -m\frac{\partial\phi}{\partial x^i} \tag{23.4}$$

Solving these equations numerically, for a large number of particles such as 1000^3, leads to a beautiful and physically accurate matter distribution.

A computationally efficient and widely used method to solve these equations is the **leapfrog scheme**, where density and velocity are evaluated with an offset of half a time step:

$$\mathbf{x}^{(i)}(t) \quad \text{and} \quad \mathbf{p}_c^{(i)}(t - \Delta t/2) \tag{23.5}$$

After generating the initial conditions (usually using LPT as discussed in Sec. 21.3), the algorithm proceeds as follows:

1. Compute the gravitational potential generated by the collection of particles, and take its gradient to obtain $\nabla\phi(\mathbf{x}, t)$.

2. Change each particle's momentum ("kick") by

$$\mathbf{p}_c^{(i)}(t + \Delta t/2) = \mathbf{p}_c^{(i)}(t - \Delta t/2) - m\nabla\phi^{(i)}(\mathbf{x}, t)\Delta t. \tag{23.6}$$

3. Move each particle position ("drift") by

$$\mathbf{x}^{(i)}(t + \Delta t) = \mathbf{x}^{(i)}(t) + \frac{\mathbf{p}_c^{(i)}(t + \Delta t/2)}{ma^2(t + \Delta t/2)}\Delta t. \tag{23.7}$$

4. Repeat.

In general, the more time steps, the more accurate the results. There are ways to optimize the time steps.

From the particles, one usually proceeds to evaluate the matter density on a regular grid using a **mass assignment scheme (MAS)**. A particle does not only contribute to the mass on the nearest grid point, but can contribute to the surrounding nodes (usually 8 in 3d). The most commonly used way to do this is the **Cloud-In-Cell (CIC)** scheme.

23.2 Evaluating the potential

The computational bottleneck is to evaluate the gravitational potential efficiently. In principle, calculating the gravitational potential for every particle requires of order N^2 operations where N is the number of particles. This is not computationally tractable. Instead one uses a **particle mesh (PM)** algorithm, where densities are interpolated to a regular grid (using the MAS), and one can then solve the Poisson equation in Fourier space using a 3d FFT.

A problem with the PM method is that it does not scale well at very high resolution, i.e. one would need a very high resolution grid to take into account local pairwise interactions between nearby particles. On these small local scales, one thus uses a different method called the **tree algorithm**. The tree algorithm generates a hierarchical tree of meta-particles. Seen from far enough away, a collection of nearby particles can be replaced by a single meta particle which combines their mass. For better accuracy one can also carry along multipoles of the mass distribution in the meta particles. The tree algorithm has the complexity $N \log N$. Modern code usually combine the PM method on large scales with the tree method on small scales. Perhaps the most widely used code is GADGET (version 2 to 4).

One additional subtle point is that, because we sample the density with a finite number of points, if points come too close there would be in principle infinite attraction between them, as an artifact of the point sampling. To avoid this one smooths the density field using a **force softening kernel**.

23.3 Baryonic simulations

To take into account baryonic forces, one uses magneto-hydrodynamic (MHD) simulations. These can be implemented using **Smoothed-particle hydrodynamics (SPH)** simulations (i.e. still using particles, but with additional forces), or with a **(moving) mesh**. Unfortunately it is not possible to simulate these forces from first principles (e.g. how an AGN blows out gas), so one needs to approximate them with a so-called **subgrid model**. There are different subgrid models that lead to different answers. For example, in the CAMELS simulations, the same initial conditions but different subgrid models can change the galaxy density by 30% or so. So while dark matter simulations, given enough resolution, are in principle arbitrarily accurate, the same is not true once we include baryonic physics. This is a key difficulty in simulation-based inference on small scales.

24 Halos and Galaxies

Perturbation theory is only valid if $\delta \ll 1$, which is only true on the largest scales over the entire history of the universe. Perturbation theory can never describe the formation of galaxies or galaxy clusters, which form when matter collapses to a small region in space with $\delta \gg 1$. Fortunately there are nevertheless analytic methods that help us to understand this domain. The methods we briefly discuss now are used in practice in cosmology to analyze data and forecast experiments.

Halo formation is a rich subject that includes concepts like the subhalos, merger trees and the halo occupation distribution, and we can only give a brief outline. Halo/galaxy formation is sensitive to cosmological parameters, astrophysical parameters and properties of dark matter. In addition, by patching together halo statistics and perturbation theory of the matter field in the so-called **halo model**, one can arrive at a theoretical description that covers all scales of cosmology. The halo model allows us to model observables and forecast measurements on very non-linear scales. Its predictions agree with simulations and data rather well.

24.1 Halos and Halo mass profile

Halos are structures of (dark) matter that are gravitationally bound, which formed by gravitational collapse of over-densities. Such over-densities will ultimately virialize. Galaxies are hosted inside much larger **dark matter halos**. There are different ways to precisely define halos, given a dark matter distribution. The most widely used method is the **friends-of-friends** algorithm and its refinement called **ROCKSTAR algorithm**. These algorithms are also called **halo finders**. The result is a **halo catalogue** with various masses, and various other properties, such as center-of-mass position and velocity. Every particle is assumed to be inside only one halo.

It turns out that to good approximation the spherically averaged mass density profile of a dark matter halo is described by the **Navarro-Frenk-White (NFW)** profile, given by

$$\rho(r|m,z) = \frac{\rho_s}{(r/r_s)(1+r/r_s)^2} \tag{24.1}$$

It is a function of halo mass m and red shift (or time). The scale radius r_s and the density ρ_s can be expressed in terms of the halos mass. Note that this profile needs to be cut off at some radius for the integral to be finite. The NFW profile is well-known enought that there is a plot on Wikipedia (https://en.wikipedia.org/wiki/Navarro%E2%80%93Frenk%E2%80%93White_profile).

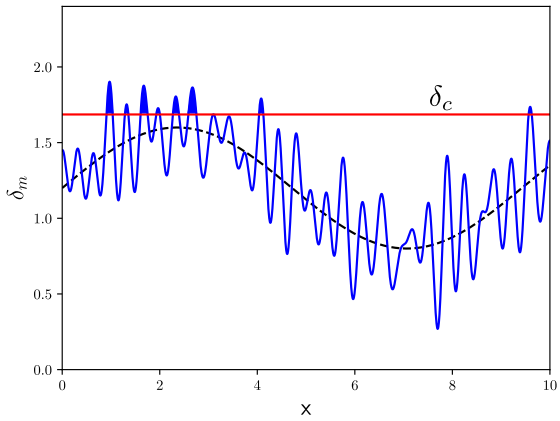


Figure 29. Halo formation. Halos form where the smoothed density field crosses the critical density. For illustration, we plot a single large-scale mode (dashed) and a few small scale modes. Figure adapted from Baumann's Cosmology book.

24.2 Halos mass function

The main statistical property of halos is their abundance. It is described by the **halo mass function** $n(m, z)$ which gives the differential number density of halos with respect to mass at a given mass m and red shift z. It is possible to calculate the halo mass function approximately using a method called the **Press-Schechter formalism**. The main ideas are the following:

- Matter perturbations on large scales are Gaussian and grow with the growth factor $D(z)$.
- We can smooth the density field on various scales R.
- In spots where the smoothed density field crosses the **critical density** δ_c, a halo will form. Because perturbations grow, new halos will form in time. It turns out that the critical density is independent of the halo mass or smoothing scale and is about $\delta_c = 1.6$, which can be derived from Newtonian gravity. This is illustrated in Fig. 29.
- Since this picture depends on the smoothing scale R, in principle smaller halos can be contained in larger ones. This is handled more carefully in the **extended Press-Schechter formalism**.

We don't have time to derive the mathematical results, but I want to show you the widely used result. The halo mass function can be expressed as

$$n(m, z) = \frac{\rho_m}{m^2} f(\sigma, z) \frac{d \ln \sigma(m, z)}{d \ln m}, \tag{24.2}$$

where ρ_m is the mean matter density. The quantity $\sigma^2(m, z)$ is the variance of mass within a sphere of radius $R(m)$ defined as

$$\sigma^2(m, z) = \frac{1}{2\pi^2} \int_0^\infty dk \; k^2 P^{\text{lin}}(k, z) W^2(kR) \tag{24.3}$$

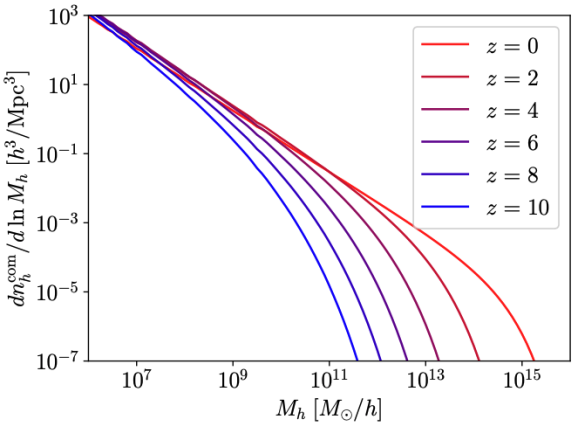


Figure 30. Sheth-Tormen mass function at different redshifts (from 2108.04279).

Here, $R = R(m)$ and the window function in Fourier space is

$$W(kR) = \frac{3\left[\sin(kR) - kR\cos(kR)\right]}{(kR)^3} \tag{24.4}$$

where m and R are related by the mean density as

$$m = 4\pi\rho_m R^3/3. \tag{24.5}$$

R can be interpreted as the radius we need to collect primordial mass from to form the halo. The term $f(\sigma, z)$ is called the **halo multiplicity** and one often assumes the **Sheth-Tormen** halo multiplicity function:

$$f(\sigma, z) = A\sqrt{\frac{2a}{\pi}}\left[1 + \left(\frac{\sigma^2}{a\delta_c^2}\right)^p\right]\frac{\delta_c}{\sigma}\exp\left[-\frac{a\delta_c^2}{2\sigma^2}\right] \tag{24.6}$$

with $A = 0.3222$, $a = 0.75$, $p = 0.3$, and $\delta_c = 1.686$. The resulting mass function is plotted in Fig. 30.

The halo mass function, as a function of cosmological parameters, can also be "learned" from simulations. This is done for example in 1804.05866, 2003.12116. By measuring the HMF from the data and comparing it to the theoretical expectation from simulations one can then in principle measure cosmological parameters. This is called cluster abundance or cluster counting. While small halos may be very sensitive to unknown baryonic physics, the largest halos are dominated by gravity and might provide reliable measurements.

24.3 Halo bias

Press-Schechter formalism can also be used to calculate the halo bias as a function of mass. This can be done using the **peak-background split** argument. The basic idea is to split perturbations into long modes (background) δ_b and short modes (peaks) δ_h as

$$\delta = \delta_h + \delta_b \tag{24.7}$$

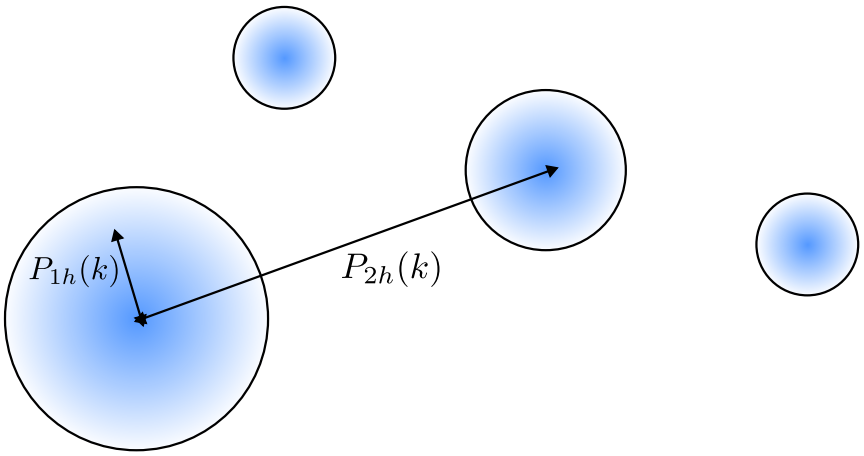


Figure 31. Example of contributions to the 1-halo and 2-halo power spectra.

The short modes will eventually form halos. The long modes can be interpreted as locally shifting the required critical density for the short modes to form halos. This is illustrated in Fig. 29 (dotted line is the long mode). By expanding the mass function to linear order in δ_b one can derive the linear halo bias. This leads to:

$$b_h(m,z) = 1 + \frac{1}{\delta_c}\frac{d\log f}{d\log \sigma} \tag{24.8}$$

Note that the halo bias satisfies a consistency relation:

$$\int_{-\infty}^{\infty} d\ln m\; mn(m,z)\left(\frac{m}{\rho_m(z)}\right) b_h(m,z) = 1, \tag{24.9}$$

i.e. the total matter field comprised of all halos is unbiased. Note that bias can be smaller than one (and even negative, for voids, which preferentially form in underdense regions). The bias of typical galaxies in a survey is larger than one.

24.4 Halo model

The halo model is the standard tool to forecast (and sometimes analyze) observables in the non-linear regime. Despite being a phenomenological description, it agrees rather well with numerical results in many cases. In the halo model, one makes the fundamental assumption that all the dark and baryonic matter is bound up in halos with varying mass and density profiles. The correlation function for density fluctuations then receives two contributions: a "two halo term" which arises from the clustering properties of distinct halos, and a "one halo term" which arises from the correlation in density between two points in the same halo. This is illustrated in Fig. 31. A review of the halo model can be found in astro-ph/0206508. This section is based on appendix A of 1810.13423.

24.4.1 Dark matter

In Fourier space, the dark matter power spectrum is given by

$$P_{mm}(k,z) = P^{1h}_{mm}(k,z) + P^{2h}_{mm}(k,z) \tag{24.10}$$

$$P^{1h}_{mm}(k,z) = \int_{-\infty}^{\infty} d\ln m\ mn(m,z)\left(\frac{m}{\rho_m}\right)^2 |u(k|m,z)|^2 \tag{24.11}$$

$$P^{2h}_{mm}(k,z) = P^{\rm lin}(k,z)\left[\int_{-\infty}^{\infty} d\ln m\ mn(m,z)\left(\frac{m}{\rho_m}\right) b_h(m,z)u(k|m,z)\right]^2 \tag{24.12}$$

In these expressions, m is the halo mass, ρ_m is the present day cosmological matter density, $n(m,z)$ is the halo mass function (i.e. the differential number density of halos with respect to mass), $u(k|m,z)$ is the normalized fourier transform of the halo profile, $P^{\rm lin}(k)$ is the linear matter power spectrum, and $b_h(m,z)$ is the linear halo bias. The one halo term is the shot noise convolved with the profile.

We need $u(k|m,z)$, the Fourier transform of the dark matter halo density profile, which for spherically symmetric profiles is defined as

$$u(k|m,z) = \int_0^{r_{\rm vir}} dr\ 4\pi r^2 \frac{\sin(kr)}{kr}\frac{\rho(r|m,z)}{m}. \tag{24.13}$$

We assume that halos are truncated at the virial radius, and have mass

$$m = \int_0^{r_{\rm vir}} dr\ 4\pi r^2 \rho(r|m,z) \tag{24.14}$$

Note that with this definition of mass, $u(k|m,z) \to 1$ as $k \to 0$. Returning to the two-halo term and using the consistency relation in Eq. (24.9), this property of $u(k|m,z)$ ensures that $P^{2h}_{mm}(k,z) \simeq P^{\rm lin}(k,z)$ in the limit where $k \to 0$, as it should.

24.4.2 Baryons, Galaxies and other observables

It is straight forward to **generalize the halo model to other fields than dark matter**. For example, the baryonic gas distribution in the halo model is modelled by assuming gas is bound within dark matter halos, having density profiles $\rho_{\rm gas}(m,z)$ which we assume to be a function of the host halo mass and redshift only. The gas power spectrum is given by Eq. 24.10 with $u(k|m,z)$ calculated through Eq. 24.13 by replacing $\rho(m,z)$ with $\rho_{\rm gas}(m,z)$. The halo model can thus be used to calculate small scale observables such as kSZ, tSZ and gravitational lensing, as well as their cross-correlation.

There are also extensions of the halo model that can be used to calculate the distribution of galaxies. The complication for galaxies is that we often observe them in groups of smaller galaxies surrounding a larger galaxy. The standard treatment of galaxies in the halo model assumes that a dark matter halo is filled up with galaxies according to the **halo occupation distribution (HOD)**. This HOD often assigns a **central galaxy** to the center of the halo, and a distribution of **satellite galaxies** around them, where more massive halos have more satellite galaxies. The HOD can be calibrated with observations. The HOD is also used to **populate dark matter simulations with galaxies**.

Halo model power spectra can be calculated with various codes, such as https://github.com/borisbolliet/class_sz. The halo model can also be used to calculate **higher N-point functions such as the bispectrum**. While the halo model is powerful, remember however that the assumptions of a set of spherical halos that includes all matter is not a very realistic one.

25 Analyzing a Galaxy Survey Power Spectrum

To measure cosmological parameters from the two point function for galaxy surveys, both the correlation function in position space and the power spectrum in harmonic space are frequently used as a basis for the likelihood. Under sufficient conditions, both analyses should give the same result. The harmonic space analysis is more directly related to the perturbation theory, and this is the approach we are discussing here.

There is also a difference in the coordinate basis between photometric surveys and spectroscopic surveys. Spectroscopic galaxy samples are usually binned into a small number of 2d maps, as discussed in Sec. 20.8. The analysis then works similar to the one of the CMB (e.g. one can use the PyMaster code to take the power spectrum in the bin). On the other hand, spectroscopic surveys are done in 3d red shift space. We are focussing on the spectroscopic case here.

The goal of power spectrum analysis is of course to fit a theoretical parameterization of the power spectrum, such as the EFT model discussed in Sec 22.6, to a measurement of the power spectrum, using some likelihood. We already discussed the likelihood step in Sec. 10.3.1 for N-body data, and for real galaxy surveys it works conceptually the same.

25.1 Power spectrum estimator

As we have seen in our N-body analysis, if we could measure the universe uniformly without a mask or noise or RSD, power spectrum estimation would just be taking the Fourier transform and squaring the modes. For a real galaxy survey, the analysis is more complicated due to the mask and noise properties of a real galaxy survey.

The most widely used power spectrum estimator is the so-called **FKP estimator (Feldman-Kaiser-Peacock)**. It is decribed in the original paper astro-ph/9304022, and a modern version is discussed for example in 1505.05341, 1704.02357. The FKP estimator is computationally tractable and near optimal for current surveys. As is the case for the CMB, one can also define an optimal quadratic estimator (quasi maximum likelihood QML estimator), which is computationally more involved. A summary of power spectrum estimation, and a discussion of the optimal quadratic estimator, is given in Oliver Philcox' PhD thesis http://arks.princeton.edu/ark:/88435/dsp01v692t9422.

Let's sketch the FKP estimator. To take into account the mask of a galaxy survey, the common procedure is to generate a **random catalog (also called synthetic catalog**) of galaxy positions from the mask of the survey. The random catalog accounts for the angular mask and radial selection function of the survey. It is generated by throwing random galaxies into the survey volume in a Poisson process, which is not modulated by the cosmological power spectrum. I.e. it gives us galaxy positions as we would observe them if galaxies were unclustered. You can often download such random catalogs in addition to the true data from a galaxy survey colllaboration.

We begin by defining the weighted galaxy density field in redshift space ($\mathbf{r}$) ,

$$F(\mathbf{r}) = \frac{w(\mathbf{r})}{I^{1/2}} \left[n(\mathbf{r}) - \alpha n_s(\mathbf{r}) \right], \tag{25.1}$$

where n and n_s are the observed number density field for the galaxy catalog and synthetic catalog of random objects, respectively. Here we have assigned the galaxies to a regular grid using some mass assignment scheme such as CIC. The factor $w(\mathbf{r})$ is a general weight factor which we discuss shortly. The factor α normalizes the synthetic catalog to the number density of the galaxies, so that $\langle F \rangle = 0$. The field $F(\mathbf{r})$ is normalized by the factor of I, defined as $I \equiv \int d\mathbf{r}\ w^2 \bar{n}^2(\mathbf{r})$.

The estimator for the multipole moments (recall that we are in red shift space) of the power spectrum is

$$\hat{P}_\ell(k) = \frac{2\ell+1}{I} \int \frac{d\Omega_k}{4\pi} \left[\int d\mathbf{r}_1 \int d\mathbf{r}_2\ F(\mathbf{r}_1) F(\mathbf{r}_2) e^{i\mathbf{k}\cdot(\mathbf{r}_1 - \mathbf{r}_2)} \mathcal{L}_\ell(\hat{\mathbf{k}} \cdot \hat{\mathbf{r}}_h) - P_\ell^{\rm noise}(\mathbf{k}) \right], \tag{25.2}$$

where Ω_k represents the solid angle in Fourier space, $\mathbf{r}_h \equiv (\mathbf{r}_1 + \mathbf{r}_2)/2$ is the line-of-sight to the mid-point of the pair of objects, and $\mathcal{L}_\ell$ is the Legendre polynomial of order ℓ. The shot noise $P_\ell^{\rm noise}$ is

$$P_\ell^{\rm noise}(\mathbf{k}) = (1+\alpha) \int d\mathbf{r}\ \bar{n}(\mathbf{r}) w^2(\mathbf{r}) \mathcal{L}_\ell(\hat{\mathbf{k}} \cdot \hat{\mathbf{r}}), \tag{25.3}$$

This expression can be simplified further to obtain the **Yamamoto estimator**.

FKP showed that a good power spectrum estimator can be obtained with the **FKP weight**

$$w(\mathbf{r}) = \frac{1}{1 + \bar{n}(\mathbf{r}) P(\mathbf{k})} \tag{25.4}$$

where $\bar{n}$ is the unclustered mean number density and $P(\mathbf{k})$ a fiducial power spectrum. This weight thus down-weights regions that we observed deeply. The FKP estimator $\hat{P}_\ell(k)$ estimates the power spectrum convolved with $w(\mathbf{r})$. So when we compare the measured power spectrum to the theory, we also need to convolve the theory prediction with $w(\mathbf{r})$.

25.2 Covariance matrix estimation

To extract cosmological parameters for the estimated $\hat{P}_\ell(k)$ we lack one last crucial ingredient, the covariance matrix of $\hat{P}_\ell(k)$. Once we have that, we can make a Gaussian likelihood as we did in Sec. 10.3.1 for N-body data, and fit our theory model to the estimates.

Calculating the convariance matrix analytically is in general not possible although approximations exists. Not only do we need to take into account the survey geometry, but unlike the CMB, now also the observed galaxy modes are correlated due to non-linear evolution. The covariance matrix is thus estimated from simulations, sometimes called **mock catalogs**. These simulations have gravitational clustering in them, they are not the same thing as the random catalogs above. Making realistic simulations of galaxy survey volumes is computationally intense and one uses simplified simulations rather than full N-body dynamics. The required number of simulations is order 1000 for a typical power spectrum pipeline.

The covariance matrix can be extracted from the mocks as

$$C_{ij}^{(\ell\ell')} = \frac{1}{N_m - 1} \sum_{n=1}^{N_m} \left[P_{\ell,n}(k_i) - \bar{P}_\ell(k_i) \right] \left[P_{\ell',n}(k_j) - \bar{P}_{\ell'}(k_j) \right], \tag{25.5}$$

where N_m is the number of mock catalogs and $\bar{P}_\ell(k)$ is the mean power spectrum,

$$\bar{P}_\ell(k) = \frac{1}{N_m} \sum_{n=1}^{N_m} P_{\ell,n}(k). \tag{25.6}$$

This is done at a fiducial cosmology. There are some subtleties with covariance matrix estimation, see in particular the **Hartlap factor correction** which affects the inverse covariance matrix at the level of a few percent.

26 Non-Gaussianity

Let's briefly discuss going beyond the power spectrum. Here we are concerned not primarily with primordial non-Gaussianity (see Sec. 16 in the CMB unit), but rather with gravitational and baryonic interaction.

26.1 Tightening measurements of cosmological parameters

In galaxy surveys, unlike the CMB, higher N-point functions are non-zero even in the absence of primordial non-Gaussianity. This is of course because of non-linear coupling. Two fields that have the same power spectrum can look very different, because the difference can be encoded in higher order correlation functions. It is now quite common to measure also the **galaxy bispectrum**

$$B_g(k_1, k_2, k_3) \sim \langle \delta_g(\mathbf{k}_1) \delta_g(\mathbf{k}_2) \delta_g(\mathbf{k}_3) \rangle \tag{26.1}$$

and extract cosmological parameters from it, together with the power spectrum. Note that the bispectrum and power spectrum estimators have a covariance, they are not independent, due to mode coupling. At perturbative scales which we can use for cosmological analysis, including the bispectrum improves cosmological parameters by 10 to 30% (2206.08327). Bispectrum parameter estimation works the same as power spectrum parameter estimation, i.e. we need a bispectrum estimator, a theoretical model of the bispectrum and a likelihood with covariance.

Of course there are even higher point correlation functions. The next is the **galaxy trispectrum**

$$T_g(k_1, k_2, k_3, k_4) \sim \langle \delta_g(\mathbf{k}_1) \delta_g(\mathbf{k}_2) \delta_g(\mathbf{k}_3) \delta_g(\mathbf{k}_4) \rangle \tag{26.2}$$

The trispectrum is not yet normally used for galaxy survey analysis, but should squeeze some more signal-to-noise out of cosmological parameter constraints (in particular by breaking degeneracies with biases). Higher N-point functions become progressively more difficult to model theoretically and more computationally difficult to estimate in the data. In the perturbative regime, higher N-point functions have progressively less signal-to-noise, since they are higher order in the small initial perturbations. So there is no point in continuing this to ever higher order correlators. On non-perturbative scales, it is likely that N-point functions are not the right thing to do, as we discuss below.

26.2 Primordial non-Gaussianity

Higher N-point functions are also a way to measure primordial non-Gaussianity (e.g. review 1412.4671). As in the CMB, in general the most promising observable is the bispectrum. The problem with non-Gaussianity estimation is to tell apart the signal coming from non-linear evolution and that of primordial origin. The degeneracy of the two signals severely degrades constraints on primordial non-Gaussianity from galaxy surveys. Even next generation galaxy surveys can only about equal (2211.14899) existing constraints from Planck for equilateral and orthogonal non-Gaussianity. However in the far future, we hope that intensity mapping of the dark ages can improve constraints by orders of magnitude (1610.06559).

The situation is better for local non-Gaussianity, or any signal that peaks in the squeezed limit. Interestingly, in that case there is an observable signal in the galaxy power spectrum called **scale-dependent bias**. Scale dependent bias is likely to improve the constraint on f_{NL}^{local} by a factor of 10 or so over Planck, within the next 10 years. Scale-dependent bias leads to a characteristic kink of the primordial power spectrum on large scales. I have spent a lot of time with this signal in my own research and hope to add a discussion here later.

27 Galaxy Weak Lensing

So far we have focussed our discussion of large-scale structure on the galaxy density. There is a second probe of the matter distribution using galaxies, which is **weak lensing of galaxies**. Reviews of galaxy weak lensing include 1612.06535, 1710.03235, 2007.00506.

The key to weak lensing is measuring the subtle **distortions in the shapes of galaxies caused by gravitational lensing** due to the matter distribution between the source galaxy and us. These distortions are typically small (a few percent or less). The lensing effect is thus weak and not detectable in individual galaxies but becomes apparent when analyzing the shapes of many galaxies statistically. As is the case for CMB lensing, one can use the measured distortions to reconstruct the underlying mass map. Of course, measuring galaxy shapes is difficult and many systematic errors have to be overcome, such as atmospheric distortion, imperfections in the telescope optics, and the intrinsic shapes of galaxies and their intrinsic alignment with eachother.

The result of these statistical shape measurements is the **galaxy lensing convergence field $\boldsymbol{\kappa_{\mathbf{g}}}$**. Once measured, its power spectrum can be used to constrain cosmological parameters. Lensing has an advantage over galaxy clustering in that it is mostly sensitive to dark matter, and thus much less sensitive to baryonic physics than galaxy clustering. For that reason one can use lensing to probe somewhat higher k scales reliably than is possible with galaxy clustering.

Galaxy weak lensing is often used in cross-correlation with galaxy clustering and even CMB lensing. Let's now discuss these cross-correlations. We use capital Roman subscripts to denote observables, $A, B \in \{\delta_{\rm g}, \kappa_{\rm g}, \kappa_{\rm CMB}\}$, where $\delta_{\rm g}$ denotes the density contrast of lens galaxies, $\kappa_{\rm g}$ the lensing convergence of source galaxies, and $\kappa_{\rm CMB}$ the CMB lensing convergence. Using the galaxy data only we get the so-called **3x2 analysis** which includes

- galaxy clustering ($C_\ell^{g_i g_j}$)
- galaxy-galaxy lensing ($C_\ell^{g_i \kappa_{{\rm gal},j}}$)

- cosmic shear tomography ($C_\ell^{\kappa_{\rm gal},i\kappa_{\rm gal},j}$).

The indices i and j indicate red-shift bins. Adding the CMB lensing convergence field, we can extend the data vector with 3 more two-point functions:

- galaxy-CMB lensing ($C_\ell^{g_i\kappa_{\rm CMB}}$)
- CMB lensing-galaxy lensing ($C_\ell^{\kappa_{\rm CMB}\kappa_{{\rm gal},j}}$).
- CMB lensing power spectrum ($C_\ell^{\kappa_{\rm CMB}\kappa_{\rm CMB}}$)

This can be called a **6x2 analysis**. The angular power spectrum between redshift bin i of observable A and redshift bin j of observable B at Fourier mode ℓ (using the Limber approximation) is given by

$$C_{AB}^{ij}(\ell) = \int d\chi \frac{W_A^i(\chi) W_B^j(\chi)}{\chi^2} P_m\left(\frac{\ell+1/2}{\chi}, z(\chi)\right) \ , \tag{27.1}$$

where χ is the comoving distance, $P_m(k,z)$ is the matter power spectrum, and $W_A^i(\chi), W_B^j(\chi)$ are weight functions of the observables A, B given by

$$W_{\delta_{\rm g}}^i(\chi) = b_g^i \frac{n_{\rm lens}^i(z(\chi))}{\bar{n}_{\rm lens}^i} \frac{dz}{d\chi} \ , \tag{27.2}$$

$$W_{\kappa_{\rm g}}^i(\chi) = \frac{3H_0^2\Omega_m}{2c^2} \frac{\chi}{a(\chi)} \int_{\chi_{\rm min}^i}^{\chi_{\rm max}^i} d\chi' \frac{n_{\rm source}^i(z(\chi'))}{\bar{n}_{\rm source}^i} \frac{dz}{d\chi'} \frac{\chi'-\chi}{\chi'} \ , \tag{27.3}$$

$$W_{\kappa_{\rm CMB}}(\chi) = \frac{3H_0^2\Omega_m}{2c^2} \frac{\chi}{a(\chi)} \frac{\chi^*-\chi}{\chi^*} \ , \tag{27.4}$$

where $\chi^i_{\rm min/max}$ are the minimum and maximum comoving distance of the redshift bin i. Here $a(\chi)$ is the scale factor, Ω_m the matter density fraction at present, H_0 the Hubble constant, b_g^i is the galaxy bias in bin i, and χ^* the comoving distance to the surface of last scattering. Note that the weight function of $\kappa_{\rm CMB}$ does not depend on redshift bins. The galaxy density and CMB convergence weight functions we have encountered before in these lectures. The galaxy lensing weight function integrates the lensing effect over the source density in a bin i. Details on cross-correlation analyses are given in 1607.01761 and 2108.00658. Of course, one can also consider bispectra involving the three signals.

28 Modern Inference Methods

In this section we discuss modern inference techniques that leverage machine learning and/or auto-differentiation. Machine learning in cosmology is reviewed for example here: 2210.01813. I will cite some references here, with preference to papers I know (i.e. the list is not ordered by precedence or importance).

28.1 Overview

In recent years, a lot of effort is made in the community to go beyond power spectra, bispectra and Gaussian likelihood approximations. The hope of course is to extract more sensitive parameter constraints from the data. The broad tools we use for this include simulations, optimization (auto-differentiation), and the many forms of machine learning. I will try to give you a broad overview with suitable references to study more. Despite massive effort, it is still somewhat debated whether these methods really allow us to get better parameter constraints from real experiments. This is because the methods need to be robust with respect to non-linear small-scale physics, which is difficult to achieve. Currently most state-of-the-art constraints still come from a more traditional analysis. See 2405.02252 for a recent quantitative comparison of some of these methods with traditional approaches.

Modern methods can be broadly classified into two different categories, which we dicuss in more detail below:

- **Simulation-based inference (SBI)**, also called **Likelihood-free inference (LFI)** or **implicit inference** uses simulations to learn how a **summary statistic** depends on cosmological parameters. The summary statistic can itself be learned. For example, we can train a neural network on simulations of galaxy catalogs to infer cosmological parameters from them.

- **Probabilistic forward modelling**, also called **explicit inference** or **Bayesian Hierarchical Modelling**. This means that we keep track of all latent variables of the simulator. In practice it means that we jointly reconstruct the initial conditions of the universe together with cosmological parameters, as we will see.

In both cases we need a **forward model**, which can be a simulator, a neural network, or even analytic perturbation theory. The **forward model maps cosmological parameters and initial conditions to observable data**. Of course a crucial aspect of the forward model is that it is accurate at the scales we are interested in and that it is computationally tractable to evaluate (as often as required by the chosen parameter inference method). These conditions are not easy to meet.

28.2 Simulation-based Inference

For a review of SBI see 1911.01429. A code implementation of SBI is https://sbi-dev.github.io/sbi/. SBI can also be implemented using more general probabilistic machine learning packages such as https://docs.pyro.ai/. More details on SBI can also be found in my ML in Physics lecture slides here https://ai.physics.wisc.edu/teaching/.

28.2.1 Summary Statistics

The first step of SBI involves **compressing the data d** into **summary statistics x**. The maximum possible compression is a summary statistic of the same dimensionality as the parameters Θ we want to learn (1712.00012). We want summary statistics to be as good as possible, ideally they would be **optimal** which means they lose no information.

Here are some summary statistics that are useful in cosmology:

- **Power spectrum and bispectrum** of course.
- **Cluster or Galaxy number density and mass distribution.** Cluster counting, especially of very massive clusters which are less sensitive to baryonic physics, can be used to constrain cosmological parameters.
- **Moments of the matter distribution** $\left\langle \delta_m^N(x) \right\rangle$ as a function of smoothing scale.
- **Wavelet scattering transform.** Uses the distribution of wavelet transform coefficients for various scales.
- **Topological data analysis.** Aims to use the distribution of topological features ("simplices") in the data.
- **Minkowski functionals.** Some other morphological characterization.

This list is not complete but probably covers the most important ones.

28.2.2 Learned Summary Statistics

A learned summary statistic, e.g. a **3d convolutional neural network, or a graph neural network** applied to the galaxy field and trained to measure cosmological parameters, is in principle optimal assuming that

- It has enough capacity (it can represent the required function with its weights).
- There was enough training data to learn the required function.
- The optimizer did find the global minimum.
- The simulation can be trusted on the scales that the neural network gets to see (e.g. we can filter out small scales first to make it more robust, but losing sensitivity).

These conditions don't neccessarily hold in practice so there is still some interest in coming up with new "hand made" summary statistics. A neural network is usually trained to directly give estimates of the cosmological parameters, while for other summary statistics there is a second step involved in mapping them to cosmological parameters. A neural network can also be trained to estimate error bars and covariances for its measurements. However, a more robust approach is to learn these error bars after training in a second step, which we discuss now.

28.2.3 Neural Density Estimation (NDE)

After having obtained a measurement of the summary statistics $\mathbf{x}_i$, either learned or not or a combination thereof, the second step is to infer parameters Θ. Sometimes it is a good approximation to assume that the likelihood is Gaussian with a fiducial covariance matrix, which we can estimate from simulations as discussed for the power spectrum above. More generally, the relation has to be learned probabilistically using a (neural) density estimator.

To do so, we first need to create a dataset of parameters and simulated summaries

$$(\boldsymbol{\theta}_n, \mathbf{x}_n)\}_{n=1}^{N_{\mathrm{sim}}} \tag{28.1}$$

Now we want to learn either of the following functions:

- In **Neural Likelihood Estimation (NLE)** we learn the likelihood $p(\mathbf{x}|\boldsymbol{\theta})$, i.e. the conditional probability of the summary $\mathbf{x}$ given the parameters $\boldsymbol{\theta}$. If we have learned the full distribution, we can do fast **amortized inference**, which means that we do not need to run new simulations to get the posterior for a new set of observations. On the other hand, it can be too expensive to run enough such simulations, in which case one can use a **Sequential Neural Likelihood Estimator (SNLE)** which focuses on learning the likelihood near the observed data. Once the neural likelihood is learned, we multiply by a prior and run the ususal MCMC.

- In **Neural Posterior Estimation** one learns directly $p(\boldsymbol{\theta}|\mathbf{x})$. One thus does not have to run an MCMC anymore.

- In **Neural Ratio Estimation** we learn the likelihood-to-evidence ratio using a classifier.

To learn either of the likelihood or posterior we need a **parametric density model**, that is some function that we can fit to the data set Eq. (28.1) by adjusting its parameters. The state-of-the-art to do this is to use so-called **normalizing flows**, for example the **masked autoregressive flow (MAF)**. A normalizing flow is a neural network that transforms a simple base distribution (usually a Gaussian) into a complicated target distribtion by learning a series of diffeomorphisms (i.e. an invertible and differentiable change of coordinates). Fitting/Training the normalizing flows works by adjusting its weights using auto-differentiation, in the same way as ordinary neural networks are trained (though with a different loss function). By now SBI methods have been fairly well established and, as in the case of MCMC, you don't neccessarily have to understand the methods in great detail to use them. A key challenge is to check that the learned likelihood or posterior is correct (especially that it is not over confident). One approach to do so is called **Simulation-based calibration**. My lecture slides on AI in Physics https://ai.physics.wisc.edu/teaching/ contain more details of the above methods, for example the required neural network training objectives.

28.2.4 SBI results in cosmology

One large project in cosmology that uses various summary statistics and SBI for parameter inference is the SimBig project. Recently this project analyzed BOSS data in 2211.00723, 2310.15246. The simulation training data comes from the ~ 20.000 Quijote dark matter simulations, which are forward modeled to include galaxies (through an HOD), survey geometry and observational systematics. The result is that they can tighten parameter constraints by a moderate factor over the power spectrum analysis. This is a large and encouraging effort. However it is still somewhat unclear how it compares to an EFT based analysis. The improvement factor depends on the non-linear cutoff scale, and in simulations it is not possible to marginalize over biases in the same way as we can in the perturbative analysis. So as soon as we tighten constraints over the EFT (including power spectrum, bispectrum and perhaps trispectrum) one has to be careful about the robustness of the measurement. Further, the Quijote simulations have limited volumes, span a limited model and parameter space, and do not include any hydrodynamics. In the foreseeable future we will use both EFT and SBI analyses and see how well they agree.

28.2.5 Theory emulators

An approach that is related to SBI and getting popular in cosmology is the generation of **neural network based emulators** of summary statistics, in particular of the power spectrum. Even for linear physics, running CAMB or CLASS at each point in the Monte Carlo chain is annoyingly slow. To speed this up, cosmologists have trained neural networks to emulate the Boltzmann solver, i.e. provide the power spectrum $P^{theo,lin}(k,\Theta)$ as a function of Θ. An example of this is Cosmopower 2106.03846. Using an emulator, the MCMC will run much faster than with the full Boltzmann solver.

Using simulations with different cosmological parameters, one can also make a power spectrum emulator of the non-linear matter power spectrum. This was done for example here 2207.12345. It is also possible to combine dark matter simulations with a biasing model (1910.07097) to obtain a bias dependent emulator of the halo (or galaxy) power spectrum. Emulators can be paired with SBI by learning their likelihood with NDEs.

28.3 Probabilistic Forward Modeling at Field Level

A second approach to cosmological parameter estimation removes summary statistics and instead works with the entire PDF of the data. This approach has the advantage that there are no black boxes (such as learned summary statistics) and systematics are easier to model since everything is treated probabilistically. The downside is that it takes considerably more, often intractable amounts of, computational resources. Let's see how this works.

Let's first assume that we have a deterministic simulator $f(\mathbf{s},\Theta)$, our **forward model**. It maps the (near-)Gaussian initial conditions $\mathbf{s}$ (the primordial density perturbations, usually including the BAO feature) to the observable smoothed galaxy density field $\mathbf{d}$ assuming cosmological parameters Θ (e.g. Ω_m, H_0). For example, f can be a dark matter simulation with a bias model for galaxies, and s would be the N^3_{side} initial displacements. Second, we assume that our observed galaxy density $\mathbf{d}^{obs}$ is equal to the true density $\mathbf{d}$ plus some uncorrelated Gaussian noise n (e.g. shot noise):

$$\mathbf{d}^{obs} = \mathbf{d} + \mathbf{n} \tag{28.2}$$

where the noise has covariance N. The likelihood of observing $\mathbf{d}^{obs}$ given s is thus given by

$$\log \mathcal{L}(\mathbf{d}|\mathbf{s},\Theta) = -\frac{1}{2}(f(\mathbf{s},\Theta) - \mathbf{d}^{obs})^T N^{-1} (f(\mathbf{s},\Theta) - \mathbf{d}^{obs}) + \text{const.} \tag{28.3}$$

We want to turn this around and get the posterior $\mathcal{P}(\mathbf{s},\Theta|\mathbf{d})$. This will give us the joint PDF of the initial conditions of the universe and the cosmological parameters, and thus measure both of them. We thus need a prior on $\mathbf{s}$, which is that it is a Gaussian field with some primordial parameters Θ' (e.g. A_s, n_s) that define the primordial power spectrum. The prior is

$$\log \mathcal{P}(\mathbf{s},\Theta') = \log \mathcal{P}(\mathbf{s}|\Theta') + \log \mathcal{P}(\Theta') \tag{28.4}$$

$$= -\frac{1}{2}\mathbf{s}^T S(\Theta')^{-1}\mathbf{s} - \frac{1}{2}\log|S(\Theta')| + \log \mathcal{P}(\Theta') \tag{28.5}$$

$$\tag{28.6}$$

Using Bayes theorem we get the posterior

$$\log P(\mathbf{s}, \Theta, \Theta'|\mathbf{d}) = -\frac{1}{2}(f(\mathbf{s}, \Theta) - \mathbf{d}^{obs})^T N^{-1}(f(\mathbf{s}, \Theta) - \mathbf{d}^{obs}) - \frac{1}{2}\mathbf{s}^T S^{-1}(\Theta')\mathbf{s} \tag{28.7}$$

$$- \frac{1}{2}\log|S(\Theta')| + \log \mathcal{P}(\Theta') + \log \mathcal{P}(\Theta) + \text{const.} \tag{28.8}$$

Usually we don't care about the initial conditions (the "phases") of the perturbations and only want to know the cosmological parameters. In this case we need to marginalize to get

$$P(\Theta, \Theta'|\mathbf{d}) = \int d\mathbf{s} P(\mathbf{s}, \Theta, \Theta'|\mathbf{d}) \tag{28.9}$$

The posterior Eq. (28.7) is not so different conceptually from the posterior of the power spectrum, but here our variables are the entire cosmological field. The reason it is even possible to write down a posterior PDF for the field is that we know the PDF of the initial conditions, and we can forward model the field to late times. Assuming that our forward model is correct, this analysis will be statistically optimal, i.e. it includes all available information. The problem is to handle such a huge computation problem. The data $\mathbf{d}$ and initial conditions $\mathbf{s}$ can be easily 1000^3 dimensional. A normal MCMC would never converge. Fortunately there are techniques for extremely high dimensional inference. Before discussing inference let's get an overview of the forward models. Forward modelling does not have to be done with the galaxy density of course, weak lensing or intensity mapping is also an appealing target.

There are **different types of forward models** that are being used, in particular:

- **Pertubation theory at field level** plus a bias expansion (1808.02002). This case is the most tractable, but can of course not go beyond the regime of perturbation theory. It is still somewhat unclear to what extent a PT forward model can outperform the traditional analysis (2307.04706).
- **Differentiable simulations** (2010.11847,2211.09958), which include both structure formation and some approximation of halo formation.
- **Neural network emulators** of structure and halo formation (1811.06533,2206.04594).
- **Hybrid EFT** (1910.07097), a combination of dark matter simulations and a bias expansion in Lagrangian space.

Recall however that to draw a single sample of the posterior, we need to call the forward model. So our forward model should be fast enough to be evaluated millions of times. Above we have not written out nuisance parameters of the forward model and one can also include stochastic sampling to model galaxy formation, but our discussion captures the essential features. As a side note, these same forward models (minus the requirement for differentiability) can also be used to get the SBI training data in Eq. (28.1).

Next to the forward model we need an **inference algorithm** to approximate the posterior and/or sample from it. All of these require that the forward model is differentiable with respect to all parameters, including the initial conditions $\mathbf{s}$. This is where auto-differentiation, for example Jax or pytorch, comes in. Inference algorithms that have been proposed include:

- Finding the MAP and making a Gaussian approximation around it for error bars (1706.06645). Finding the MAP by gradient descent is faster than sampling, but it is hard to get reliable error bars.
- Hamiltonian Monte Carlo (HMC), (1203.3639). This is the most reliable but also most computationally intense approach. Recently a different variant of Monte Carlo was used that is also promising: 2307.09504.
- Variational inference or Variational Self Boosting with normalizing flows (2206.15433).

In all of these cases it is difficult to deal with a **multimodal posterior** which is expected at small scales. Even without that problem, it is difficult to generate enough independent samples and be sure the posterior surface is well covered. Also, not all parameters are created equal and it is hard for example to sample from band powers of the initial power spectrum. While appealing in principle, it is still very hard to use this approach in practice on data, especially with a non-perturbative forward model. On the other hand, forward modeling can in principle strongly improve constraints by breaking parameter degeneracies present in N-point function analysis (2112.14645) and is the only provably optimal approach.

28.3.1 Reconstruction of initial conditions

I want to briefly look at the problem above from the point of view of **solving an inverse problem**. Assume that we want to reconstruct the unobservable initial perturbations $\mathbf{s}$ from the observed noisy data $\mathbf{d}^{obs}$, for known fiducial cosmology parameters. This kind of problem is called an inverse problem in statistics and computer science and many algorithms exist to solve such problems, making various approximations. Note that the problem is ill-conditioned. Because we don't observe the transverse velocities of galaxies, and for other reasons, we cannot simply simulate the observed field backwards in time. In our **forward reconstruction** of initial conditions above, this problem is solved by having a prior on the initial conditions, that makes the problem probabilistically invertible. There is a second class of algorithms to reconstruct the initial conditions, called **backwards reconstruction**. This methods starts from the observed field and uses a sort of backwards perturbation theory to reconstruct the initial conditions. The simplest form of this algorithm is called **standard (BAO) reconstruction**. It was devised to sharpen the BAO feature which is being somewhat "washed out" by large-scale "bulk movements" of matter. By first undoing the Zeldovich displacement (1-LPT), one makes the feature sharper and can thus improve cosmological parameters from a power spectrum measurement. This method is widely used in galaxy survey analysis. There are also more advanced **iterative initial condition reconstruction** methods (1704.06634). In addition to probing BAO, one can imagine running a bispectrum estimator on the reconstructed initial conditions to search for primordial non-Gaussianity.

Forward reconstruction is more powerful in principle because it is clear how to include systematics (by adding them to the forward model). Backwards reconstruction on the other hand is far less computationally demanding. Neural networks can also be trained to perform backwards reconstruction, see e.g. 2104.12864,2305.07018. One can also learn a probabilistic initial conditions reconstruction with a diffusion model from N-body simulations (2304.03788) .

28.4 Generative Machine Learning at field or point cloud level

I want to briefly mention one more major area of machine learning research in cosmology, that of generative modeling at the field level. A generative model can **emulate a simulation** and be potentially much faster than the original simulation that it was trained on. Making simulations (or emulations if you prefer) is also possible without generative (probabilistic) modelling. For example one can train a deterministic U-net to go from initial conditions, which are very fast to generate, to the late time matter distribution. Generative modeling on the other hand includes a step of random sampling, i.e. every time we run the machine learning model we get a different result.

The main machine learning models that can generate cosmological distributions such as $\delta_m(\mathbf{x})$ at high resolution are

- **GANs and WGANs** (2001.05519)
- **Diffusion Models** (2311.05217)
- **Normalizing flows** (2105.12024, 2202.05282)

and all of them have been used in cosmology. One can model either density fields or displacement fields of particles. Very recently, people also work with point cloud models (2311.17141), that don't generate a field (image) but a set of points (galaxies). All three generative methods can also work with point clouds.

A main use for generative models is to **speed up simulations**. For example, one may be able to upgrade a low-resolution dark matter simulation to an emulated high resolution hydro simulation. Or one may populate a low resolution dark matter simulation with realistic galaxies. Of course, to train all of these models one needs to have some high-resolution training simulations. The generation process can also be conditioned on cosmological parameters. Such conditional generative models can also be run in inverse mode to **estimate cosmological parameters**. There is a rapidly growing literature in this field and I have only cited a small subset thereof.